UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Using Social Media to Help Spanish Speakers
Improve Word and Sentence Stress:
The Case of Facebook and YouTube

El Uso de las Redes Sociales
para que los Hispanohablantes Mejoren
su Acento Léxico y Prosódico:
el Caso de Facebook y YouTube

D. Kim Heineth Jensen

2021

UNIVERSIDAD DE MURCIA

FACULTAD DE LETRAS

Using social media to help Spanish speakers improve word and sentence stress: The case of Facebook and YouTube

El uso de las redes sociales para que los hispanohablantes mejoren su acento léxico y prosódico: El caso de Facebook y YouTube

Director: Dr. José Antonio González Mompeán

D. Kim Heineth Jensen

2021

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgements

Although writing a Ph.D. can be a lonely experience, there is no doubt that many people have been involved in the process over the past four years. First of all, I would like to thank my supervisor, José Antonio Mompeán González, without whom this project would never have got off the ground. You have been such an important part of developing my initial proposal from a basic idea to the final product that is presented here. Your wealth of knowledge has been a constant inspiration, and your many… many comments and corrections have improved my initial drafts significantly.

I also want to direct a big thank you to some of the professors at the University of Murcia. Thank you to Rachel Criado Sánchez, Inma Arboleda Guirao, and María del Mar Vilar García for allowing me to recruit participants from your courses. These have been essential for the completion of this project.

I would also like to give special thanks to Jan Kennedy. Without you, I would not have had a place to call home for large parts of this project, and your constant support and interest in my study have encouraged me to keep at it – even through periods when I have found the project particularly challenging.

I am also thankful to James Munro and Helena Kennedy for providing speech samples for the test and training materials used in this thesis. In addition to lending your voices to my project, you have also lent me your ears when I have needed to vent my frustrations – whether they have been related to my project or to life in general. I hope that from now on, 'bitter' will only apply to my pints and less so to my persona.

I am also grateful to Gang Yao. Your help with the linguistic barriers I have encountered on the programme has been incredibly valuable. Most of all, I am thankful for your friendship over the past years. Being able to meet up with you and discuss anything from *t*-tests to Chinese cuisine has been more helpful than you know. I only hope that I have helped you as much as you have helped me.

My final and most heartfelt thanks go to Helena Kennedy. I have already mentioned you once, but no one deserves to be mentioned and praised like you do. Without your love and support, I doubt I would have finished this project. You have always been there for me – even when we were apart. You have made countless sacrifices to allow me to pursue my dream, and I am forever grateful to you. I hope and wish that my work will lead to a better future for the both of us.

Abstract

In the past few decades, pronunciation teaching has been a neglected skill in ESL/EFL classrooms (Elliott, 1995; Kelly, 1969). This is of course a problem as learning both segmental and supra-segmental aspects of a new language is vital to successful communication (Celce-Murcia, Brinton, and Goodwin, 2010). Various reasons have been identified to explain why pronunciation has suffered a Cinderella existence in recent years such as a lack of quality materials, insufficient teacher training, and poorly designed curricula forced onto teachers by school boards or principals (Henderson et al., 2012). One solution to these problems could be to take parts of the pronunciation teaching out of the classroom and allow learners to work independently online at a time of their choosing.

To investigate two possible avenues for doing so, this thesis explores the applicability of two popular social media platforms (Facebook and YouTube) in teaching specific aspects of English supra-segmental phonology. These platforms have been used to teach other aspects of English linguistics with writing receiving more attention than any other skill. However, very few studies have focused on pronunciation, and those that have, have looked solely at student perceptions and therefore failed to gauge the effectiveness of the training provided. Thus, in the two studies presented in this thesis, applicability was measured through students' ability to learn the given rules, their improvement from pre- to post-test as well as the students' evaluation of the course in the post-test questionnaire.

For the two studies, a total of 147 participants were recruited and divided into two experimental groups and a control group which served as control group for both studies. The experimental participants took part in a four-week training course delivered through the aforementioned social media platforms. Depending on the study they participated in, the participants were tasked with learning either word

stress rules (Facebook) or sentences stress rules (YouTube). The two courses had the same structure with videos being posted on the Monday and Thursday of each of the first three weeks and with the last week used for additional practice. The topic of the word stress rules were compound nouns describing placenames, food and ingredients or OBJECT/AGENT relationships (e.g. 'bricklayer'). The sentences stress rules dealt with three types of non-final tonicity, namely objects of general reference, event sentences, and contrasts. Both courses used dialogues with textually enhanced audio scripts as the main tool for helping the participants notice the stressed words in the dialogues (Sharwood Smith, 1993).

Although not all results were statistically significant, the results of both studies generally show that the experimental participants improved from pre-test to post-test in their ability to identify and produce either word or sentence stress correctly. This was a surprising finding in the sense that the studies also found that many of the participants failed to accurately state the rules that had been presented to them on the course. Instead, the results seem to indicate that simply working with the materials improved the participants' perception and subconscious control of the language despite not teaching them the rules.

In order to evaluate the viability of using social media for English pronunciation teaching, the participants were asked to fill out a post-test questionnaire in which they rated several aspects of the course and could give recommendations for improvements. The participants' responses were very encouraging indeed as the vast majority of participants stated that they found the course useful. However, a small tendency was observed for participants in both studies to favour YouTube as a mode of delivery. Implications of the test results and questionnaire responses are discussed.

Resumen

Dominar un idioma extranjero o un segundo idioma (L2) no es una tarea fácil. Entre otros aspectos, requiere que el alumno aprenda vocabulario nuevo, comprenda un nuevo tipo de sintaxis y perciba y produzca sonidos y ritmos del nuevo idioma. La pronunciación, en particular, a menudo plantea un desafío para los estudiantes. Probablemente esto es porque hay múltiples factores involucrados en lo que respecta a la pronunciación. Por ejemplo, los alumnos pueden tener que desarrollar la capacidad física para producir nuevos sonidos. Sin embargo, lo que puede ser aún más difícil es la capacidad de percibir y producir contrastes segmentales que no se utilizan en la lengua materna del alumno (Flege, 1995). Lo difícil que pueden ser algunas partes del aprendizaje del idioma extranjero se refleja en la cantidad de formación que han recibido algunos estudiantes en las investigaciones anteriores sobre la pronunciación. En algunos casos, los participantes han trabajado varias horas en un solo contraste fonémico (Akahane-Yamada et al., 1998). Sin embargo, ni siquiera después de una formación extensa, los participantes rara vez desarrollan un 100% de precisión en estos contrastes.

Las mismas cuestiones desde el punto de vista de la percepción y producción también se aplican a aspectos como el acento y la entonación, conocidos como rasgos suprasegmentales. Históricamente, menos trabajo se han centrado en la fonología suprasegmental en comparación a la fonología segmental. Sin embargo, se han propuesto dos modelos. Uno, el "Stress Deafness Model (SDM)" (Dupoux et al., 1997) sugiere que los hablantes que no usan el acento de manera contrastiva (por ejemplo, hablantes del francés) no podrán percibir cambios en los modelos del acento a nivel léxico. Otro modelo, el "Typology Stress Model (SDM)" (Altmann, 2006), va un paso más allá al hacer predicciones tanto para la percepción como para la producción del acento léxico. Según este modelo, un conjunto de parámetros fonológicos determina

qué dificultad encuentra un alumno la percepción y producción del acento léxico en un segundo idioma.

Considerando lo difícil que es aprender modelos de pronunciación, es casi irónico que, en las últimas décadas, la enseñanza de pronunciación se ha convertido en una habilidad frecuentemente ignorada en las clases de inglés como lengua extranjera (Elliott, 1995; Kelly, 1969). Evidentemente, esto representa un problema porque el aprendizaje de aspectos tanto segmentales como suprasegmentales de un idioma nuevo es vital para una comunicación exitosa (Celce-Murcia, Brinton y Jones, 2010). Diversas razones explican por qué durante los últimos años la pronunciación ha sido desatendida inmerecidamente, tales como una falta de materiales de calidad, una insuficiente formación del profesorado, y un currículo pésimamente diseñado por juntas o directores educativos (Henderson et al., 2012).

El último de estos factores presenta un obstáculo real para los profesores porque un currículo mal diseñado puede hacer que los profesores carezcan del tiempo necesario para centrarse en la enseñanza de la pronunciación. Este es un problema considerable ya que, como se indicó anteriormente, la pronunciación de una segunda lengua es un aspecto particularmente difícil de adquirir.

Una solución para los problemas mencionados anteriormente se puede encontrar en la tecnología. Al sacar partes de la enseñanza de la pronunciación del aula y permitir que los alumnos trabajen de forma independiente en línea, los profesores pueden disponer de más tiempo para concentrarse en otras partes del currículo (Tanner y Landon, 2009). El uso de la tecnología en la enseñanza de idiomas podría tener varios beneficios adicionales. Por ejemplo, es probable que el aprendizaje de idiomas asistido por ordenadores (CALL) haga que los alumnos se sientan más cómodos porque pueden trabajar al ritmo que les conviene y porque pueden trabajar en un entorno que es menos intimidante que el de un aula normal

(Eskenazi, 1996). Varios estudios ya han demostrado que la enseñanza de la pronunciación asistida por ordenadores (CAPT) puede ser eficaz (Elimat y AbuSeileek, 2014; Neri et al., 2008; Seferoglu, 2005). Se pueden adoptar diversos enfoques. Algunos utilizan sistemas de reconocimiento del habla (ASR) para proporcionar a los alumnos una evaluación sobre su habla. Otros se centran en podcasts o ejercicios de “shadowing”. Además, otros recurren a las redes sociales como un medio para enseñar la pronunciación.

Como resultado, y con el objetivo de investigar dos opciones de las muchas existentes, esta tesis se centra en la posible utilidad de dos redes sociales populares, Facebook y YouTube, en la enseñanza de características concretas de la fonología inglesa suprasegmental. Además de ser populares, estas dos plataformas tienen una variedad de características que pueden hacerlas adecuadas para la instrucción de la pronunciación. Facebook, en particular, es una herramienta que se presta a fines didácticos. Esto se debe a una función recientemente añadida llamada "Unidades de Aprendizaje Social". Esta función permite a los profesores estructurar los cursos lección por lección e incluso permite que las lecciones se publiquen a la hora especificada por el profesor. Aunque YouTube carece de esta función, contiene elementos que podrían ser valiosos en un contexto de aprendizaje, a saber, la capacidad de cualquier usuario de subir videos y textos.

Ambas plataformas se han utilizado para enseñar otros aspectos de la lengua inglesa aparte de la pronunciación. La escritura, especialmente, ha recibido mucha más atención que cualquier otra habilidad (p.ej., Yunus y Salehi, 2012). Los resultados en esta área han sido alentadores.

Muy pocos estudios se han centrado en la pronunciación, y los que lo han hecho, han estudiado únicamente la comprensión de los estudiantes. En otras palabras, parece haber una tendencia muy fuerte en el área de investigación a preguntar a los participantes qué pensaban sobre el uso de una determinada tecnología, sin medir realmente la efectividad de la intervención.

Esta tesis da un primer paso para llenar este vacío en la literatura. Por lo tanto, en los dos estudios presentados en esta tesis, la aplicabilidad de estos recursos se midió a través de la capacidad de los estudiantes de español para aprender tres reglas de la pronunciación relacionadas con el acento léxico y prosódico en inglés. Además de medir la efectividad del tratamiento, también se analizaron las percepciones y evaluaciones de los participantes sobre el curso. Esto se realizó mediante un cuestionario después del test final utilizando ítems en una escala de Likert que les permitió calificar varios aspectos del curso. Además de estos, el cuestionario después del test final también contenía una serie de preguntas abiertas en las que los participantes podían dar recomendaciones para mejorar el uso de los recursos utilizados.

Un total de 147 estudiantes de inglés como lengua extranjera participaron en los dos estudios. Se trataba principalmente de estudiantes de varios niveles en la Grado en Estudios Ingleses de la Universidad de Murcia (también se invitó a participar a participantes de escuelas de idiomas locales, pero solo se inscribió un número muy reducido).

Los participantes se dividieron en dos grupos experimentales (un grupo que aprendió a través de Facebook y otro que aprendió a través de YouTube) así como un grupo de control. Los participantes del grupo de control actuaron como controles para ambos estudios.

Siguiendo un diseño de test inicial / final, los participantes primero tomaron un test inicial seguido de cuatro semanas de formación o ninguna intervención. Para garantizar que todos los participantes tuvieran la oportunidad de mejorar al ser parte del proyecto, todo el material se puso a disposición en línea una vez finalizado el curso. El test final se llevó a cabo inmediatamente después de que finalizó el curso de cuatro semanas.

Los dos cursos tenían la misma estructura. El lunes de cada semana, se subió un video en la plataforma correspondiente. Este video contenía de 10 a 13 diálogos entre dos hablantes nativos de inglés y todos estaban relacionados con el tema tratado en esa semana. Se pidió a los participantes que encontraran tres diálogos que mostraran un modelo diferente al resto (un descubrimiento guiado). Por ejemplo, si el tema de la semana era los nombres de lugares, siete diálogos contendrían un sustantivo compuesto con un acento secundario y tres contendrían un sustantivo compuesto con un acento principal. Además de identificar los elementos que eran diferentes, también se les pidió a los participantes que crearan una regla de pronunciación basada en sus respuestas (p. ej., 'Los nombres de lugares en inglés tienen un acento secundario a menos que terminen en calle; en este caso, el nombre del lugar tiene un acento principal). El jueves siguiente, se subió la regla de pronunciación junto con un nuevo video que presentaba los mismos diálogos. Sin embargo, esta vez, el texto de audio se había mejorado textualmente mediante el uso de letras mayúsculas y negrita. Esto se hizo para aumentar la conciencia de los participantes sobre las reglas de la pronunciación (Sharwood Smith, 1993).

En el Estudio 1 (Facebook), la atención se centró en el acento léxico. Más específicamente, los temas incluidos fueron sustantivos compuestos que describen localidades, comida e ingredientes, y relaciones entre objetos y agentes (por ejemplo, “bricklayer > "albañil"). Se eligió este tema porque los sustantivos compuestos en inglés y español tienen acento diferente (Hualde, 2007). Por lo tanto, era probable que a los participantes les resultara difícil aplicar correctamente el acento léxico en inglés.

En el Estudio 2 (YouTube), la atención se centró en el acento prosódico. Las reglas de acento prosódico trataron tres tipos de tonicidad no final, específicamente objetos de referencia general, frases con sucesos y contrastes. Los dos primeros fueron elegidos en base a la opinión de Ortiz-Lira (1995), quien encontró que los objetos de referencia general y frases con sucesos eran casos de tonicidad no final que sus participantes encontraron particularmente difíciles. Además de estos, también se incluyó el acento contrastivo, ya que el inglés y el español generalmente usan diferentes herramientas lingüísticas para enfocar un elemento. Mientras que el inglés a menudo desplaza el núcleo (es decir, la sílaba con más acento en una frase), el español tiende a utilizar el orden de las palabras para mantener el núcleo al final de la frase (Gutiérrez-Díez, 2005). Por lo tanto, se consideró que estos tres tipos de oraciones probablemente serían difíciles para los hispanohablantes.

Como se mencionó anteriormente, los videos utilizados en los cursos utilizaron diálogos con textos de audio mejorados textualmente como la herramienta principal para ayudar a los participantes a notar las palabras acentuadas en los diálogos. Además de esta herramienta, se incluyeron una variedad de otros factores en el diseño del curso para ayudar a los participantes a aprender las reglas. Estos incluían la práctica de la producción (Swain, 1995), la instrucción explícita (Saito, 2012) y el input auditivo (Krashen, 1983). Por último, también se esperaba que el hecho de que los participantes pudieran elegir el entorno en el que trabajar tendría un impacto positivo en la experiencia de aprendizaje de los participantes (Neri et al., 2002).

Los estudios revelan algunos hallazgos interesantes. En primer lugar, la predicción realizada en el Estudio 1 (que los participantes encontrarían más fáciles las palabras del acento secundario porque el español tiene compuestos del acento secundario) resultó no ser precisa. Estos ítems no se planearon inicialmente para ser parte del análisis, pero como el investigador notó algunas producciones inesperadas de estos ítems durante los tests, se hicieron parte del análisis en la medida que el diseño del estudio lo permitiera.

Un hallazgo que fue algo decepcionante, además de inesperado, fue que una gran cantidad de participantes no pudieron decir con precisión las reglas que les habían sido presentadas durante el curso. Aunque no se les dijo a los participantes que tendrían que producir las reglas después del curso, se esperaba que pudieran aprenderlas. Lo interesante, y digno de una mayor investigación, fue que a pesar de que los participantes no aprendieron las reglas, generalmente mejoran desde el test inicial hasta el test final. Entonces, parece que los resultados indican que simplemente al trabajar con el material, la comprensión y el control subconsciente del lenguaje de los participantes mejoraron, a pesar de que no les enseñaron las reglas.

Aunque no siempre resultaron significativos estadísticamente, los resultados de ambas investigaciones en general muestran que tanto Facebook como YouTube pueden ser herramientas efectivas para enseñar, al menos, algunas partes de la pronunciación del inglés. Sin embargo, se necesitarán más estudios para investigar si algún elemento de la pronunciación en inglés se puede enseñar de esta manera.

En cuanto a la evaluación de los cursos por parte de los participantes, de hecho, las respuestas fueron favorables porque la gran mayoría de los sujetos declaró que el curso les resultó útil. No obstante, una pequeña tendencia en ambos estudios apunta a que YouTube fue la red social preferida como herramienta de presentación. Además, hubo varios participantes que respondieron que preferirían hacer ejercicios de pronunciación en el aula en lugar de hacerlos en línea. La razón de esto fue generalmente que sintieron que la presencia de un tutor les ayudó. Estos comentarios apuntan a una debilidad potencial en el tipo de curso utilizado aquí. Sin embargo, con el rápido desarrollo de los sistemas de reconocimiento del habla, existe la esperanza de que los alumnos que trabajan en línea dispongan de comentarios en un futuro no muy lejano.

Basándonos en los resultados de los tests de percepción y producción, así como del cuestionario después del test final, esta tesis concluye que las redes sociales podrían ser una herramienta viable para la enseñanza de la pronunciación del inglés. Sin embargo, todavía es necesario realizar más investigaciones para arrojar luz sobre varias áreas de este campo. Por ejemplo, la demografía de los estudiantes involucrados podría resultar muy importante. Debido a la disponibilidad de los participantes, el estudio no pudo determinar si los estudiantes fuera del grupo de edad de 18 a 25 años tienen una experiencia diferente con las plataformas utilizadas en esta tesis.

Chapter 1 – Learning and Teaching Pronunciation

It is a truth, universally acknowledged, that a language learner in possession of a good vocabulary, must be in want of good pronunciation. Although this adaptation of the famous Jane Austen quote may be rather crude, it nonetheless describes the experience of millions of language learners across the globe. It is generally accepted that pronunciation is the most difficult part of a foreign language to get right. There are multiple reasons why this is the case. Learning second language (L2) pronunciation requires both physical motor skills and perceptual practice (Saito et al., 2020). In addition to this, a range of psychological factors may influence how successful learners are in their attempt to master this aspect of L2 learning (Pennington and Richards, 1986).

Not only is L2 pronunciation difficult to learn, but it also has a reputation for being difficult to teach. As with learning L2 pronunciation, teaching involves a range of facets that must be right in order for the instruction to be effective. These include quality materials, subject knowledge, and, of course, time. The latter of these, in particular, is an issue as the first two become irrelevant if classroom time is taken up by other things.

In recent years, a myriad of technological tools has been suggested as a solution to the problems in learning and teaching pronunciation. Technology offers many advantages that traditional learning and teaching often lack. One of the main advantages described early in the literature is that technology is 'patient' (Pennington, 1999). What is meant by this is that the learner can make the same mistake countless times and still receive the same answer, whereas a human teacher might get annoyed with the learner.

Technology also offers new types of feedback (see 1.5.1). This includes visual feedback that allows learners to see their voice represented in various ways as well as audio feedback which may allow the learner to hear their own voice or a native-speaker model played back to them. They can also have automatic speech recognition (ASR) judge their pronunciation accuracy (i.e., the extent to which their production resembles an NS model). This can give them an idea of how close their pronunciation is to that of native speakers of English. Moreover, technology can be used to help learners improve their perception through high variability phonetic training (HVPT – see 1.5.2). Furthermore, with technology, learners can now work in a space and at a pace they find ideal. With the advances in computer technology in recent years, learners can work from the comfort of their own home if they so prefer. Finally, an obvious further advantage that has come about with improvements in technology is that teachers no longer have to be present to deliver instructions to students, as the instructions can be accessed online – either as real-time lectures or as recordings (e.g., Estebas-Vilaplana, 2015). Teachers can also post pronunciation instructions that students can access on their smartphones using applications such as Twitter (Mompeán and Fouz-González, 2016). This means that learners can access pronunciation materials anytime they want due to the vast amounts of content that can be found on the Internet. A very recent tool for language learning – and indeed pronunciation learning and teaching – is social media. Results so far have been encouraging, but there are still many areas of this technology that need exploring.

This thesis presents two studies on English pronunciation teaching using the social media platforms Facebook (Study 1) and YouTube (Study 2). The structure of the thesis is as follows. The remainder of Chapter 1 first outlines the history of pronunciation teaching. Then it looks at some of the issues and debates in learning and teaching pronunciation. Finally, it takes a closer look at the affordances provided by various types of technology from the past to the present day. Chapter 2 provides an overview of some key elements of English and Spanish word and sentence stress patterns. It presents two models of L2 stress perception, the Stress Deafness Model (SDM) and the Stress Typology Model (STM), the latter of which also discusses L2 stress production. Chapter 3 presents the main objectives and rationale for the two studies and also provides the research questions used to guide the studies. Chapter 4 presents the methods used for the execution of the two studies. Chapter 5 gives an overview of the results, which will, in turn, be discussed in Chapter 6. Finally, Chapter 7 presents some conclusions that can be drawn from the two studies, while acknowledging limitations observed in the studies; implications and recommendations for future research are also provided.

1.1. History of Pronunciation Teaching

Second language (L2) pronunciation is an area of second language acquisition (SLA) that is notoriously difficult to learn and has a reputation for being difficult to teach (Fraser, 2001). This is a problem worthy of attention considering the important role pronunciation plays in communication. Essentially, if a speaker's pronunciation is poor, communication can become impossible. For this reason, it would make sense if pronunciation was placed high on the list of priorities of any language teacher, but this is far from the case. In fact, pronunciation has often been referred to as the neglected skill (Celce-Murcia, Brinton and Goodwin, 2010; Elliott, 1995) or the Cinderella (Kelly, 1969) of language teaching. The importance assigned to pronunciation in foreign language teaching has varied considerably throughout history. In the following section this varied importance will be explained first through a historical view of pronunciation teaching, and then by looking at some practical challenges that teachers face in teaching pronunciation.

Exactly how pronunciation was regarded before the 1800s is difficult to say, due to the scarcity of sources to help us fully understand the views and trends of this time (Kelly, 1969; Murphy and Baker, 2015), but materials for English pronunciation instruction can be found as early as 1617 (Munro and Derwing, 2011). It would seem that while scholars may have been advocating the inclusion of speaking practice for centuries, teaching in classrooms focused exclusively on the written word up until the mid-19th century (Smith, 1893). The aim of teaching a foreign language was to provide students with sufficient knowledge about the structure of the language (i.e., its grammar) and its lexicon to produce translations between their L1 and L2 – hence the term 'Grammar-translation Method' (Celce-Murcia et al., 2010). As a consequence, speaking and pronunciation did not receive much attention, if any.

A change in focus began – as the world became more globalised – around the mid to late 19th century. While the Grammar-translation Method may have worked well for students who needed to learn how to produce written texts, in a more globalised world, the focus became increasingly centred around learning how to *speak* a foreign language. This is clearly illustrated by Krause (1916), who writes:

> [t]he man who is able to read a page of Taine, or perhaps of Anastole French, and who finds himself in a French business-house or a French drawing-room without the ability to express his wants or his thoughts in a single well-formed and intelligible sentence, feels like a fool, and he deserves to feel like a fool. (p. 102)

The method chosen to achieve this aim became known as the 'Direct Method'. The method was based on the assumption that a second language can be learned the same way as a first language. For example, Franke (1890) argued that children learn their first language without the use of dictionaries, grammars or producing translations, and yet learn to produce idiomatic language with ease. By way of extension, this was thought to be applicable to L2 learners as well. Thus, the learning style in the Direct Method was very much one that emphasised imitation and repetition with strict attention to pronunciation (Krause, 1916). The Direct Method was adopted by Berlitz, whose schools are still around to this day. It has been suggested, though, that only highly motivated students tend to succeed using this method (Asher, 1969). In addition, Krause (1916) warned that the method could be too heavily reliant on having an animated teacher who is able to engage the learners.

In the late 19th century, the Reform Movement, spearheaded by Henry Sweet, Paul Passy, and Wilhelm Viëtor, shared the disdain for the shortcomings of the Grammar-translation Method. One thing that set their theories apart from the Direct Method was that they refuted the idea that first and second language acquisition were identical in nature (Sweet, 1899). Their focus was the spoken language, and in their view, the means to be able to speak a foreign language was a detailed language analysis. Hence, one of the most important contributions of this movement was the development of the International Phonetic Alphabet (IPA). It was suggested that by using the IPA, teachers would be better able to describe and teach new sounds in the foreign language (Sweet, 1899; Jespersen, 1904).

Made possible by new advances in technology such as the tape recorder, the 1940s saw the rise of the Audio-Lingual Method (ALM). This method was based on the framework and approaches laid down by the scholars and practitioners of the Direct Method and the Reform Movement. Additionally, it incorporated ideas from structural linguistics as well as behavioural psychology. The ALM would gain in prominence for another two decades, and elements of its teaching practices can still be found in modern-day classrooms. In its essence, it is very similar to its forerunners with its imitative style of teaching. However, the Audio-Lingual Method had some new developments to offer as well. One of the new techniques that were introduced was the minimal pair drill (Celce-Murcia et al., 2010). This type of exercise was initially restricted to focusing on segmental differences only; that is, differences that involve one speech sound (e.g., *beat-bit*). This difference should ideally be a contrast that is not used in the L2. Later, the minimal pair drills were made to include suprasegmental features as well (Bowen, 1972, 1975). This means that the drills were used to practise features that involve more than an individual speech sound. These are features such as sentence stress, rhythm, and intonation.

In the 1950s and 1960s, when the popularity of the Audio-Lingual Method was at its peak, pronunciation was seen as a key component of language learning (Murphy and Baker, 2015), and it was widely assumed that listening to, and repeating after, a native speaker (NS) would make L2 learners' pronunciation native-like (Derwing and Munro, 2015). However, the results of teaching through the ALM were generally poor, and ALM is perhaps one the most heavily criticised teaching methods to date. The view of the ALM started to change when the theoretical underpinnings of the method were brought into question by Noam Chomsky's criticism of Skinner's 1957 book 'Verbal Behaviour'. Skinner's book described how language learning is a product of habit formation as was believed to be the case by behaviourist psychologists. Chomsky, however, argued that behaviourist theory could not account for the development observed in first language learners such as the fast vocabulary acquisition (Chomsky, 1959). Vocabulary is rarely taught explicitly to children, but even so, they manage to acquire a large number of words in their first few years of life. Furthermore, children produce sentences that they have not been exposed to; in Chomskyan terms, there is a 'poverty of stimulus'

(Chomsky, 1980) in the child's environment, which would lead to reduced linguistic ability in the child if behaviourist accounts were to be upheld. It should be added, though, that despite the ALM being heavily criticised and becoming less popular, the listen-and-repeat practice is still used to this day in, for example, the Pimsleur method.

Out of Chomsky's linguistic theories grew the cognitive approach, which treated pronunciation with very little regard, as ultimate attainment in a second language was viewed as impossible (Celce-Murcia et al., 2010). Although Chomsky's generative framework did not directly lead to the development of new teaching methods (Derwing and Munro, 2015), it did contribute significantly to the decline of the influence and popularity of behaviourist psychology in language teaching. It can be argued that this decline led to a vacuum in the area of teaching methodology. This vacuum, in turn, helped generate many new ideas about how to teach a foreign/second language in the 1970s and 1980s such as Total Physical Response (Asher, 1969) and The Silent Way (Gattegno, 1972), neither of which advocated direct pronunciation instruction.

Among the new pronunciation teaching methods of the 1970s and 1980s, The Silent Way deserves a special mention, though. In this method, the teacher speaks as little as possible and tries to illicit the correct words and phrases from the student through the use of a variety of tools such as coloured word- and sound charts as well as coloured rods, which are known as *Cuisenaire rods.* These tools allow teachers to show both sound and stress patterns in English without modelling these themselves. The absence of a language model is central to the theory behind The Silent Way, as it assumes that languages (first and second) are learnt, not through imitation, but through mirroring (Messum, 2007; Messum and Howard, 2015). This means that the learner produces some sort of output and obtains a response from the teacher. The response will then inform the learner about whether their initial output needs modifying. In some respects, The Silent Way was similar to the ALM in the sense that pronunciation was regarded as highly important. However, The Silent Way differed from the ALM because it also focused on suprasegmentals and banned the use of specific transcriptions systems and overt rules (Celce-Murcia et al., 2010).

Further contributing to the decline of both the ALM and the general neglect of pronunciation teaching was the fact that some scholars at the time (e.g., Scovel, 1969) claimed that pronunciation instruction did not make a difference past the puberty stage due to maturational constraints on the human brain (see 1.3.1). This idea was presented by Penfield and Roberts (1959) as the Critical Period Hypothesis (CPH) and later made famous by Lenneberg (1967). Although the idea of a CPH has been contested by several researchers (e.g., Bialystok, 1997; Bohn, 2005; Flege, 1987a; Flege, Munro, and MacKay, 1995; Singleton, 2005), it still has widespread support (DeKeyser, 2006: Long, 2005).

The 1970s and 80s also saw the emergence of Krashen's theories on second language acquisition. Krashen is probably most famous for introducing the distinction between *learning* and *acquisition*. He argued that SLA – similarly to first language acquisition – happens unconsciously, whereas second language learning is a conscious effort often involving explicit instruction (Krashen, 1982). An important thing to note about these two concepts is that what is learnt cannot be acquired. The ability to acquire language is maintained throughout life, and thus does not end at puberty as argued by the proponents of the Critical Period Hypothesis (Krashen, 1982). Unlike Chomsky, Krashen used his hypotheses of language learning to create a foundation for teaching practice which he – in collaboration with Tracy Terrell – named 'The Natural Approach' (Krashen and Terrell, 1983). In The Natural Approach, pronunciation is not seen as something that should be taught. There are two reasons for this: firstly, pronunciation is believed to improve as the student is exposed to the right input; and secondly, native speakers do not expect non-native speakers (NNSs) to be accent-free, so ultimate attainment should not be a goal for the learner (Krashen and Terrell, 1983).

As a final note on Krashen, it should be mentioned that although widely criticised for being unscientific, his hypotheses and teachings had a significant impact on how his contemporaries viewed pronunciation teaching (Levis and Sonsaat, 2017; Lightbown and Spada, 2006), and he has remained an authority in the field of SLA to this day.

In today's English as a Foreign Language (EFL)/English as a Second Language (ESL) classrooms, English is overwhelmingly taught through what is known as Communicative Language Teaching (CLT) or the Communicative Approach. As the terms imply, the focus is on language for communication. The foundation of what became CLT can be found in the discussion of Chomsky's notion of 'linguistic competence' and Hymes' 'communicative competence' (Interestingly, although this discussion became important in the area of SLA teaching, both Chomsky and Hymes were talking about L1 acquisition). Hymes (1972) argued that in focusing on *linguistic* competence, Chomsky failed to incorporate the social aspects of language learning. What is more, by assigning more importance to 'competence' than to 'performance' (Chomsky, 1968), he failed to incorporate the communicative goals of everyday language use.

CLT was (and to a large degree still is) critical of approaches such as the Grammar-translation Method for not focusing on communication. Similarly, the Chomskyan notion of linguistic competence was not endorsed as it excludes the focus on learners' proficiency in using language in real contexts for communicative purposes. Furthermore, at this time, teachers and researchers alike had started to realise that the Audio-Lingual Method was obsolete (Richards and Rodgers, 2001).

Communicative Language Teaching evolved during the 1970s and 80s to become a widely accepted approach to teaching (Arnold and Brown, 1999; Howatt and Widdowson, 2004). Canale and Swain (1980), building on Hymes' ideas of 'communicative competence', played an important part in shaping the approach. In particular, their development of this concept in into distinct components (i.e. Grammatical Competence, Sociolinguistic Competence, Discourse Competence, and Strategic Competence) has inspired research in second language pedagogy to this day.

Rather than being a move away from the structural approach of ALM, CLT tries to combine structural elements with functional ones to give students the tools to both analyse and use the language (Littlewood, 1981). Unfortunately, this approach to teaching English, with its focus on language as a means for communication, pushed pronunciation to the background (Benrabah, 1997; Derwing and Munro, 2005; Pennington and Richards, 1986) with some authors viewing pronunciation as an 'extraneous' feature when it comes to language teaching (Brown and Yule, 1983, p. 53).

A further factor contributing to the neglect of pronunciation teaching in this era was the fact that very few people knew how to teach pronunciation in a communicative context (Murphy and Baker, 2015; Pennington and Richards, 1986). For instance, in the ALM, pronunciation was taught through drilling speech sounds and speech patterns, but as this method had been heavily discredited, drilling was no longer seen as something to be used liberally or at all (Celce-Murcia et al., 2010; Finocchiaro and Brumfit, 1983). This left language teachers at a loss with regard to how to teach pronunciation.

Since the early 1990s, an ever-increasing number of studies has shown that pronunciation instruction can indeed be effective (Saito and Lyster 2012; Saito, 2012). Furthermore, it has been argued that pronunciation is important for ESL learners' self-confidence and that it plays a vital role in the learners' language use outside the classroom (Morley, 1991; Setter and Jenkins, 2005). What is more, some studies indicate that students themselves often request pronunciation instruction (Derwing and Rossiter, 2002; Levis and Grant, 2003). Regrettably, although there has been a significant increase in the number of studies focusing on pronunciation, teachers are still calling out for materials and training to help them teach pronunciation in their classrooms (Henderson et al., 2012).

In summary, as can be seen from the discussion above, the neglect of pronunciation teaching can be explained through a variety of contributing factors. Some of these are theoretical and have to do with pronunciation not being promoted because the skill of speaking was not part of the curriculum (Grandgent, 1892) or because it was believed that teaching pronunciation was not possible (Lenneberg, 1967; Purcell and Suter, 1980). Yet others believed that pronunciation would develop as the learner was exposed to the proper input (Krashen, 1982). Furthermore, the failure of the ALM to deliver the results could be taken as evidence that even with constant drilling, native-like pronunciation is unachievable. Finally, while the communicative classroom of the 1980s and 1990s did focus on the skill of speaking, specific pronunciation was not promoted.

Although many of the factors leading to the neglect of pronunciation teaching and research occurred either simultaneously or in rapid succession to produce what Derwing and Munro (2015, p. 48) have labelled 'a perfect storm', the tide has now shifted, and the number of publications in the area of pronunciation research is increasingly rapidly (Munro and Derwing, 2015; Saito and Plonsky, 2019). Moreover, academic conferences focusing specifically on various aspects of pronunciation have appeared in the past couple of decades (e.g., PSLLT, EPIP, and ACCENTS). In addition to this, specialist journals have also been founded that focus on pronunciation learning and teaching (e.g., JSLP). All in all, the field of L2 pronunciation is stronger than it has been for many years – at least academically. The following section will look more closely at the importance of pronunciation as well as the theory and practice of pronunciation teaching.

1.2. Why Teach Pronunciation?

There is an inherent assumption in this question, namely that it is possible to teach pronunciation. For now, this assumption will be regarded as correct, but as described above, it is far from an undisputed one.

One of the most influential proponents of pronunciation teaching in English language teaching, Adrian Underhill, has listed several reasons why pronunciation should be taught (Underhill, 2010). One of his reasons relates to the fact that pronunciation is, in part, a physical skill. As with any other motor skill, pronunciation must be practised for the learner to master it. Although not all pronunciation issues are physical (see 1.3.2), there is no doubt that active practice can have a substantial impact. Furthermore, pronunciation is not simply about speech production. Pronunciation can help learners in other areas of language use such as perception-related issues. For example, even though the sound-spelling correspondence in English is limited, being able to pronounce words correctly can help the learner with many spelling issues such as the voicing distinction in *path-bath*. Underhill also states that pronunciation helps the learner when listening to the foreign language; or as he puts it 'the mouth teaches the ear'.

As will be shown later, the link between perception and production is not entirely clear with some studies suggesting that perception leads production and other studies suggesting the opposite. However, it seems likely that working actively on voicing contrasts, for example, could help the learner perceive these better. The last reason for teaching pronunciation that Underhill mentions is that it builds self-esteem. The amount of research directly focusing on self-esteem and language learning may be limited, but there is certainly evidence that psychological factors are important when it comes to L2 pronunciation. One only needs to look to Krashen's Affective Filter Hypothesis (1.3.2) to see that some learner confidence and self-esteem are likely to matter in language learning.

Before getting deeper into the topic of whether pronunciation should be taught, it might be useful to define a few terms which are often used in the discussion about pronunciation teaching. These are *comprehensibility*, *intelligibility*, and *accentedness* (Munro, 2008). If a speaker is considered comprehensible, listeners find that speaker easy to understand. However, this does not mean that actual comprehension has taken place (Derwing and Munro, 2005). Intelligibility, on the other hand, is a measure of how accurately listeners can identify the words in the speech of a talker. Finally, accentedness is the perceived degree of accent assigned to a talker by native listeners. Contrary to what some might expect, there is no clear relationship between accentedness, comprehensibility and intelligibility. This means that some non-native speakers can have a very strong foreign accent, but still be highly intelligible (Munro and Derwing, 2002).

It seems fair to assume that when people speak, they want to be understood. In order to be understood, a speaker's pronunciation has to reach a certain level of intelligibility. If a speaker fails to reach this threshold, communication breakdowns are likely to occur (Morley, 1994). Communication breakdowns of this type are likely to be detrimental to the learner's self-esteem, and, if experienced frequently, may put the non-native speaker off engaging in conversation altogether.

When native speakers hear a foreign accent, they are often able to tell that it is indeed a foreign accent. In fact, studies have shown that a foreign accent can be identified even in very short speech samples (Flege, 1984) and in even speech played backwards (Munro et al., 2003). It might be more difficult for them to say what it is about the accent that makes it foreign. Even so, a foreign accent can have detrimental effects on L2 speakers in a variety of ways.

Speaking with a foreign accent carries a stigma in many contexts (Henderson, 2004), and it is probably not a coincidence, as pointed out by Max von Sydow, that in films, people 'with an accent are bad guys' (quoted in Derwing and Munro, 2015, p. 131). This stigmatisation is very real and can impact learners negatively, so it is not something language teachers should ignore. As a case in point, Derwing (2003) found that a third of her participants believed they had been discriminated against based on their accent. Similarly, Henderson (2004), states:

> [e]xclusion from the workforce (ranging from professional gatekeeping via unrealistic accreditation requirements to more blatant racial prejudice) as a result of discrimination against a marked accent which signals immutable characteristics (that is, country or area of origin, ethnicity and visibility) has been identified in New Zealand research. (p. 16)

The reason behind this exclusion and stigmatisation can in part be explained by the traits some listeners ascribe to accents. For example, Flege (1988) found that listeners may make negative judgements about personal qualities based on accent. However, Flege also points out that the process by which native speakers assign qualities to foreign accents depends on a variety of factors. It has been suggested that the process by which accents are stigmatised results from listeners placing L2 speakers with a foreign accent in a less prestigious socio-economic category, which places a stigma on the L2 speaker. However, what is particularly important – and encouraging – in this context is that the process seems to be slightly more complicated than that, as the degree of foreign accent seems to play a part as well. This should give some comfort to non-native speakers, and most certainly underlines the importance of teaching pronunciation.

A further aspect that research has shown that can have negative consequences for the learner is that listeners also tend to draw conclusions about a speaker's linguistic ability based on the person's social identity. This can cause a decrease in a listener's ability to understand a speaker if the listener is told that the speaker is from a foreign country, even if the speaker is actually a native speaker (Kang and Rubin, 2009). Although there is little teachers can do to help in this respect, a first step in the right direction is to help learners reach their full potential linguistically. For the reasons outlined above, it should be clear that pronunciation is indeed a worthwhile focus in language learning and teaching – provided that it can be taught effectively. This will be the topic of the following pages.

As can be seen from the paragraphs above, L2 pronunciation learning is the most difficult aspect of learning a new language. Furthermore, not being able to master pronunciation can have some very negative consequences. Therefore, we need to look for effective ways to teach pronunciation. As stated by Chun (2002):

> it is precisely because of the fact that pronunciation is one of the most difficult facets of language in which to achieve native or near-native competence that we must seek ways to help learners overcome some of the most challenging aspects of pronunciation. (p. 89)

While effective ways of teaching pronunciation seem necessary in contemporary L2 instruction, the assumption is that pronunciation can, indeed, be taught – a point to which we now turn.

1.3. Can Pronunciation be Taught?

There is certainly reason to believe that pronunciation can indeed be taught. Whether late learners can achieve native-speaker competency is another matter. The fact that late learners rarely learn to speak their L2 as NSs has been one of the most frequently used arguments for the proponents of the Critical Period Hypothesis. The CPH will be addressed first in this section. The subsequent to subsections look at factors involved in learning the pronunciation of an L2 and the effects of instruction, respectively.

1.3.1. Foreign Accent and the Critical Period Hypothesis

It is commonly acknowledged that everybody has an accent. For native speakers of English, the accent can be influenced by a variety of socio-linguistic factors such as geography, social class, age, and ethnicity (Trudgill, 2001). Unless a child is suffering from a speech impairment of one form or another, they will learn to produce their L1 with native-speaker competence. The same thing is true for immigrant children who arrive in their new country at a very young age and are exposed sufficiently to the new language (Flege, 2009). However, the picture changes significantly when looking at late learners (i.e., learners who did not learn the language from a young age). These are the language learners who receive the first exposure to the L2 after the critical or sensitive period (see below). These learners will often experience that their language competence fossilises at a certain level well below that of a native speaker (Selinker,

1972).

Several factors have been suggested to influence the degree of foreign accent. Among these variables are age of learning, amount of L1 use, formal instruction, language learning aptitude, and motivation (Piske, MacKay, and Flege, 2001). These authors suggest that age of learning is the strongest predictor of foreign accent in an L2. However, it is important to note that the decline in L2 pronunciation ability shows a steady decline with age rather than an abrupt drop after a certain age. Piske et al. (2001) also reported on a study of 72 Italian immigrants to Canada and how their pronunciation varied depending on L1 use and age of arrival in Canada. Furthermore, upon re-examining their data, Flege, MacKay, and Piske (2002) found that some early bilinguals (age of arrival = 7-8 years) who primarily used English, produced English without a detectable accent, leading the researchers to suggest that it is possible – in some instances – for a non-native speaker to reach a native-like performance level.

An opposing explanation for the foreign accent of late L2 learners can be found in the Critical Period Hypothesis. As mentioned in 1.1, the foundation for the Critical Period Hypothesis was laid by the work of Penfield and Roberts (1959). Their research based on cortical stimulation during brain surgery led Lenneberg (1967) to propose the hypothesis. Although Lenneberg (1967) did not include lexical learning as an area affected by a critical period, the CPH has been claimed to pertain to all aspects of language learning (Long, 1990), but here, only pronunciation will be considered.

Although Penfield and Roberts (1959) concluded that the left hemisphere of the brain plays a major role in speech production, it was Scovel (1969) who proposed that lateralisation was the mechanism responsible for the decline in oral proficiency among late learners. Hence, the problems L2 learners face were believed to be intrinsic and not something external factors could influence. Scovel (1969) states that:

> [t]he fact that children can acquire language with native-speaker fluency and adults cannot, and the fact that the production and recognition of foreign accents is a trait and not a skill, forces us to conclude that it is *nature* and *not nurture* which determines our ability to speak without a foreign accent. (p. 249)

The CPH has probably generated more research than any other hypothesis in SLA. While a large number of studies have claimed to find evidence in support of the hypothesis, an equally large number of studies have challenged it (Bohn, 2005). Among the criticisms is the fact that the cut-offs for the onset and termination of the critical period are not agreed upon by the proponents of the CPH. Some argue that the onset is at the age of two, even though babies are already going through a significant linguistic development at that age. As for the termination of the critical period, some claim that ultimate attainment is impossible after the age of two, whereas others set the termination of the critical period much later. Furthermore, work by Flege and colleagues leading to the development of the Speech Learning Model (SLM, Flege, 1995) has suggested that even adult learners are able to change their phonetic inventory to some degree. Although this work is not proof of ultimate attainment, it does indicate that the concept of maturational constraints on speech learning should be abandoned.

One of the problems scholars seem to face when debating the CPH is that they do not seem to agree on what level is sufficient for the subject to pass as a native-like speaker in a study. This has led proponents of the CPH to question the methods used in some of the studies claiming to show ultimate attainment. One issue often brought forth is that the participants in these studies are often tested on read-out-loud tasks, which are less demanding than spontaneous speech. For example, in a study of late learners' acquisition of Dutch, Bongaerts, Mennen, and van der Silk (2000) claimed to have identified two of their 10 participants that native listeners rated as high or higher than native speakers. However, Abrahamsson and Hyltenstam (2009) disregarded this finding claiming that the participants would not have passed as native speakers had they been tested thoroughly enough. This conclusion was reached in part based on their own study (ibid), which initially found that some late learners received ratings as high as those of native speakers. However, when presented with a more cognitively demanding speech production task, all of these participants received lower ratings than those of native speakers. The question remains how proficient a non-native speaker has to be before they can be considered to sound native-like, and the answer could be that it depends on who is listening.

As can be seen from the discussion above, much of the evidence either supporting or challenging the CPH has been indirect to some degree. This is, to a large extent, because the claims regarding cerebral plasticity have been impossible to measure directly due to insufficient technological means. However, in recent years, advances in technology have allowed researchers to observe changes in the human brain that occur with age. Research in the field has shown that the notion of reduced neural plasticity cannot be upheld (DeLuca et al., 2019). As a consequence, an explanation for L2 foreign accents must be found elsewhere.

To sum up, there is most certainly a case to be made that late learners generally do worse than early ones, when it comes to acquiring native-like L2 pronunciation. However, there is little evidence to support a strong version of the CPH suggesting that this is due to cerebral maturational constraints. Instead, a range of other factors are needed to explain why late learners find pronunciation difficult. In 1.3.2, some of these factors will be discussed.

1.3.2. Factors Involved in L2 Pronunciation Learning

As most learners of a second language will know, pronunciation is usually the most challenging element of L2 learning, with very few learners ever reaching truly native-level proficiency (Flege, Munro, and MacKay, 1995; Long, 2005). Thus, a very basic yet important question is why second language pronunciation is so difficult for learners to master. This subsection takes a closer look at some of the important facets involved in this area of SLA.

As can be seen from the quotation below, pronunciation stands out compared to other aspects of language learning:

> [p]ronunciation is the only part of language which is directly physical and which demands neuromuscular programming. Only pronunciation requires an incredible talent for sensory feedback of where the articulators are and what they are doing. And only pronunciation forces us to time and sequence motor movements. All other aspects of language are entirely cognitive or perceptual in that they have no physical reality. (Scovel, 1988, p. 62)

Despite placing pronunciation in a category of its own, the quote nonetheless fails to include many of the factors that influence pronunciation learning in a second language, including age, quantity and quality of input, perceptual factors, motivation, and other personal factors.

As described above, age is perhaps the most important factor when it comes to predicting degree of foreign accent. As the debate regarding the influence of age was discussed in detail above, the only thing that should be added here is the recommendations made from past years' research in the field. Although there is little evidence in favour of a strong interpretation of the CPH, there can be little doubt that learning an additional language at an early age may provide the learner with a distinct advantage when it comes to pronunciation.

It is uncontroversial that experience is required in order to learn a language. When investigating English as a second language, experience has been measured as 'length of residence' in a new country as done by Flege and his colleagues. However, it has been found that this might not be the most appropriate way of measuring this variable, as living in a new country does not necessarily equal exposure to the new language (Flege, 2009). The situation is further complicated by the fact that the age of arrival (AOA) may influence the effect of exposure over time. For example, a 2-year-old and a 20-year-old will have very different linguistic experiences over the course of ten years. Hence, it might not be the AOA, per se, that is important, but the learner's linguistic experience and context.

In a Spanish context, English as a foreign language experience can in most cases be equated with years of formal learning. However, it is important to add that this does not simply mean teaching the additional language a few hours a week in early childhood as L1 use plays a significant part in determining degree of foreign accent as shown by Flege, MacKay, and Piske (2002). In fact, it is interesting to note that age of learning may not play as significant a part in second-language learning depending on the circumstances. This idea was proposed by García Lecumberri and Gallardo (2003), who investigated the effects of formal education in relation to age among three groups of L1 Spanish learners. The study found that the learners who started learning English later were judged to be more intelligible and have less of a foreign accent. In addition, the older learners performed better on both vowel and consonant discrimination. The authors concede that their findings may be influenced by the less advanced cognitive development of the youngest group of learners (age 9-11). They, nonetheless, maintain that age of learning is not an important factor in an FL context. It thus seems as though more research is needed to elucidate the effects of age in an FL context.

One thing that most authors agree upon is that it is crucial that the input learners receive is adequate in terms of both quantity and quality (Flege, 2009). No one has argued more strongly in favour of the importance of input than Krashen. As discussed in 1.1, his theories on language learning played a big part in the demise of pronunciation teaching. Krashen's input hypothesis was part of his Monitor Model, which consisted of four hypotheses: the *acquisition-learning hypothesis*; the *monitor hypothesis*; the *natural order hypothesis*; the *affective filter hypothesis*; and finally, the *input hypothesis*. The latter of these hypotheses argues that a second language is learnt when the learner is exposed to input that is at a level of *i* + 1 compared to the learner's level – Krashen uses the term 'stage' instead of level in his original work (Krashen, 1982). In other words, the required input that the learner is exposed to should be only slightly higher than their existing level. However, Krashen notes that it should not be an aim in and of itself for the teacher to provide this input as he states that input 'need not contain only *i + 1*. [...] if the acquirer understands the input, and there is enough of it, *i + 1* will automatically be provided' (Krashen, 2003, p. 21).

What Krashen did not assign much importance to was the effects of output i.e., the opportunities L2 learners have to actively practise using the language. As mentioned above, output practice should be considered an important part of language learning and teaching. Hence, Krashen's views were challenged by the Comprehensible Output Hypothesis (Swain, 1985). As Swain and Lapkin (1995) explain, when learners produce language, they are more likely to notice problems in their linguistic ability. Only through noticing these issues can learners push themselves to improve their productive capabilities (Schmidt, 1990). In addition to the concept of noticing, the comprehensible output hypothesis also claims that producing speech helps learners test their hypotheses about language. This is done through the feedback they get from their interlocutors. If their speech is not comprehensible, they will need to adjust their language. Finally, the hypothesis claims that language use enables the learner to create meta-linguistic knowledge of learner language. This is done when learners reflect on the language they have used and internalise what they have learnt.

A great deal of research has looked at perceptual factors to explain issues on L2 pronunciation. From a perceptual point of view, L1 phonological categories of speech sounds are established very early on in the learner's life (Kuhl et al., 1992; Werker, 1989). Once these categories are in place, they can be said to function as a phonological sieve (Trubetzkoy, 1971) through which L2 input is interpreted. In a similar vein, the existence of a 'perceptual magnetic effect' has been suggested (Kuhl and Iverson, 1995). In this analogy – the Native Language Magnet (NLM) model – L1 prototypes are said to warp the perceptual space in listeners and draw L2 input towards these prototypes (Kuhl, 1992, 1994). However, Frieda et al. (1999) show that this analogy does not provide the best model for L2 speech perception as listeners seem to perceive L2 tokens slightly differently from what the model predicts. This said, the model proposed by Flege and colleagues, the SLM, makes very similar predictions to the NLM in that it posits that L2 speech perception is indeed affected by L1 phonetic categories (Flege, Frieda, and Nozawa, 1997). However, it should be noted that reliance on L1 categories does not tell the whole story, as L2 learners may at times use vowel duration rather than spectral differences to distinguish between non-native vowel categories (Bohn, 1995).

Closely related to this line of thinking is the cognitive approach to pronunciation learning (Fraser, 2001). According to the cognitive approach, one of the reasons that L2 learners struggle with pronunciation is that they lack the right phonological concepts, and thus have to rely on the phonological concepts on their L1 for speech perception and production. According to Fraser (2001):

> [b]y far the majority of pronunciation problems stem not from physical, articulatory causes, but from *cognitive* causes. In other words, the problem is not that the person can't physically make the individual sounds, but that they don't *conceptualise* the sounds appropriately – discriminate them, organise them in their minds, and manipulate them as required for the sound system of English. (p. 20. Italics originally in bold)

It must be pointed out that concepts in this approach are different from the perceptual categories described above. In this cognitive view, perception is important, but perception is only seen as a process whereby the learner receives some degree of information about the surrounding world, and thus represents a vague awareness. The cognitive concepts, on the other hand, represent knowledge.

Hence, one way to help learners improve their pronunciation is by helping them form the needed concepts. This is not a simple thing to do, and it requires both time and practice (Couper, 2011; Fraser, 2006a). For this reason, meta-linguistic knowledge is considered a powerful tool in this approach as it is through the meta-linguistic knowledge, that the learner can work on concept formation.

Although the cognitive approach argues that the vast majority of pronunciation issues in L2 are due to cognitive rather than motor factors, there can be little doubt that motor issues can cause considerable problems for some language learners. As stated in the quote above, these are factors that are unique to this aspect of language learning. Pennington and Richards (1986) describe the development of second language phonology as a ‘dynamic process involving cognitive, psychomotor, linguistic, and interactive factors’ (p. 214). As most people who have tried to learn Spanish as a foreign language know, if learners do not use the trilled /r/ in their L1, it is likely to be very difficult for them to produce this sound. However, in terms of teaching, it is a fairly simple task for the teacher to describe the sound and suggest exercises to train the muscles involved in its production.

The last group of factors to be discussed are psychological in nature. One such factor is foreign language anxiety (Horwitz, Horwitz, and Cope, 1986). Foreign language anxiety ‘is categorized as a situation-specific anxiety, similar in type to other familiar manifestations of anxiety such as stage fright or test anxiety’ (Horwitz, 2010, p. 155). As such, it is a condition that can make it extremely difficult for a learner to experience gains from classroom-based pronunciation teaching and has, in fact, been considered as one of the most reliable predictors of language learning failure (MacIntyre, 1999). To overcome the effects of foreign language anxiety, teachers will have to make sure that the learning environment is as safe and comfortable as possible. However, this might not be an easy task in a classroom full of students. Regrettably, this means that even the most well-designed materials can prove useless if the learner has a high level of foreign language anxiety (Arnold and Brown, 1999; Dewaele and MacIntyre, 2014).

While foreign language anxiety is typically viewed as a negative factor in L2 language learning, it can in some cases be experienced – perhaps surprisingly – as beneficial (Lightbown and Spada, 2006). However, this is mostly in relation to preparation for tests or presentations because a small amount of FLA can help increase motivation. In the classroom, FLA is mainly negative as it can distract learners by making them think more about how their speech is perceived than on how to produce it. Furthermore, it can keep learners from interacting altogether, meaning that they will produce no output and thus rob themselves of the opportunity to notice gaps in their linguistic capabilities.

Another psychological factor which is likely to affect pronunciation learning is *language ego*. The term was coined by Guiora (1972) drawing on the Freudian concept of *body ego*, thus describing the child's acknowledgement of its linguistic development. As the child grows, the language ego solidifies and becomes an obstacle to pronouncing words and sounds in a foreign language (Larsen-Freeman and Long, 2014). According to Guiora (1972), '[p]ronunciation is the most salient aspect of the language ego, the hardest to penetrate (to acquire in a new language), the most difficult to lose (in one's own)' (pp. 144-145). This aptly depicts the struggles both teachers and learners face when it comes to acquiring L2 phonology. Whereas psycho-motor problems can be overcome through instruction and exercise, issues relating to the leaner's personality and perception of self are much tougher hurdles. A consequence of the language ego could be that the learner feels awkward or self-conscious when practising pronunciation in a classroom. It has thus been suggested that the difference in flexibility of the language ego between children and adults explains why it is relatively easier for children to learn to produce the sounds of a foreign language (Guiora et al., 1972).

Related to the idea of language ego is the learner's perception of social affiliation. This refers to how the learners see themselves as part of a group, both socially and linguistically. This can be a very strong force indeed and can result in learners deliberately not producing new language to the best of their ability because sounding 'different' would make them feel less as part of the group (Laroy, 1995). This was also observed by Jenkins (2000), who reported that some of her learners directly stated that they felt embarrassed about speaking English to learners with whom they shared an L1. In some cases, linguistic conversion can also be used as a tool to solve communication issues. This exact scenario was described by Walker (2005), who gives the example of a minimal-pair drill in which learners produced the word *road* as /ro-at/ in order to distinguish between *road* and *wrote*.

Considering the large number of factors at play when learning to pronounce a foreign language, it is not surprising that pronunciation can be difficult to teach; nor is it surprising that many scholars have claimed it impossible to teach. The next section looks at the effects of pronunciation teaching.

1.3.3. Effects of Instruction

As mentioned in 1.1, it was believed for a long time that pronunciation could not be taught to adult learners. One explanation was that late learners would be unable to improve due to maturational constraints on the adult brain (Lenneberg, 1967). Another theory posited that pronunciation had to develop on its own through exposure to the right input (Krashen, 1982). Because Krashen distinguishes between learning and acquisition, he argues that pronunciation instruction only leads to learning, which is in essence declarative knowledge. As, in Krashen's view, declarative knowledge cannot be transformed into procedural knowledge, pronunciation instruction is pointless. Yet another argument was that, statistically, instruction did not show any effect (Purcell and Suter, 1980).

Purcell and Suter's study was particularly influential because rather than being theoretical in nature, it purported to use detailed statistical analyses to draw its conclusions at a time when empirical research on pronunciation was not as common as it is currently. The study was a re-examination of the data obtained by Suter's (1976) study on variables predicting foreign accent. It claimed to expand on its findings by applying different statistical methods to the ones used in the initial study. Of the 20 variables under investigation, only four were found to be reliable predictors of pronunciation accuracy, namely First Language, Aptitude for Oral Mimicry, Residency, and Strength of Concern for Pronunciation Accuracy. The variables 'Number of years of formal classroom training in English' and 'Number of months of intensive formal classroom training in English', and 'Number of weeks of formal classroom training focused specifically on English pronunciation' on the other hand, were not found to correlate with pronunciation accuracy. Although these findings had a significant influence on teachers and scholars at the time, and have often been used as evidence that second language pronunciation instruction is not worth focusing on in the classroom, some researchers have taken issue with the conclusions drawn by Purcell and Suter. Despite the researchers' focus on presenting a methodologically rigid study, there are a few flaws worth noting. For one, some of the variables created for the study are not particularly well defined, making it difficult to say exactly what is being measured (Derwing and Munro, 2015; Pennington, 1998). Derwing and Munro (2015) further point out that the study is correlational in nature and thus does not allow for firm conclusions regarding causality.

Proving the non-existence of something is not an easy task, and unsurprisingly the number of studies arguing in favour of an effect of pronunciation training vastly outnumber the studies that claim no such effect exists (Abe, 2011; Saito and Hanzawa, 2016). In the past two decades, a host of studies on the effects of instruction on various aspects of pronunciation have been carried out. Although not all of them have been successful, this is by no means evidence that pronunciation instruction is not effective. It merely shows that pronunciation can indeed be difficult to teach.

Indeed, the findings from the research that has been carried out to shed light on the effectiveness of pronunciation teaching are by no means monolithic in nature. For example, in an often-cited study, Macdonald, Yule and Powers (1994) found that their three experimental groups (drilling, self-study, negotiation of meaning) showed no statistically significant improvement when compared to a control group. It must be said, though, that the duration of the treatment was very limited, namely one session of 10-30 minutes (the self-study group was allowed 30 minutes as opposed to only 10 minutes for the other two groups), so it might not come as a great surprise that no improvements could be found.

A different and slightly more indirect approach was taken by Fullana and Mora (2009). These researchers compared three groups of learners who varied in terms of the age at which they started learning English as well as the amount of instruction they had received. Thus, in this study, there was no direct intervention. Instead, the three groups of learners were tested on their perception and production of voicing contrasts in English word-final obstruents. Their analysis found no indication that the amount of instruction the learners had received predicted their performance on the tests.

Because of the many components involved in producing spoken language, pronunciation can be approached in a variety of ways. It can be taught by developing with the students a metalanguage about pronunciation which allows the teacher to give clear instructions about how students can improve their pronunciation (Couper, 2011). Other than improving communication between the students and their teacher, the creation of a metalanguage also helps raise the students' awareness of the features in focus as explained above.

Another tool that teachers can use is corrective feedback. Li's (2010) meta-analysis of corrective feedback in SLA in general showed that corrective can be effective. Unfortunately, this study did not specifically identify pronunciation as an area of focus. The effectiveness of corrective feedback specifically in L2 pronunciation teaching was the focus of a study by Dlaska and Krekeler (2013). This study looked at both implicit feedback (students listened to their own recordings) and explicit corrective feedback (the tutor corrected the students' pronunciation). The authors concluded that the learners who received explicit corrective feedback improved significantly more than the learners who only received implicit feedback. Saito and Lyster (2012) explored the effects of corrective feedback in combination with form-focused instruction (FFI) as well as FFI on its own. As the name implies, FFI is a type of instruction where form is brought to the fore and meaning plays little or no role at all. As stated by Fraser (2001), there is little point in correcting a learner if the learner does not understand what is being corrected. The researchers found that FFI on its own was effective in improving their participants' pronunciation of /ɹ/. They further found that the effect was higher when combined with corrective feedback. It could be argued that not focusing on meaning robs the learners of an opportunity to learn, but research has shown that in some aspects of pronunciation learning, adding meaning as a focus can have a detrimental effect on learning (Trofimovich and Gatbonton, 2006).

Another tool which has been advocated for, for many years, is the use of basic phonetics and phonetic notation (e.g., Jespersen, 1904). Jespersen was very clear that phonetic terminology should be avoided, but that being able to talk about articulators, for example, would prove advantageous in pronunciation teaching. However, because the use of IPA symbols became regarded as tools focusing on rigid from rather than fluent communication, it fell out of use for several years.

Despite the scepticism some have shown towards this type of teaching tool, researchers are now starting to change this view. For example, Mompeán and Lintunen (2015) provided a reassessment of phonetic notation in pronunciation instruction. The authors state that there are several potential advantages of using this type of instruction. At a very basic level, it introduces a level of standardisation in terms of orthography as each symbol has a fairly clearly defined sound as opposed to conventions that try to spell out the sounds with ordinary letters. A further advantage to these systems is that phonetic notation is 'systematic and comprehensive enough to deal with both segmental and suprasegmental features' (Mompeán and Lintunen, ibid. p. 308). In addition to the advantages already mentioned, introducing phonemic script will help raise students' awareness of the differences between speech sounds that might assimilate to the same L1 category as is the case with /I – i/ in L1 Spanish speakers or /s – z/ in L1 Danish speakers. As for the place of phonetics in L2 pronunciation, knowing how a sound contrast is produced can be crucial for its correct realisation as in the aforementioned case of /s/ and /z/ as voicing contrasts are impossible for a teacher to show directly. Furthermore, being aware of the difference between two phones and being able to use phonetic script allows learners to look up lexical items in a pronunciation dictionary to confirm their pronunciation, thus providing the learner with better opportunities for autonomous learning.

Despite the clear potential of using phonetics instruction, only a small number of studies have tested its effect empirically. Focusing on L1 American English learners of Spanish, Kissling (2013) found an effect of specific phonetics instruction with regard to improving learners' production of a selection of Spanish stop consonants, approximants, and rhotics. The experimental participants in this study were compared to a group of learners working with similar materials in class, but who did not receive explicit phonetics instruction. The study found that although the experimental group improved, the improvements were no greater than those made by the comparison group. However, this should not necessarily be taken as an argument against including elements of phonetics in pronunciation teaching, as a focus on this area has several other advantages as pointed out above. As a final note on this topic, it should be added that in a later study, Kissling (2015) found that phonetics instruction was, in fact, effective in improving learners' perception of L2 speech sounds as well as production.

The studies mentioned so far have focused exclusively on the effects of phonetics instruction on segmental phonology. Speculating that suprasegmental features might develop faster than segmental ones, Gordon and Darcy (2016), took a broader approach to the study of the effects on instruction. Their study compared three groups of learners: a group receiving instruction on segmentals, a group receiving instruction on suprasegmentals, and a control group. More specifically, the group receiving instruction on suprasegmentals focused on word stress and sentence stress, linking, rhythm, and vowel reductions. The study found that focusing on these features helped the participants improve their intelligibility. These results are particularly encouraging because the participants only received four hours of instruction across three weeks. Thus, there can be little doubt that at least some understanding of basic phonetic concepts can be a great help in learning and teaching pronunciation. However, more research is still needed in order to elucidate which methods work best in various areas of pronunciation.

Phonetics teaching for L2 pronunciation training has also been done using computers in a language laboratory. As a case in point, AbuSeileek (2007) had his participants (N = 50) work with the software *Mouton Interactive Introduction to Phonetics and Phonology.* The study investigated whether working with exercises on stress placement in word and sentences would improve the participants' pronunciation. After 12 weeks, during which the participants had worked with the software for approximately one hour per week, this experimental group was compared to a control group consisting of students from the same programme, who had not used the software. The data indicated that the experimental group made significant improvements compared to the control group, although some aspects showed larger improvements than others.

Interestingly, focusing on production might not always be the best way to improve speaking accuracy. This was demonstrated by Counselman (2010), who compared two groups of native speakers of English learning Spanish. One group received classroom-based pronunciation instruction and were assigned tasks focusing on speech production, whereas the other group received the same pronunciation instruction but were assigned tasks focusing on perception. The results of the study showed that the group having worked with perception-based assignments outperformed the other group in the following production test.

It is important to keep in mind what is meant when talking about improvement. Although there are general trends in how pronunciation studies are carried out, the way participants' pronunciation is scored or judged may differ considerably from study to study. As such, Canadian researchers Derwing and Munro as well as some of their colleagues have looked extensively at the constructs of *intelligibility* and *comprehensibility* as the target for improvement (Derwing, Munro, and Thomson, 2007), whereas other studies have focused specifically on the *accuracy* in the production of a speech sound or intonation pattern. This accuracy may be based on native speaker judges, non-native expert judges (e.g., trained phoneticians), acoustic analysis, or any combination of these. The considerable variation in participant assessment could be taken to mean that if a study focusing on pronunciation instruction fails to report an improvement, it is possible that the treatment was not a waste of time, but that the subject may have improved in ways that were not measured.

It should also be mentioned that important differences are likely to exist between second language and foreign language learning contexts. In particular when it comes to pronunciation, it can be argued that second language learners have a distinct advantage over foreign language learners because they have large amounts of native speaker input available to them whereas NS input is limited in foreign language classrooms (Best and Tyler, 2007; Larson-Hall, 2008). While it is true that the Internet has made finding examples of native speaker input much easier, it still takes a conscious effort on the part of the FL learner, whereas most SL learners need but leave their house in order to be exposed to NS input. It is thus very likely that English instruction is indeed effective, but that foreign language learners suffer from a lack of opportunities to experience and practise the language due to the learning context in which they are. This situation is further aggravated by instruction often being limited to a few sessions of less than an hour each week (Muñoz, 2008). As a consequence, the effects of foreign language instruction may fail to materialise.

It is thus clear that although pronunciation teaching can be effective, there is no guarantee that it will be, as it seems to depend on both the amount of instruction and possibly the type of instruction. Although many factors play a part in determining how successful learners are in improving their pronunciation, there can be little doubt that any learner, given the right circumstances, will be able to improve at least to some extent. With this in mind, there is one question left to answer, namely what to teach when teaching pronunciation. This will be the focus of the following section.

1.4. Issues in Teaching English Pronunciation

By now, it should be clear that a range of factors are involved in teaching pronunciation. This means that teachers face a variety of challenges which can be both theoretical and practical in nature.

Challenges that can be considered somewhat theoretical include considerations about what to teach. Because time is of the essence when it comes to teaching, it is imperative that teachers know what to focus on in their teaching (Gordon and Darcy, 2016). Hence, this section looks at some of the issues and debates regarding the focus and aims of pronunciation teaching. More specifically, three debates will be reviewed, namely the nativeness vs. intelligibility debate, whether to teach for an English as a Foreign Language (EFL) or English as a Lingua France (ELF) context, and finally, whether teachers should focus on segmental or suprasegmental features. Following the discussion of these three topics, some more practical issues that teachers face will also be discussed.

1.4.1. Nativeness or Intelligibility

Some researchers have claimed that the only aim for second language pronunciation teaching should be increased comprehensibility and intelligibility (Munro and Derwing, 2015; Thomson and Derwing, 2014). First of all, these researchers point out that for the vast majority of learners, sounding like a native speaker is an unattainable goal. Another reason for their scepticism of pronunciation teaching focusing on nativeness can be found in their opposition to the number of unqualified pronunciation coaches in the language learning industry. Many of these charge large fees while providing pronunciation instruction (accent reduction) that may do more harm than good while claiming to improve learners' accents and making them sound more native-like (Derwing and Munro, 2009). Finally, their views also stem partly from research on the relationship between intelligibility and accentedness, where results have shown that L2 speakers can be very heavily accented without this compromising their intelligibility (Munro and Derwing, 1995).

While it is certainly true that even heavily accented speech can be intelligible, arguing that accent reduction measures are always a waste of time and money might be a step too far. The reason for this is that the dismissal of the importance of accent seems to miss the reality that NNSs face. The goal of a teacher might be to make their students able to speak intelligible English, but learners often want either more than that, or simply something different than that, namely to approach a native-like model (Dimitrova and Chernogorova, 2012). For some, particularly in a foreign language context, this may only be due to a personal idea about what is 'good' and what is 'bad' English – however misconstrued this notion is – but in a second language context, sounding like a native speaker can be a matter of feeling included in the group or community (Gluszek and Dovidio, 2010).

One argument in favour of adding accent reduction measures to EFL teaching also relates to the effect of linguistic stigmatisation discussed earlier in Section 1.2. For example, research suggests that a foreign accent can affect perceived grammaticality. In other words, NS listeners may judge foreign-accented speech as less grammatical than it actually is (Kennedy, 2015). Although further research is needed to firmly establish the conclusions drawn in this study, the implications for how English is taught in both NS and NNS communities should be clear.

The discussion above may suggest that whereas the arguments in favour of focusing on intelligibility to a large degree are based on empirical research, the arguments in favour of focusing on nativeness are mainly based on opinion. Hence, there can be little doubt that, as a general rule, the focus of pronunciation teaching should be on intelligibility. However, this is not to say that teachers should refuse to help learners with pronunciation issues they rate as important. In particular, students who are learning English in order to find work in an English-speaking country may want to get as close to native-speaker pronunciation patterns as possible. At the same time, it is important that learners are made aware of how much work may be needed in order to approach native-like competence in L2 pronunciation.

1.4.2. EFL or ELF

The debate over whether to teach English as a Foreign Language or English as a Lingual Franca is, to some extent, political in the sense that the main models of English used in EFL are not based on ease of learning, but used for cultural and historical reasons. This debate is beyond the scope of this dissertation, so, instead, this subsection looks at ELF from a phonological perspective.

For many years, foreign languages were taught with the aim of equipping the learner with the required tools to engage in conversation with native speakers. With the increased realisation that most learners will need English to talk to other non-native speakers, Jenkins (2000) suggested a change in focus from English as a Foreign Language to English as an International Language (EIL) or English as a Lingua Franca (ELF). In line with the research by Derwing and colleagues discussed above (e.g., Derwing and Munro, 1995; Derwing and Munro, 2009; Thomson and Derwing, 2014), the focus of ELF is on intelligibility. However, in contrast to EFL, ELF puts the importance on mutual intelligibility between non-native speakers. In order to achieve this mutual intelligibility, Jenkins formulated the Lingua Franca Core (LFC). The LFC is an extensive phonology with guidelines as to what elements of phonology should be prioritised in English language teaching.

The features that should be prioritised, according to Jenkins' ELF proposal, are referred to as core features, whereas the features that should not be prioritised are regarded as non-core. One clear difference between teachings of the LFC and traditional EFL relates to consonant sounds. The LFC state that all consonants should be taught, with the exception of /θ/ and /ð/. Many learners of English will remember being corrected in the EFL classroom when failing to produce these sounds. It is important to note that Jenkins does not say that non-core features must not be taught, as she clearly acknowledges the personal choice of the learner (Jenkins, 2009).

The LFC has been the topic of some debate due to some of the claims made by Jenkins (2000). One such claim is that nuclear stress (i.e., the addition of stress to a particular word in an utterance to make this word stand out) should be a core feature, but word stress should not. The reason for this assessment is that word stress: is regarded as impossible to teach, varies among NS varieties of English, and does not cause intelligibility loss between non-native speakers. As will be shown in Chapter 2, at least the last of these claims is inaccurate.

One very appealing, but perhaps slightly naïve, element of ELF is that it puts the demands on native and non-native speakers alike. As stated by Jenkins (2009), 'no matter which circle of use we come from, from an ELF perspective we all need to make adjustments to our local English variety for the benefit of our interlocutors when we take part in lingua franca English communication' (p. 201). Unfortunately, anecdotal evidence suggests that many native speakers primarily grade their language by speaking more loudly.

A criticism often levied against the LFC is that it is based on a rather small data set. The initial core rules were formulated based on 'five classroom or social interactions in which there was a communication breakdown, five information exchange tasks, and recorded social exchanges' (McCrocklin, 2012, p. 250). Some research has been conducted on the phonology of LFC (e.g., Lewis and Deterding, 2018), but there can be little doubt that more research would be helpful to inform this area. Even so, ELF remains an interesting alternative to EFL. For many teachers who are unsure about what aspects to teach, ELF provides a useful framework for pronunciation teaching.

1.4.3. Focus: Segmentals or Suprasegmentals?

In 1.3, the case was made that pronunciation can indeed be taught. This has been shown to be the case for both segmental (Elliot, 1995; Saito and Lyster, 2012) and suprasegmental (Pennington and Ellis, 2000) aspects of English. An extension of these findings is to see how they affect pronunciation.

Segmental and suprasegmental features of an L2 differ in the challenges they pose to the learner. With the limited time available for teachers to teach pronunciation in the classroom, knowing what aspects of pronunciation to focus on would offer teachers a great deal of help. Unfortunately, when it comes to pronunciation instruction 'there is no agreed-upon system of deciding what to teach, and when and how to do it' (Darcy, Ewert, and Lidster, 2012). In recent years, researchers have looked at whether a focus on segmentals or suprasegmentals yields the most improvement in the L2 English skills, but the results have been somewhat contradictory.

On the one hand, a series of studies have suggested that segmental instruction significantly contributes to learners' oral skills. As mentioned above, based on conversational data, Jenkins (2000) proposed that when non-native speakers interact with one another, consonant sounds are the most important and that nuclear stress seems to be the only suprasegmental feature that might interfere with intelligibility. However, this is far from the whole picture.

Bent, Bradlow and Smith (2007) investigated the relationship between segmental errors and intelligibility ratings. While the study revealed that segmental errors in vowels have a negative effect on intelligibility, errors in consonants showed no such effect. Zielinski (2008) found, in turn, that segmental errors in stressed syllables were particularly crucial in predicting low intelligibility ratings.

A large number of authors have argued in favour of focusing on suprasegmental aspects of pronunciation instruction (Anderson-Hsieh and Kohler, 1988; Field, 2005; Hahn, 2004). For example, McNerney and Mendelsohn (1992) state that 'a short term pronunciation course should focus first and foremost on suprasegmentals as they have the greatest impact on the comprehensibility of the learner's English' (p. 132 – Quoted from Hahn, 2004). Fraser (2001) rightly points out that it makes little sense to help learners with their production of speech sounds if their control of word and sentence stress is poor, as the inaccurate word stress in itself will render the learners unlikely to be understood regardless of how accurate their segmental production is.

In relation to the discussion of foreign accent, studies have shown that untrained listeners can detect a foreign accent even when segmental features are obscured by low-pass filtering – a technique that removes the listener's ability to distinguish segmental features (Munro, 1995). This could be taken as evidence of the importance of teaching suprasegmentals in the foreign language classroom if the aim of the instruction is to help learners with their accents. Moreover, misplaced nuclear stress has been found to cause communication breakdowns between non-native speakers of English (Jenkins, 2002) and make speech less intelligible for NS listeners (Hahn, 2004). It is very plausible that some suprasegmental features are more important than others. For example, Field (2005) looked at how misplaced word stress affects intelligibility and concluded that, while a change in lexical stress does affect intelligibility, 'pronunciation teaching programs should rank word stress at a medium level of importance' (p. 419).

It is important to acknowledge that not all suprasegmental features may be equally easy to learn as shown by Pennington and Ellis (2000). Their study compared the role of attention on four different aspects of English prosody. Although the participants' performance was generally poor, contrastive stress did stand out as an area that the learners seemed to grasp better than the others. Their findings led Levis (2005) to claim 'that some elements of intonation, such as nuclear stress, appear to be learnable, other elements, such as pitch movement marking boundaries and the intonation of sentence tags, are not' (pp. 369-370). This might be a stretch, though, as the study in question only focused on one type of training of EFL learners with Cantonese as their L1, but the findings do indicate that not all elements of suprasegmental phonology are equally easy to learn.

It should be pointed out that, while in some studies on the topic of segmentals versus suprasegmentals have explored either one or the other, others have compared the two in the same study. Anderson-Hsieh, Johnson, and Koehler (1992) found evidence that both segmental and suprasegmental features affected global pronunciation ratings (intelligibility and acceptability), but that suprasegmental features generally showed a larger effect. In a similar vein, Derwing, Munro, and Wiebe (1998) looked at segmental and suprasegmental instruction in comparison to no instruction at all. The aim of the study was to find out how the different types of instruction affected both accentedness and comprehensibility over the course of 12 weeks. The participants were rated on both a sentence reading task and a picture-guided narrative task. In the latter case, fluency was also assessed. After the 12 weeks had passed, both groups receiving instruction improved on both accentedness and comprehensibility when reading sentences out loud. However, only the group receiving instruction on suprasegmental features showed improvement in spontaneous speech, but only in fluency and comprehensibility. The participants' degree of accentedness in spontaneous speech did not change as a result of the instruction. In the read-out-loud task, on the other hand, the group receiving segmental instruction improved significantly more than the two other groups. Of these two, the group receiving no instruction made no improvements, whereas the group receiving instruction on suprasegmentals made some improvements.

The point has been made by researchers that the debate about which of the two foci is better is not a helpful one. According to this view, what some studies have not sufficiently acknowledged is the interdependence of segmental and suprasegmental features (Zielinski, 2015). In this view, the discourse needs to change entirely because it fails to include the role of the listener. This can be exemplified with a study by Zielinski (2008). In this study, NS listeners were asked to transcribe the speech of NNSs with different language backgrounds. Cases of unintelligibility were analysed from the point of the speaker as well as the point of the listener. One of the findings of the study was that in some cases, the listener reported loss of intelligibility due to a missing segment. However, the speaker had left out the segment because it broke the rules of their L1 syllable structure, and could thus be said to have a suprasegmental origin.

Field (2005) makes the point that many studies fail to acknowledge the many different factors that comprise suprasegmental phonology. In an attempt at shedding more light on how individual elements contribute to intelligibility, he investigated the role of lexical stress. According to his data, intelligibility is compromised for both native and non-native listeners when stress is shifted from one syllable to another, but primarily if the direction of the shift is to the right. Additionally, he reported that a change in vowel quality seemed to lessen the effect of the stress change. This finding was both counter-intuitive and unexpected. First of all, one would not expect additional deviation from the standard form of a word to improve the intelligibility of that word. Moreover, the results ran counter to the findings of Cutler and Clifton (1984), who found that deviating vowel quality had a detrimental effect on intelligibility (see Chapter 2).

To sum up, there can be little doubt that instruction in both segmental and suprasegmental features can be helpful, although one has to keep in mind that they may help with different aspects of the learner's L2 phonology. Segmental instruction may help learners with their accuracy, but this does not necessarily translate to increased intelligibility. Suprasegmental instruction may lead to improvements in both areas, but that does not mean that suprasegmental instruction will help learners in all aspects of segmental phonology. This is not to say that segmental instruction is not worthwhile, as many learners directly request help with specific speech sounds. In an ideal world, teachers would naturally focus on both, but the reality is that, in many cases, teachers lack sufficient time to do so. As will be argued later, technology could be part of the solution to this problem.

1.4.4. Other Factors

The teaching and learning of L2 pronunciation is a multi-faceted issue. When looking at pronunciation teaching from the teachers' perspective, a few factors appear to be important. As such, teachers face a number of practical issues with regard to teaching pronunciation. Many EFL teachers feel they lack proper training and knowledge to teach pronunciation (Breitkreutz, Derwing, and Rossiter, 2002; MacDonald, 2002). This can be explained by the fact that the pronunciation modules on teacher training courses often lack a comprehensive phonology component, and only scratch the surface of what is required to teach pronunciation, and in some cases, there is no pronunciation module at all (Henderson et al., 2012). As a consequence, teachers are left to rely on their intuition to teach pronunciation (Thomson, 2012). This should come as no surprise considering that the majority of pronunciation research has been conducted in non-teaching contexts (Wang and Munro, 2004), and published in journals that have very few, if any, language teachers among their readers (Derwing and Munro, 2005). On top of this, many teachers feel that good resources for pronunciation teaching are hard to come by (Foote, Holtby and Derwing, 2011; Henderson et al., 2012) making it even less likely that teachers will make pronunciation a part of their lessons. While there are materials available such as *Jazz Chants* by Graham (1978) as well as books with pronunciation rules and exercises such as the 'English Pronunciation in Use' series (Hancock, 2003; Hewings, 2007; Marks, 2007), research has done little to help teachers gain useful tools to use in their classrooms (Couper, 2006). Additionally, some of the resources that have been developed for teachers have not necessarily been based on sound classroom-based research, thus putting their applicability into question (Baker and Murphy, 2011). Furthermore, pronunciation as a research topic received very little attention for many years, and has only seen a revival in the past decade or so (Thomson and Derwing, 2014).

Apart from the factors already mentioned, there is a variety of other factors that many teachers have little or no control over. One of these is the L1 background of the students being taught. This is an issue in an ESL context, in particular, as the learner's L1 heavily influences their strengths and weaknesses when it comes to L2 pronunciation. This is less of an issue in an EFL context, but should nonetheless be considered. Completely monolingual classrooms are becoming increasingly rare, and a student whose L1 is different from that of, say, most students in the classroom should not suffer a disadvantage because of their L1. This means that pronunciation teaching should ideally be tailored to the specific multilingual context of teaching so that all students in the classroom benefit equally from the instruction, irrespective of their L1.

In many EFL classrooms, however, this idea may be at odds with the reality language teachers face due to time constraints (Anderson-Hsieh, 1990; Kelly, 2000; Setter, 2008). Since many teachers have to get through a set curriculum, they may not feel they have the time to teach pronunciation (Gilbert, 2008; Munro and Derwing, 2006). This issue is further aggravated by funding being cut in many institutions meaning that schools have to provide the same, or an even better service, with fewer resources (Ducate, Lomicka, and Lord, 2012). With classroom time being reduced, it becomes clear that more learning has to take place outside the classroom, and pronunciation has already been suggested to be an area of language learning that is best learnt through study outside the classroom (Acton, 1984; Anderson-Hsieh, 1990; Thomson, 2011).

The issue of time constraints is nothing new, it should be added. One of the arguments for the continued use of the Grammar-translation Method was that it was believed that teachers had insufficient time in their schedules to teach the spoken language (Grandgent, 1892). Ironically, even if curricula were changed to allow for more pronunciation teaching, and teachers were taught what aspects of pronunciation to focus on, some teachers might be set in their ways to such an extent that they would be reluctant to implement the curricular changes (Darcy, Ewert, and Lidster, 2012). If this is indeed the case, taking parts of pronunciation teaching out of the classroom might be an ideal solution. Finally, when practising pronunciation in the classroom, it is paramount that the students feel comfortable (Morley, 1991). However, in a class with 10 or more students of different ages, genders, and backgrounds, creating the right kind of atmosphere in the classroom can be a challenge for even the most experienced teacher.

A solution to several of the problems described above can be sought in technology. For example, by allowing learners to learn a language using a computer, the teacher can ensure that the learner is able to work at their own pace in a stress-free environment (Neri et al., 2002). As will be discussed in section 1.5, a large variety of technological tools now exist, which are likely to enhance the learning experience in some way.

Students of today have been labelled 'digital natives' (Prensky, 2001), which means that they have grown up in a digital world and generally know the ins and outs of mainstream types of technology. While the extent of the implications of this proposed new generation is still being debated (Bennett, Maton, and Kervin, 2008), there can be little doubt that most people are generally becoming more comfortable with online technology such as hyperlinks, learning management systems, and social media.

Finally, this section has mainly dealt with the challenges teachers face. However, it must not be forgotten that many of these challenges also pertain to the learners. For example, some of the linguistic features such as highly variable word and sentence stress patterns make English even more difficult to learn for L1 speakers of languages that do not share these features (see Chapter 2 for discussion). In addition, learners of English face the issues of the intricate sound-grapheme correspondence in English. As a consequence, learners cannot generally rely on the spelling of a word to tell them how it is pronounced leading learners to pronounce the /l/ in words like *walk, talk, and palm* (Mompeán and Fouz-González, 2016). Another consequence is that many teachers are tasked with difficult questions such as why *doom and door* have different vowel sounds.

In summary, there is a long list of reasons why English pronunciation is difficult to teach and difficult to learn. Some are very basic, practical issues such as time constraints and lack of training on the teacher's part. Other issues are linguistic in nature and simply stem from the fact that the L1 and L2 are different. Finally, as discussed in 1.3.2, some obstacles are to be found within the learner and may range from social to perceptual factors. As in many other areas of life, technology has been suggested as a tool to make the task at hand more manageable. Although technology cannot yet solve all these issues mentioned so far, great strides have been made in recent years to aide teachers and learners. The next section looks at how technology has been applied in the past to assist pronunciation teaching.

1.5. Technology and L2 Teaching and Learning

Section 1.2 above described how poor pronunciation in an L2 can have severe consequences. Fortunately, as shown in 1.3.3, it is clear that, although L2 pronunciation can be extraordinarily challenging, pronunciation instruction can be effective. However, teachers are still faced with the issue of finding time in their teaching schedule to teach it. Technology seems an obvious area in which to look for assistance in this matter. As Levis (2007) states:

> the use of computers is almost ideally suited to learning pronunciation skills. Computers can provide individualized instruction, frequent practice through listening discrimination and focused repetition exercises, and automatic visual support that demonstrates to learners how closely their own pronunciation approximates model utterances. (p. 184)

Language learning mediated by a computer is known Computer Assisted Language

Learning (CALL). Within this field, the use of computers to learn pronunciation is referred to as Computer Assisted Pronunciation Teaching (CAPT). Research in both fields has developed rather quickly over the past few decades. This is likely to be a consequence of the rapid development in technology that we are currently witnessing, as the main issues in CALL in the past were related to computer hardware. This was exemplified by Stenson et al. (1992), who struggled to save all their recorded speech data on their 30MB hard drive.

With every new gadget and every new application comes a potential for learning, which needs to be investigated. That said, it is important to note that technology should always be used with a specific goal in mind and not just for the sake of using it (Chun, Kern, and Smith, 2016).

As detailed in the beginning of Chapter 1, this chapter explores the development of CALL with a particular focus on CAPT as well as some of the advantages and possible issues in using technology for pronunciation teaching and learning. From a student's perspective, practising pronunciation using a computer must have a certain appeal as it enables the student to work alone without being judged by peers (Eskenazi, 1996). Anyone who has ever taught pronunciation to students in a classroom will also have seen the blushing face of a student who struggles with a specific speech sound or intonation pattern. With the introduction of the right technology, this could be a thing of the past as computers can provide a stress-free environment for students to practise in and thus do away with much of students' language learning anxiety in a teaching context (Horwitz, Horwitz, and Cope, 1986; Laroy, 1995; Neri, Cucchiarini, and Strik, 2003). On the other hand, some students have indicated that they did not feel comfortable working with a computer and would prefer working with a human tutor (Stenson et al., 1992). It is important to notice, however, that this particular study was published in a time when computer use was less common than it is today. As will be shown later, students generally rate technological use in an educational context quite favourably.

An advantage that may benefit learners as well as teachers is that when working with computers, teaching and learning can be asynchronous (Chun, Kern, and Smith, 2016; Estebas-Vilaplana, 2015). The teacher can thus upload instructions in the shape of links or worksheets to a virtual learning environment (VLE), and the students can work with the materials when they have time. Alternatively, teachers can also post pronunciation instructions that students can access on their smartphones using applications such as Twitter (e.g., Mompeán and Fouz-González, 2016).

A drawback to these approaches is that in pronunciation teaching, feedback/error correction is sometimes necessary. However, for many aspects of pronunciation instruction, critical listening (i.e., a technique where students record and listen to their own voice or in some cases compare their productions to a NS model) might suffice (Dauer, 1993). Finally, as will be discussed later in this chapter, automatic speech recognition (ASR) might in some cases be able to provide the needed error correction during self-study.

1.5.1 Feedback

One of the advantages often pointed out when discussing CALL is the computer's never-fading ability to provide feedback regardless of how many errors the learner makes (Eskenazi, 1999a). Hence, feedback is a potentially powerful tool in relation to CAPT because it gives the learner a chance to receive consistent information about their speech production in one form or another.

Feedback in language learning generally means that a tutor (or a computer) comments on a learners' production. This may constitute either comments affirming what a student has done correctly (positive feedback) or something the student needs to change (negative feedback). The latter is often referred to as corrective feedback. This can be given in the shape recasts (i.e., repeating a non-target-like feature in a target-like way), specifying rules that the learners need to be aware of, or any other comment that aims to help the learner improve a given linguistic feature.

Although some scholars have claimed that corrective feedback is essentially ineffective, such as Truscott (2007), who looked at corrective feedback in L2 writing, there now seems to be ample evidence that corrective feedback can indeed be effective in L2 learning (Lee et al., 2015). Hence, feedback is now seen as a crucial element of pronunciation teaching (Celce-Murcia et al., 2010).

In the L2 classroom, two common types of feedback can be identified, namely verbal feedback and written feedback. The former is usually given during lessons, while the latter is usually provided outside of lessons (e.g., written comments on an assignment). As will be shown in the two studies in this thesis, written feedback can also be provided in a digital context.

When using computers for pronunciation teaching, feedback can also be given in a variety of ways. It can be auditory, meaning that the learner receives feedback as sound (see 1.5.1.1). Alternatively, feedback can be visual, meaning that the learner can somehow see their speech production on a screen (see 1.5.1.2). Finally, the feedback can be audio-visual which combines the two modalities.

1.5.1.1. Auditory Feedback

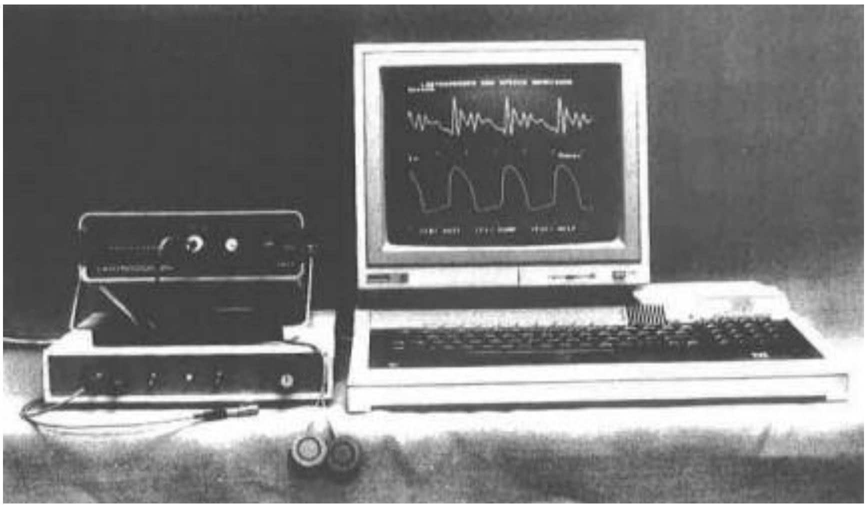
Auditory feedback can be provided by repeating the correct sound or speech pattern back to the learner. The largest number of studies using auditory feedback is probably the studies using High Variability Phonetic Training (HVPT; see 1.5.3). As this topic is covered elsewhere, these studies will not be included here. Outside the field of HVPT, there seems to be relatively few studies on auditory feedback only. In many cases, however, this type of feedback is compared to other types in studies looking at two or three modalities. One such study was Dowd et al. (1998). In this study, L1 English speakers were taught French non-nasalised vowel sounds using either auditory feedback or visual feedback. The results showed that the learners using visual feedback outperformed the group using auditory feedback. The authors described these findings as surprising since 'the auditory feedback method [...] is a major component of the method used by most children to learn to speak their native language' (p. 18). However, given that first and second language acquisition are likely to involve slightly different mechanisms, this should perhaps not be such a surprise after all.

The studies that have looked at auditory feedback only, generally yielded mixed results. For example, Jugler and Mobius (2015) used two types of auditory feedback for two different groups, namely a modified version of the participants' own voices and a native speaker model. The study found that training German learners of French to produce stop sounds in the L2 was effective for voiceless stops but not for voiced stops.

1.5.1.2. Visual Feedback

Despite being far from commonplace, computers were slowly taken into use for private, work, and educational purposes in the late 1970s and early 1980s. In an educational context, this led to new types of feedback being made possible. Among the earliest studies of the use of feedback is work using a laryngograph (Figure 1.1), an instrument that measures airflow through the larynx and displays changes in the airflow on a screen (e.g., Abberton and Fourcin, 1975; Fourcin and Abberton, 1977).

Figure 1.1.
Picture of a laryngograph – adapted from Abberton, Howard, and Fourcin (1989).



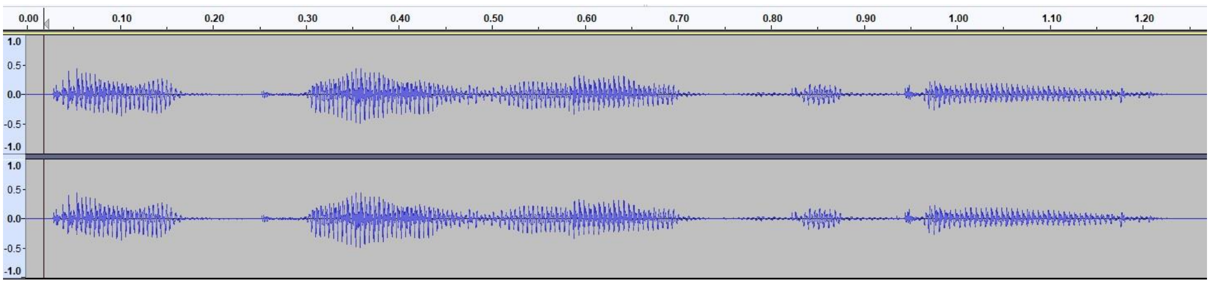
With this new equipment, learners could receive a visual display of elements in their speech which they might not be able to hear. In some cases, they would even be allowed to compare the visuals of their speech to a model produced by their teacher (Albertson, 1982). Although the studies of these tools generally reported positive findings, on closer inspection, de Bot (1980) argued that more research was needed before the applicability of this instrument could be asserted.

Waveforms

The use of visual feedback has continued since the 70s and 80s, and has been adopted by various companies in their language learning materials. One such type of feedback is the waveform, which has been widely used in commercial language learning software such as *Rosetta Stone, Tell Me More*, and *Fluenz*. In essence, the waveform is a visual display of the intensity of a talker's speech production (Hincks and Edlund, 2009). Higher intensity is reflected in higher amplitude in the signal displayed (Figure 1.2).

Figure 1.2.

Waveform depiction of one participant's production of the sentence *I quite like the guy.*



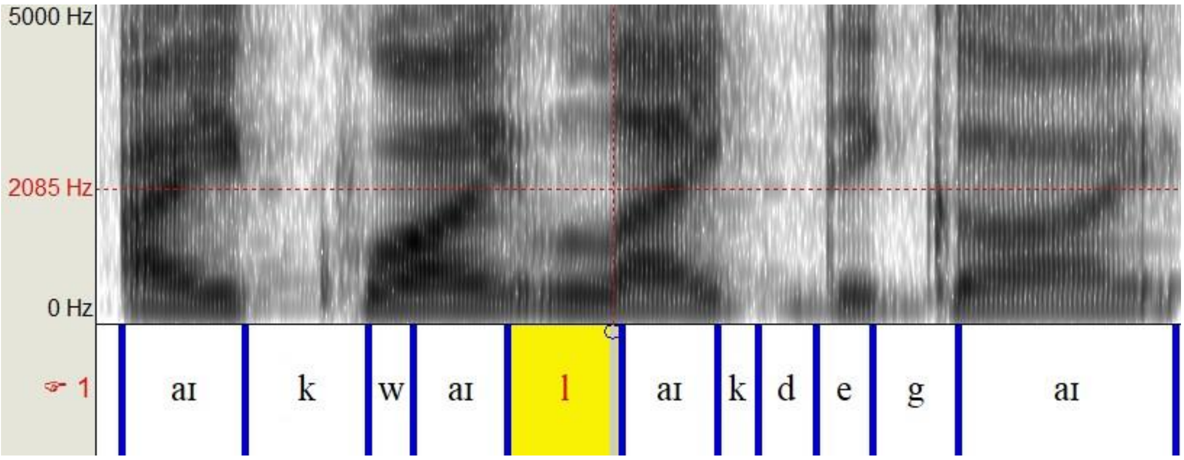
Despite their popularity among software developers, waveforms have not generally been well received by researchers who, in general, seem to have considered them 'to not be worth the pedagogical time and energy, at least when used as the sole source of feedback' (Levis, 2007, p. 190). However, their usefulness might depend on how they are employed. In commercial language learning software, the learner is often left to compare their own recorded waveforms to that of a native speaker model (Neri et al., 2002). Without any advice on how to use the waveforms, not only does it become an impossible task, but it also becomes pedagogically misleading as two waveforms can differ and yet both be the product of a correct pronunciation (Neri, Cucchiarini, and Strik, 2002). There are indications, however, that there might be a place for waveforms in L2 learning after all, as researchers have successfully used this type of feedback to teach various aspects of L2 pronunciation such as duration of geminates in Japanese (Motohashi-Saigo and Hardison, 2009) and English nuclear stress (Coniam, 2002). In the latter example, comparisons between L1 and L2 waveforms were used to draw the attention of non-native EFL teachers (L1 Cantonese) to the stress-timed rhythm of English. Whether the participants improved as a result of the training was not tested, but the feedback provided to the researcher was positive. The participants believed they had benefited from using the waveforms and reported that they found the waveforms a useful tool for illustrating the difference between stress timing and syllable timing. Finally, in this particular case, the learners required significant amounts of instruction in how to interpret the feedback, so it might not be an ideal solution for independent study.

Spectrograms

A spectrogram is a visual representation of how frequency and intensity in acoustic signals change over time. Spectrograms can use either colours or a grey scale, as in Figure 1.3 below. Here, an increase in the intensity of a frequency corresponds to the darker sections of the spectrogram. As an example, Figure 1.3 shows the production by a participant in Study 2 of the sentence *I quite like the guy* (from the pre-test). The spectrogram makes it possible to discern contrasts in intensity. As such, the darker areas are likely to be vowel sounds as these have higher intensities. Similarly, the very light areas are likely to indicate stops such as /k/.

Figure 1.3.

Spectrographic representation of the sentence *I quite like the guy.*



Note: The participant producing this item dropped the /t/ in *quite.*

Spectrograms have been used to help improve L2 learners' pronunciation, but are generally considered to be rather difficult to interpret for non-specialists (Akahane-

Yamada et al., 1998; Chun, Hardison, and Pennington, 2008). In spite of this, research has yielded positive results from employing spectrograms as a teaching tool. Akahane-Yamada et al. (1998), for example, used spectrographic visuals to train Japanese L1 learners of English the /r - l/ distinction, which typically causes problems for this group of learners. The participants improved both in production and perception – more so in the former than the latter. However, in the pre-training phase, the participants received instruction in correct tongue placement for the targeted speech sounds, which may obscure the true contribution of the spectrograms to some extent.

Lambacher (1999) also argues in favour of the use of spectrograms in pronunciation teaching although he does concede that this type of tool should be used in combination with other activities. For example, when teaching plosives like /p, t, k/, spectrograms could be used alongside an activity such as blowing away a piece of paper, to better understand the concept of aspiration. He further stresses the importance of raising students' awareness of the differences between L1 and L2 speech sounds as it is in this regard, particularly, that spectrograms can be useful. Raising students' awareness through the use of spectrograms was also the method chosen by Saito (2007), who concludes that spectrograms should be considered as a tool for helping learners think more about their tongue movement when producing English speech sounds.

Although spectrograms can be used to help students acquire the English /r - l/ or /t - d/ contrasts as well as other sound contrasts, the implementation in an EFL classroom might not be an easy one. Because extracting information from a spectrogram is such a complex process for a non-specialist, a teacher in a multilingual classroom in particular would have to allocate disproportionate amounts of time for instruction in how to interpret the feedback the students receive.

This, in turn, assumes that the teachers themselves know how to interpret a spectrogram, which is by no means a given, as Thomson (2011) states:

> [w]hen it comes to providing displays of the acoustic characteristics of individual segments (i.e., spectrograms), there is seemingly no benefit to learners. Spectrograms are uninterpretable to non-experts, and do not convey any information that can be readily used to improve pronunciation. (p. 747)

In sum, although spectrograms have proven useful in research, they should probably not be the first tool of choice for most teachers.

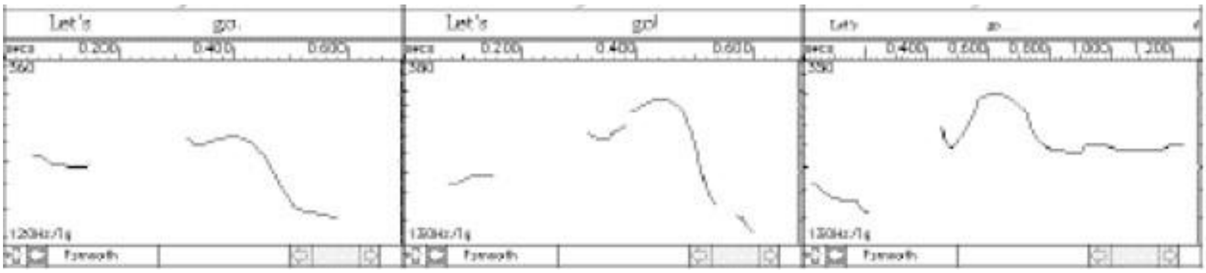
Pitch contours

Pitch contours have generally been perceived as more useful than waveforms and spectrograms in terms of pedagogical applicability. The clear advantage that pitch contours have over these other tools is that the learner can quite easily see the pitch contour move as a result of their speech production. Because of this, the feedback becomes much more salient (Busà, 2008; Chun, 1989). This is particularly important as the direction of a pitch change can be very hard to perceive (de Bot and Mailfert, 1982).

A critique that has been levied against the use of pitch tracking is that feedback can only be produced for voiced sounds. Thus, if an example sentence in an exercise contains too many voiceless sounds, it becomes more difficult for the learner to make the connection between their speech production and the feedback (Anderson-Hsieh, 1994; Chun, 1998). Notwithstanding this issue, pitch contours have nonetheless been used effectively in L2 intonation research (Spaai and Hermes,

1993) as well as for helping deaf L1 learners (Abberton and Fourcin, 1975; Spaai et al., 1994). Although this is a clear limitation to the applicability of pitch contours as feedback, they can nonetheless serve a purpose. Figure 1.4 shows three productions of the phrase *let's go* produced with three different tones. On the left the utterance is neutral, in the middle the speaker expresses impatience, and on the right the utterance is cajoling (Chun, 1989).

Figure 1.4.
Three renditions of the phrase *Let's go* using three different tone patterns – adapted from Chun (1989).



Not all studies have shown equally promising results. In fact, Vardanian (1964) found that the control participants outperformed the participants receiving treatment. However, the results were not statistically significant. In this study, Vardanian used a fundamental frequency extractor paired with an oscilloscope to provide language learners with feedback on their intonation. In a similar vein, but using newer technology, Stenson et al. (1992) found no statistical difference between participants who trained with *Windows SpeechViewer* and participants who did not use this software. On a positive note, the overall feedback from the participants was generally positive.

Several other studies have shown good results using pitch as visual feedback (e.g., Hardison, 2005; Hirata, 2004; Taniguchi and Abberton, 1999). Over the course of ten 30-minute training sessions, Hirata (2004) used fundamental frequency (essentially what listeners hear as pitch) as feedback to improve L1 English speakers' ability to perceive and produce pitch and durational contrasts in Japanese. Working first with individual words and later with full sentences, the participants further showed improvement in novel words and sentences.

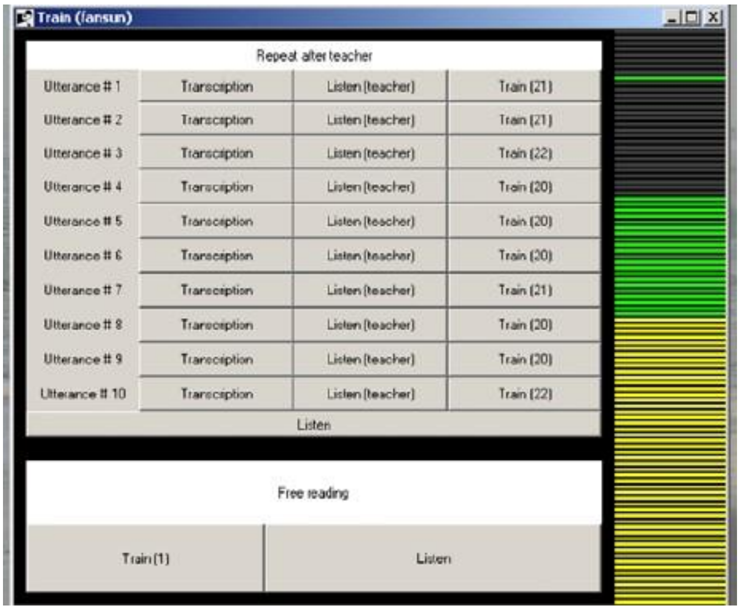
Levis and Pickering (2004) argue that intonation must be taught at discourse level as they found that native speakers use different intonation depending on whether they produced individual sentences or discourse level text. This view is supported by Hardison (2005), who found that although participants training with sentence level material improved in some areas, participants who trained with discourse level materials showed more improvement in free speech. It is worth mentioning that Hardison's participants were all advanced learners, which could be important when considering the appropriateness of this type of training, as it could be hypothesised that lower-level learners might struggle with the larger amounts of input that is provided at discourse level. Interestingly, Hincks and Edlund (2009) argue that using speech analysis software is not ideal with discourse level materials. While they acknowledge that pitch contours can be used for simpler materials, the authors state 'the standard technique can be advantageously used for practicing phrases in the type of pronunciation training done at elementary levels of language training, but is inadequate for stimulating intonational development over longer stretches of discourse' (p.34).

As a final note, it must be said that considering that several studies have shown good results using materials that never went beyond the level of individual sentences (de Bot, 1983; Hardison, 2002; Hardison, 2004; Hirata, 2004; Weltens and de Bot, 1984), it might be a stretch to claim that sentence level practice is no good. It is more likely a case of one being slightly better than the other in certain contexts.

As an alternative to pitch contours, Hincks and Edlund (2009) suggested using yellow and green flashing bars to indicate pitch variation (see Figure 1.5).

Figure 1.5.

Pitch representation used by Hincks and Edlund (2009).



The students can listen to a target pronunciation model and try to replicate this model using the flashing intonation metre to guide them. According to the authors, this is an easy-to-use solution designed for self-study. Because everything is done in real-time, and because results cannot be compared to an NS model, or indeed to the student's own productions, it is not possible for the student to receive feedback other than the immediate response delivered upon an utterance. While this is undoubtedly a novel approach to teaching intonation, and despite their good results in terms of participant feedback, visual pitch representation of this type does not seem like a ready-made, stand-alone solution to pronunciation training. Although their results showed that the participants used intonation more, their overall pronunciation decreased after training, although only very slightly. Lastly, there are a few potential pedagogical issues that need addressing; most important of which is the fact that their intonation metre did not actually assess whether the applied intonation was appropriate. It merely told the student whether they applied more or less varied intonation than in previous attempts. This approach has the obvious flaw that students could be using the intonation wrong and yet believe that they are improving their spoken English, despite the authors arguing that they:

> see the computer's role as more of a 'tool' than a virtual tutor—a tool that will provide a learning environment capable of responding interactively to learner production, without attempting to provide 'right' or 'wrong' answers to the way the student delivers the presentation. (p. 36)

Ultrasound

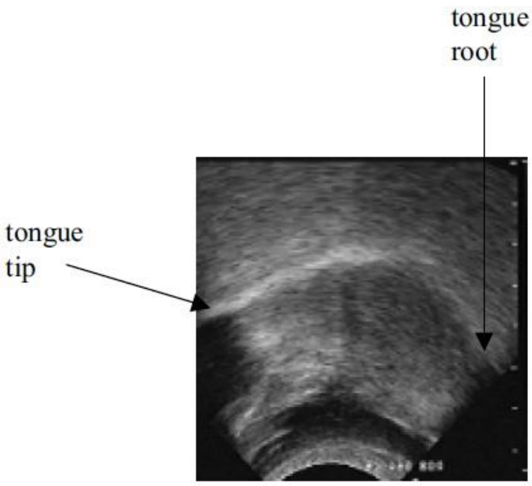
As a more recent addition to the areas of research on visual feedback in L2 acquisition, ultrasound can be mentioned. Based on positive results from speech therapy (e.g., Adler-Bock et al., 2007; Bernhardt et al., 2005), Gick et al. (2008) tested ultrasound imaging as a tool for L2 pronunciation teaching and found that all three of their participants improved post-training. The main advantage of this type of feedback is that learners can see their tongue movements in real-time as they produce speech. This also makes it easier for the instructor to point out what needs to change.

While this technology has often been viewed as too expensive and inaccessible for L2 teaching, recent advances in technological developments have now made ultrasound equipment 'lightweight, portable and relatively affordable' (Holt, 2018, p. 252).

Ultrasound imaging allows for fairly detailed depiction of the position of the tongue (see Figure 1.6), thus allowing students to work on difficult speech sounds.

Figure 1.6.

Ultrasound image of the tongue taken from Gick et al. (2008).



In a Spanish context, the distinction between /i/ and /I/ would be an obvious area of focus, but any phonemic contrast involving tongue position could, in theory, be trained. This is a great advantage because tongue position is virtually impossible for an L2 instructor to show, and can be very difficult to explain compared to features such as lip rounding (Catford and Pisoni, 1970).

Despite the promise that this technology has shown, it does have a few downsides which must be mentioned. Firstly, only one student will be able to use the equipment at any one time, so, unless the institution invests in multiple devices, training all students in a class is currently a slow process. Secondly, this type of training would only be available to the students while on the school premises, meaning that out-of-class practice would be impossible. Finally, as can be seen from Figure 1.6 above, although the image is fairly detailed, it is likely that the instructor will have to dedicate a significant proportion of class time to show students how the technology is used. For these reasons, it is probably fair to say that strong evidence of the superiority of ultrasound to other – more affordable – technologies will be needed before schools start to invest in ultrasound scanners.

Electropalatography

As a final example, electropalatography should be mentioned. Originally designed for speech therapy and phonetic research, the idea behind this technology is that learners are able to see how their tongue moves inside their mouth – much as when using ultrasound – thus giving both them and the instructor a good idea of where potential problems lie. The tongue movements are registered by a device (an artificial palate fitted with electrodes) which is inserted into the mouth of the learner (Figure 1.7).

Figure 1.7.

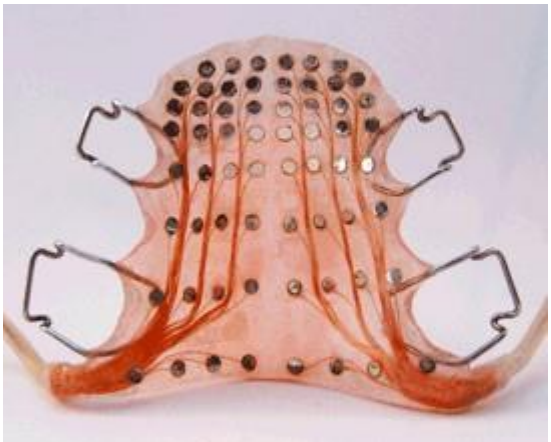
Model of how electropalatography is used.

Despite having been around since the 1960s, this is still a niche technology, and there are some obvious challenges that need addressing before it can be put into large-scale application in SLA. For example, in order for the device to be accurate, each learner should ideally have a customised one, making the technology very expensive.

Furthermore, although great efforts have been made to make the device as light as possible (see Figure 1.8), learners may still feel uncomfortable speaking with something covering their palate.

Figure 1.8.

Example of an artificial palate used in electropalatography.



As a consequence, users may not produce speech the same way they otherwise would (Hickey, 1992). This is a big problem indeed, because it can mislead both the learner and the teacher with regard to what speech production issues need working on. In spite of this, researchers have produced positive results using this technology (e.g., Schmidt, 2012). Despite not being the most popular branch of CAPT, research to improve the technology further is still ongoing, but more research is needed to unearth its full potential (Krynicki and Michalski, 2019).

It is difficult to say which of these types of feedback is better, and it might not even be a relevant question, because each tool has its advantages and disadvantages, and thus serves a different purpose to the others. Stenson et al. (1992) sum it up quite aptly when stating:

> [c]omputer-based visual display equipment has a definite participative value as a motivator for both students and instructors, but such technology is perhaps most useful as a supplement to rather than a replacement for the human interaction of teacher and student in the teaching of pronunciation. (p. 16)

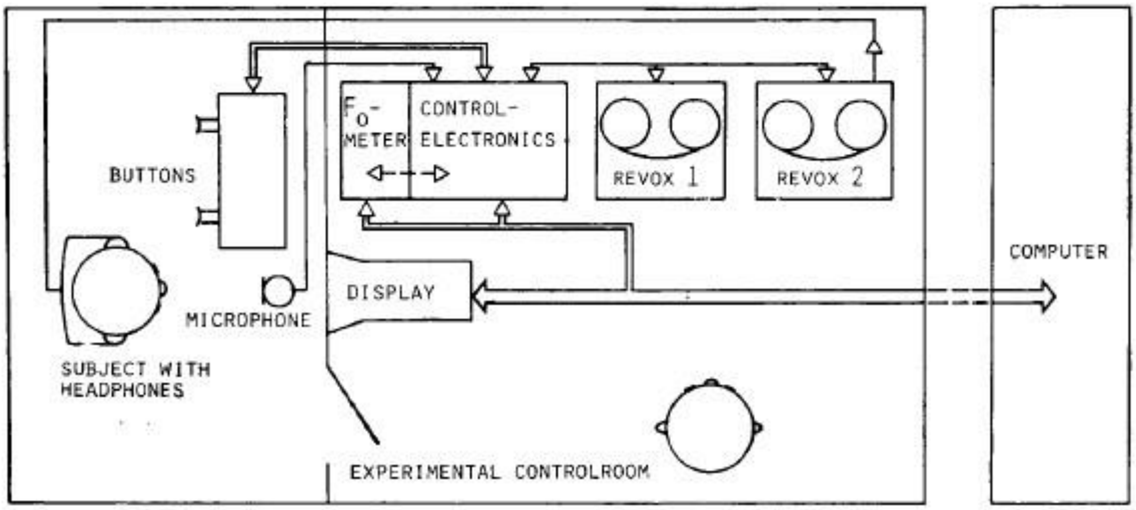
However, one commonality which all of the tools presented above share is that they take time to learn how to use. As stated previously, time is a valuable resource for many teachers, so tools that can be used without much prior instruction should be favoured. Likewise, tools which can be used outside the classroom in the learners' own time would have a distinct advantage over the rest.

1.5.1.3 Audio-visual Feedback

Audio-visual (AV) feedback, as the name implies, uses two modes for providing feedback. This means that learners can both see a visual representation of their speech and hear either their own production back or hear a model speaker to imitate. Despite initial negative results (Wichern and Boves, 1980), studies using this type of feedback have generally yielded positive findings. In a 1983 study by de Bot focusing on the pronunciation of 59 L1 Dutch students learning English, audio-visual feedback (pitch contours + own productions) was compared to auditory feedback only as well as no training at all. The setup for this study is a good example of the how this type of experiment/training works (Figure 1.9), but it should be added that improvements in technology have made more recent experiments less demanding in terms of space.

Figure 1.9.

Experimental setup for studies on audio-visual feedback.

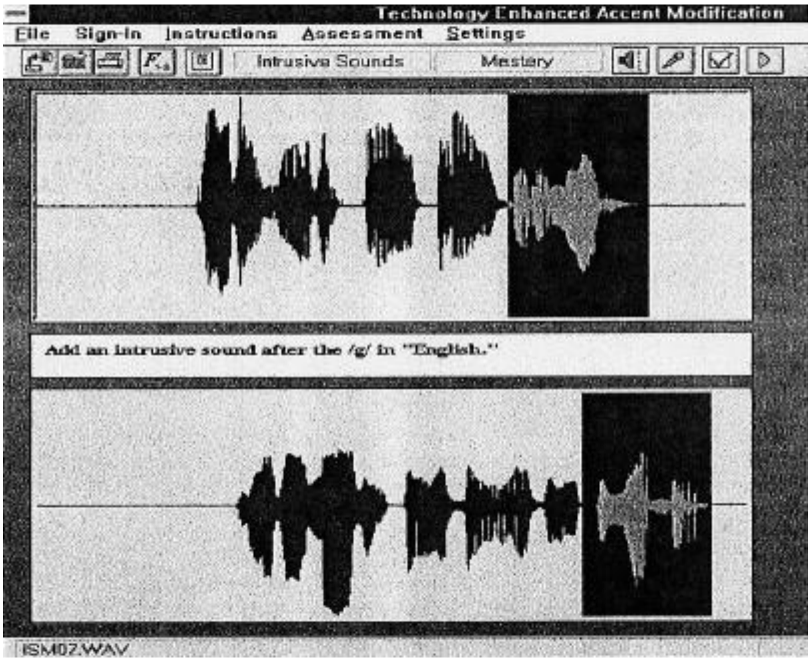


Note: In a more current setup, the functions of the 'EXPERIMENTAL CONTROLROOM' will most likely be carried out by a computer.

The study found that participants receiving feedback outperformed the control group, which only took a pre-test and a post-test, in terms of improving the use of pitch. Of the two groups receiving feedback, the group receiving AV training performed significantly better. However, de Bot commented that part of the explanation for the experimental groups performing better is probably that these groups received thorough instructions, including an introduction to intonation, prior to the beginning of the training sessions.

Focusing initially on training L2 suprasegmental phonology, Hardison (2002, 2004) found that using pitch displays helped English L1 speakers in their acquisition of French. Across 13 sessions of 40 minutes each, she had her participants produce French sentences while trying to imitate the NS pitch contour shown to them on a computer screen. In addition, the participants could see the pitch contour of their own speech and hear their own voice played back to them during training. Hardison's results showed that the participant not only improved in their prosody, but also that their speech samples were rated higher for segmental production. Additionally, it was shown that the training helped the participants recall sentences from long-term memory. In a similar vein, Hew and Ohki (2004) used pitch contours as well as waveforms as feedback in their study, and found improvements in their participants after less than an hour of training. However, the post-test was done immediately after the training session so it is very difficult to say what the effects in the long term were. Schwartz (1996) designed the *Technology Enhanced Accent Modification* software (TEAM) to improve the oral performance of international teaching assistants (ITAs) at his university. By today's standards it seems fairly basic as it was essentially a type of software that allowed learners to see, listen to, and repeat after an NS model as well as compare their own speech visually to that of the model (Figure 1.10).

Figure 1.10.
Screenshot from the TEAM software developed by Schwartz.



Note: Adapted from Kim (2012).

Despite its basic design, the software proved very effective. Students of ITAs who used the software rated their ITAs' linguistic ability more highly than did students of ITAs who had followed a traditional type of English instruction or speech therapy.

The TEAM software was later tested by Kim (2012), who conducted two case studies on Korean graduate students. In this study, Kim found that the software did bring about improvements in both learners, but made the interesting observation that the learner who improved more also reported having become more aware of his pronunciation, and as a result, felt more stressed when speaking English. She further argued that the software is more likely to be useful to visual learners as the participant who improved the most expressed a preference to a visual learning style. Considering the increasing amount of scrutiny that the learning styles theory has come under in recent years (Coffield, 2013), this is a bold claim which might warrant further investigation.

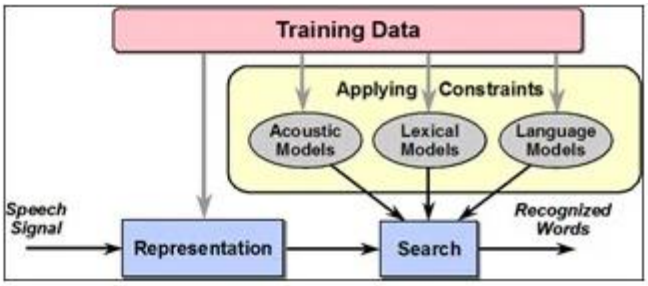
Finally, Ramírez Verdugo (2006) trained her participants by having them repeatedly record the same scripted dialogues over the course of 10 weeks. The experimental group (N = 10) could compare their recordings to that of native speakers and were also able to visually compare their pitch contours to that of the native speakers. The participants' recordings were both measured acoustically and rated by five native speakers of English. The findings from the study revealed that the trained group improved in terms of both measures compared to an untrained control group. Ramírez Verdugo further reports that her experimental participants became 'more aware of the role of intonation in spoken discourse' (p. 152). This is an important point in pronunciation teaching. Oftentimes, studies focus on drilling pronunciation rules which is repetitive and not very engaging for the learner (Hismanoglu, 2012; Amer and Amer, 2011). The type of technology mediated instruction used in the study by Ramírez Verdugo is much more likely to help the students understand how intonation is used in English. A possible caveat, however, could be training with the same dialogues in each training session could lead to some learners getting bored with the exercises. However, this might be mitigated by the fact that the learners can see their improvement from session to session in how closely their productions resemble those of the native speaker models.

1.5.2. Automatic Speech Recognition

One of the most exciting tools in CAPT is Automatic Speech Recognition. For anyone who has called a bank, utility provider, or any other larger corporation in recent years, ASR is nothing new. This is the technology that allows a computer to 'understand' what a person is saying and, ideally, respond to the user's request. In very basic terms, most ASR systems consist of an acoustic model, a pronunciation model (lexicon), a language model, and a decoder as illustrated in Figure 1.11.

Figure 1.11.

Schematic representation of an ASR system (Glass and Zue (2003).



The acoustic model tells the ASR what each phone sounds like. The pronunciation model is made by expert linguists and it tells the ASR system how each word should be pronounced, or rather what the sequence of phones should be. The language model's role is to tell the speech recogniser how probable a string of words is. Lastly, the decoder takes the information provided by the other three components and outputs the most probable string of words. These processes are usually handled through statistical modelling using what is known as the Hidden Markov Model (HMM) – sometimes alongside Deep Neural Networks (DNN) (Chan et al., 2015). More recently, speech recognisers have been proposed that are based on recurrent neural networks (RNNs) only. An RNN is essentially a DNN which is particularly good at predicting sequences (like speech input). Trained using orthographically transcribed speech data, the big difference between these and the more traditional recognisers is that they provide speech output based purely on the acoustic signal, thus leaving out the need for a pronunciation model and, with that, the painstaking work of labelling each individual word (ibid.).

ASR has already been used for second language testing allowing students, or even job applicants, to take a language test over the phone (Hincks, 2001; Townshend et al., 1998). It is easy to see how this could save a great deal of time for employers and university admission offices alike. However, it would be necessary to make sure that the test taker is indeed the person that he or she claims to be, as even high-stakes tests have seen participants cheat by having others sit their test (e.g., the Test Of English for International Communication (TOEIC) in the UK in 2014).

While not always 100% reliable, ASR usually gets it right when used in an L1 context (Juang and Rabiner, 2005), but the picture changes when considering an L2 context (Benzeguiba et al., 2007; Derwing, Munro, and Carbonaro, 2000). One of the issues is that to make sure the ASR system gives the right response, the learner's language has to be evaluated at the individual phone level – for systems targeting segmental phonology (Menzel et al., 2001). This requires a great deal of data in order to train the software, and this data is not always available. Even so, research has taken great strides in the past few decades to improve this technology (Levy, 2009), although there are still challenges ahead (Fouz-González, 2015; Goodwin, 2014).

It should be added that the reliability or accuracy of the ASR depends on the type being used. Some types of software are created to recognise individual words from a small vocabulary whereas others are required to recognise longer stretches of speech from a larger vocabulary. Needless to say, when the recogniser is asked to understand more words, the likelihood of mistakes increases.

In an L2 research context, ASR has been shown to be widely applicable for improving both segmental and suprasegmental aspects of L2 learners' language in a variety of languages e.g., Dutch (Cucchiarini, Neri and Strik, 2009), Swedish (Hincks, 2002a), Spanish (Precoda, Halverson, and Franco, 2000), Japanese (Kawai and Hirose, 2000), and English (Burleson, 2007; Walker et al., 2011; Witt and Young, 1997).

However, although ASR has been used successfully to train certain sounds and speech patterns of L2 learners, in order to be an effective tool for learning to speak a foreign language, the ASR system must meet certain standards such as being able to understand foreign accented speech (Dalby, Kewley-Port, and Sillings, 1998; Derwing, Munro, and Carbonaro, 2000) and rate the non-native speech productions the same way human listeners do (Eskenazi, 2009; Kim, 2006). Unfortunately, ASR has often failed to meet expectations when assessed against these criteria. Coniam (1999), for instance, tested the software *NaturallySpeaking* by Dragon Systems, asking his participants to read a passage of 1050 words, which was analysed and rated by the ASR system as well as human raters. He found that although ASR rating correlated with human ratings, they were much less accurate than the human listeners. These findings were echoed by Derwing et al. (2000), who also tested a software package from Dragon Systems, namely *Naturally Speaking Preferred*. They also found that the ASR system performed poorly when rating non-native speech, but do point out the important fact that the ASR package was not developed for second language learners.

A further criticism levied against ASR relates to a major concern that researchers have regarding its ability to provide feedback, which can at times be poor. Hincks (2002a), for instance, describes how one of her participants using *TellMeMore* noticed that all his productions of words beginning with /p/ were flagged up by the software, and yet the feedback had not helped him correct his pronunciation. As described above, not all types of feedback are equally easy for learners to understand. For this reason, some researchers have tried to gamify feedback in various ways. For example, Gómez et al. (2008) had their software show a dart board with the desired pronunciation as the bull's eye. Dalby and Kewley-Port (1999) used a bowling scenario in which the student's pronunciation score was reflected in the number of pins knocked over as illustrated in Figure 1.12.

Figure 1.12.

Gamified feedback used by Dalby and Kewley-Port (1998).

The obvious aspect lacking here is of course some sort of instruction as to how the learner can improve. Being told repeatedly that you are mispronouncing a word without any clues as to how or what to improve can be quite demotivating and of little help.

Derwing, Munro, and Carbonaro (2000) concluded that while ASR had potential, it still failed to perform to the same level as human listeners. Two decades have passed since their article was published, but even though improved ASR systems are continually being developed, we are yet to find anything near an ideal solution to the issues L2 learners face. As mentioned above, ASR works by comparing the input the learner provides to the models stored in the software. It can then tell the user whether the produced utterance was acceptable or not, but it is much harder for the software to accurately say which parts of the utterance need improvement. It goes without saying that using software that provides incorrect feedback – be it false positives or false negatives – can be disastrous for the L2 learner's learning experience. However, it is also important to recognise the importance of when to intervene, as it would be quite discouraging for the learner if every single non-standard pronunciation was flagged up by the software (Egan, 1999). Some students do occasionally request to be corrected every time they make a mistake. However, few – if any – teachers would actually do this, as excessive correction may discourage students. Similarly, ASR should not be programmed to correct every tiny phonological deviation. Finally, Derwing and Munro (2015) state 'off-the-shelf ASR technology is of little use in identifying production problems because it is unable to distinguish accent features that do and do not affect intelligibility' (p. 130). In their view, ASR – or indeed CALL as a whole – should only be used if it is based around the principle of improving intelligibility.

While these are issues often cited in the literature, Neri, Cucchiarini, and Strik (2003) state that much of the criticism is unfair. They emphasise the fact that testing dictation software for L1 use says nothing about what ASR can be used for in SLA. They further argue that much of the criticism directed towards ASR misses the point because it focuses on the questionable feedback that is often provided, rather than the speech recogniser itself. In their view, feedback should be seen as separate from ASR because the feedback is primarily a design issue, not an issue with the underlying ASR algorithms. Similarly, Wachowicz and Scott (1999) argue that the technology is already of a quality sufficient to be used for teaching purposes. They further point out that ASR was used successfully as early as 1982, but emphasise that it is important to use the technology appropriately for it to be effective. Undoubtedly, from a purely linguistic point of view, these arguments seem fair. After all, the technology promises no more than 'recognising' speech, and many speech recognisers are indeed getting better at recognising even accented speech. That said, for language teaching purposes, it could be argued that the two must be evaluated together. An ASR system has to be able to give good feedback, just like even native speakers of a language have to have a teaching qualification to know how to give adequate feedback. Only being able to tell that a learner's pronunciation does not match a desired model is insufficient for both human teachers and digital teachers.

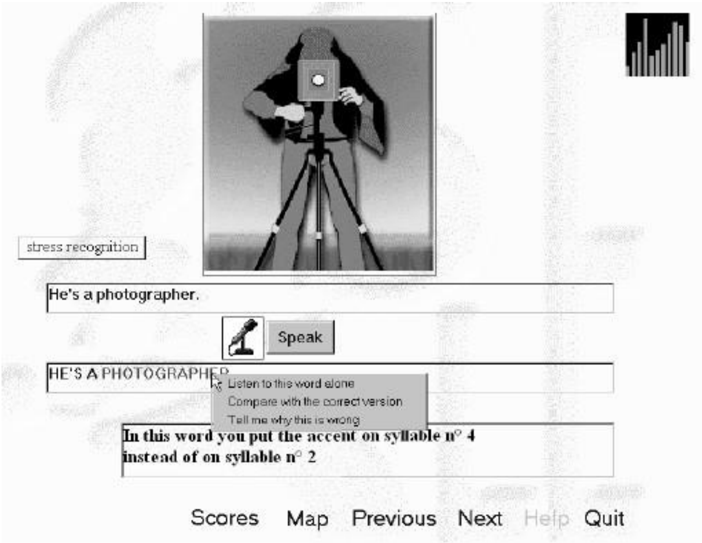
While the majority of studies seem to focus on segmental aspects of learner languages, some studies have explored the application of ASR in helping learners improve suprasegmental aspects of their L2 (Liakin, Cardoso, and Liakina, 2015). Some of the issues that have been raised in relation to the ASR for segmental learning, such as the problems with feedback, have also been applied to prosody e.g., Kaltenboeck (2001). The author argues that due to the limited performance of speech recognisers, it would be better to abandon the idea of the computer as a substitute for a human teacher, who can give concrete advice on speech production, and instead change the focus of CAPT to one that is based around perception and overall awareness of English speech production.

Somewhat counter-intuitively, many researchers working with ASR develop technology to be used in CALL, without actually testing its effectiveness on learners. As a case in point, Imoto et al. (2002) developed a speech recogniser that could detect stress patterns in the utterances of L1 Japanese learners of English. They reported that their software had an accuracy of 95.1% for native speakers and 84.1% when used to detect errors in English for L1 Japanese speakers. However, they did not investigate whether their software was able to improve the productions of Japanese learners of English. A similar study was conducted by Bang, Kang, and Lee (2013), who found that learners were generally positive about their experience with this type of learning. It has to be said that the number of participants in that study was quite low (N = 8), so more research into student responses to working with this type of software would be desirable. Obviously, a test of whether their software actually improves learners' speech production would also be interesting to see, particularly because their ASR system provides feedback on multiple levels (segments, rhythm, and phrase breaks).

With their 1999 paper, Herron et al. initiated what they called the ISLE project - Interactive Spoken Language Education (see Figure 1.13). This was a project that aimed at devising language learning software specifically aimed at native speakers of French and German (although the authors claimed it could be adapted to other languages as well). The authors stated that 'ISLE aims to create a natural learning environment in which the student is not responsible for self–diagnosis' (p. 855). The ISLE project was intended to yield a software that could diagnose both segmental and suprasegmental errors and provide accurate feedback to the learner.

Figure 1.13.

Interface of the prototype of ISLE – adapted from Herron et al. (1999).

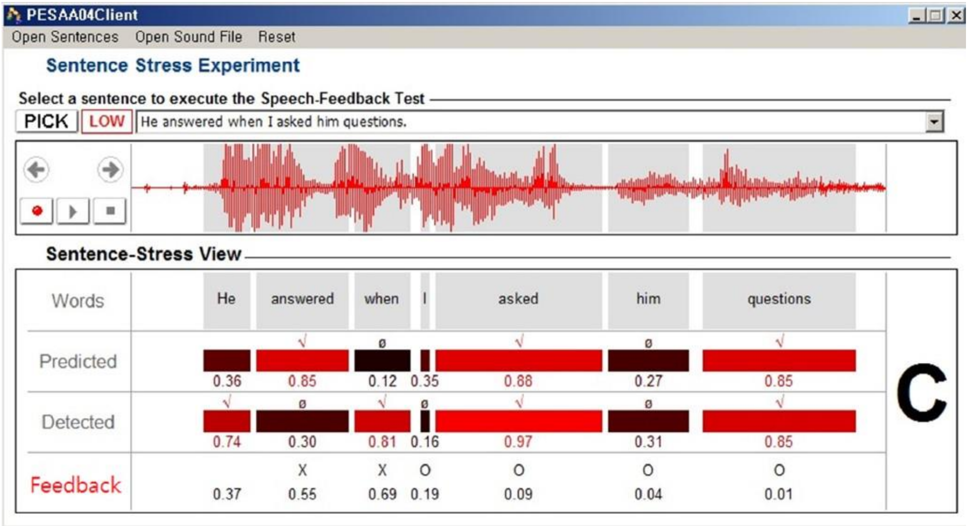


Despite their statement of intent, the project never really managed to meet expectations as the system provided an unacceptable number of false negatives (Menzel et al., 2001).

More recently, Lee et al. (2017) devised a speech recognition system to aid Korean L1 speakers improve their English prosody. As Korean is a syllable-timed language, they trained their speech recogniser to detect the incorrect use of stress in the learners' utterances and, similarly, indicate if stress had not been applied when it should have been. Figure 1.14 shows an example of the feedback provided to the learners. As the figure illustrates, the ASR compared a predicted stress pattern to a detected stress pattern. When the two matched, the learner received an '0' to indicate a correct production. However, if the predicted and the detected stress patterns differed, the learner received an 'X'.

Figure 1.14.

Visual feedback delivered by the ASR system developed by Lee et al. (2017).



The focus of their investigation was whether the learners improved their comprehensibility, accentedness, and rhythm, as judged by three native speakers of English. They found that the 20 participants in the experimental group showed significant improvement in terms of accentedness and rhythm, but not in terms of comprehensibility. The 20 participants in the control group, on the other hand, did not show any improvement at all. A minor drawback to their system is that it was 'designed to detect sentence stress errors in utterances of broad focus. Hence it is very likely that [the] system will judge sentence stress placed on a narrowly focused function word as an error' (Lee et al., 2017, p. 34). While this is of course unfortunate, it underlines very well the overarching problem in CALL, namely that it seems impossible to design one piece of software which can be used across the board.

Despite the recent advances in ASR, it is still an imperfect technology. As such, the sentiment expressed by Ehsani and Knodt (1998) is still valid, namely that what is needed are 'learning environments that exploit the strengths of speech technology while working around its limitations' (p. 54). However, it is encouraging that as the technology becomes more advanced, the limitations become fewer.

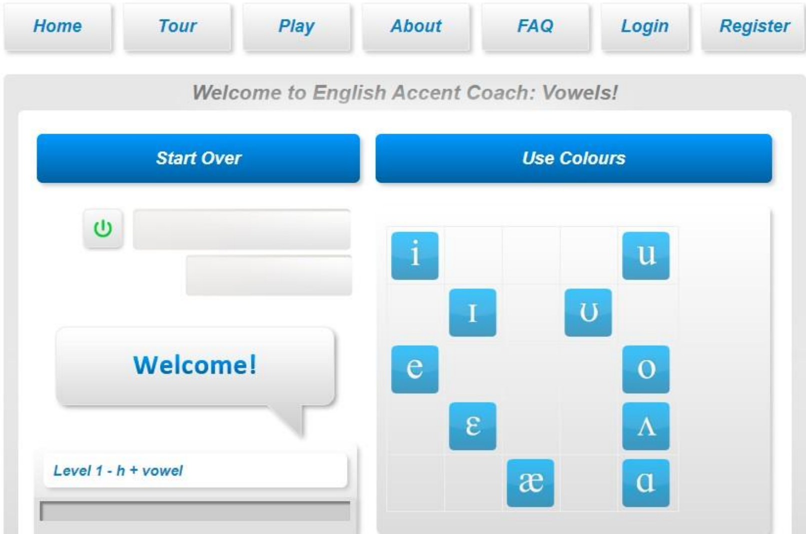
1.5.3. High Variability Phonetic Training

As mentioned in section (1.3.2), one of the issues learners face when learning an additional language in a foreign language context is the scarcity of native-speaker input to help learners form new phonetic categories in their phonetic inventory. A possible solution to this could be High Variability Phonetic Training (HVPT). Initially developed as a type of treatment in the speech and hearing sciences (Pierce, 2014), HVPT exposes the learner to large amounts of native-speaker input produced by several different speakers and in different phonetic contexts (Barriuso and Hayes-Harb, 2018). The input used can be modified in various ways, to attempt to enhance the effectiveness of the training, and some researchers have even used digital tokens rather than human ones (Strange and Dittmann, 1984). HVPT has for some time been viewed as an effective method for training non-native speakers to perceive and produce non-native speech sounds (Aliaga-García and Mora, 2009), and research suggests that both high and low proficiency learners can benefit from the training (Wong, 2015).

HVPT has some of the same advantages as the platforms in the present study in that training can be accessed for free – e.g., through EnglishAccentCoach.com (Thomson, 2012) – and whenever it is convenient to the learner. Figure 1.15 shows an example of a training session from EnglishAccentCoach.com. As can be seen, each vowel option is placed relative to one another in a manner that helps the learner understand where in the mouth the vowels are produced. This could help the learner internalise the production of the vowel.

Figure 1.15.

Example of a vowel training exercised from EnglishAccentCoach.com.



One drawback to this type of application is that it often assumes, that learners are familiar with the phonemic chart, which is not always the case. A possible alternative to phonetic symbols could be the use of keywords to represent the response options. A study by Fouz-González and Mompeán (2021) showed that keywords can indeed be used instead of phonetic symbols, but even though this might make the labels used in HVPT easier to use for learners and teachers alike, the use of phonetic symbols is still standard.

Despite the fact that IPA symbols may take some time to master, teachers should nonetheless urge their students to learn the phonemes of English, and students should try their best to learn them. This is because learning phonemes can be a great help when using a pronunciation dictionary (or any dictionary with added phonemic script). In addition, learning the IPA symbols may even help students better understand suprasegmental aspects of the language being learnt (Mompeán and Fouz-González, 2020).

HVPT usually presents learners with one of two task types, namely discrimination tasks or identification. The former can be divided into two subtypes known as AX and ABX. The AX discrimination task requires the listener to determine whether the two sounds presented are the same or different. In the ABX task the listener initially hears two sounds and has to decide whether the third sound presented is similar to the first or the second sound. It is, of course, possible to make variations of this task XAB or AXB. A final variation to this task can be referred to as the 'oddity discrimination test' (Flege and MacKay, 2004, p. 7) in which the learner hears three sounds and has to identify the one that is dissimilar to the other two. Identification tasks can be slightly more difficult because these tasks require the learner to determine exactly what sound they heard. Furthermore, in order to identify the sound, the learner has to choose from a range of orthographic representations of the sounds e.g., phonemes or other letter combinations, which can also make the task more difficult. In spite of these issues, however, studies show that training carried out using identification tasks is usually the more effective of the two (Carlet, 2017).

In a ground-breaking study marking the inception of HVPT training in the field of L2 speech learning, Logan, Lively, and Pisoni (1991) explored the hypothesis that large amounts of variable input can help direct learners' attention to the subtle differences between L2 speech sounds. The theory was that this type of training would help establish new phonetic categories in the learner's phonetic inventory. In their study, they found that all six participants (Japanese L1) improved their ability to perceive the distinction between American English /r/ and /l/. As will be described below, results corroborating their findings have since been obtained in several studies. However, other studies have failed to show an effect of HVPT (Aliaga-García, 2007) and yet others have yielded mixed results with improvements in some trained contrasts but not in others (Carlet and Cebrian, 2014).

It is clear that there are still elements of HVPT that need exploring. For example, Carlet (2019) suggests that the ideal training regime might depend on the learners' L1 and the phonetic context of the targeted phoneme. In addition, it would be beneficial to explore the effect of training. This should focus both on the number of sessions required as well as the optimal duration of each individual session. In doing so, research can ensure that sufficient training is provided at all items for all relevant contrasts.

What makes HVPT particularly interesting is that improving learners' speech perception seems to have an effect on their speech production, as suggested by Bradlow et al. (1997). As with the perception studies, the results have been promising with several studies showing that perceptual training can indeed help improve learners' production of L2 speech sounds (Bohn and Flege, 1990; Bradlow et al., 1997; Flege, 1987b). This should come as no surprise, though, as it is one of the main predictions of Flege's Speech Learning Model (Flege, 1995).

Although HVPT regimes have shown promising results, one has to keep the procedure of many of these studies in mind when judging HVPT as a viable avenue for classroom teaching or even self-guided study. What is common for many of these studies is the large amount of training that goes into obtaining the good results. To show how time-consuming pronunciation training can be, it is worth looking at some of the studies that have focused on training L2 learners through technology. For instance, the participants in a study by Lambacher and colleagues (Lambacher et al., 2005) followed six sessions of 20 minutes over six weeks to improve on the *cop-cup* distinction. Similarly, Akahane-Yamada et al. (1998) trained a native speaker of Japanese on the English /r - l/ contrast using HVPT and audio-visual feedback. The training was highly effective, but the total time spent was no less than five hours.

Although this amount of training may sound excessive, especially when comparing to most ESL pronunciation exercises, the training duration is not out of the ordinary for this type of study (e.g., Bradlow et al., 1997: 20 hours of training; Hardison, 2003: 11.5 hours of training). It should be obvious that a teacher cannot set aside five hours of teaching time to focus on one contrast – however important it might be. Seen from this perspective, although a highly exciting area of second language speech learning, the applicability of HVPT to the EFL classroom needs more exploring (Barriuso et al, 2017). In fairness, though, there are examples of training studies that have shown promising results with shorter training periods (Francis and Nusbaum, 2002: 90 minutes). It is possible that the amount of training required depends on the language feature that is being targeted, as research seems to suggest that some cues are easier to guide learners' attention to than others (Francis, Baldwin, and Nusbaum, 2000).

Whereas some of the studies mentioned above focused only on the acquisition of one phonemic contrast, other studies have explored working with several phonemic contrasts at once. Aliaga-García (2009), for example, found that HVPT training was highly effective for improving the perception of the 11 vowels included in the training. In fact, research suggests that HVPT training may be more effective if a full set of vowels is used as opposed to a subset (Nishi and Kewley-Port, 2007).

Interestingly, although high variability in the phonetic input has been shown to improve learners' perception, cases of too much of a good thing so to speak, have been found as well. As such, research seems to indicate that the high variability in the NNS input used to train NSs on foreign accent probably varies too much to be effective as training material (Jongman, Wade, and Sereno, 2003; Jongman and Wade, 2007). It is, of course, important to note the difference in pedagogical scope between these studies and the majority of HVPT studies, namely that the latter tried to teach native speakers to understand foreign accented speech. This is probably less common, but has nonetheless been advocated within the English as an International Language (EIL) framework as described in 1.4.1.

The extent to which training improves learners' perceptual and productive skills has been much more thoroughly investigated for segmental phonology than for suprasegmental phonology, but the latter has nevertheless received some attention over the past few decades. Considering the importance of suprasegmental phonology on both comprehensibility and intelligibility, this is potentially a very interesting field to investigate. Here, studies on tone and stress will be looked at in more detail as these tend to be the studies where perceptual training is used as a mode of instruction.

Using a HVPT training regime, Wang et al. (1999) attempted to help L1 English speakers perceive Mandarin tones. Using natural tokens akin to the ones used by Logan et al., their results showed that perceptual training can be effective for improving L1 English speakers' ability to perceive tonal contrasts. What is more, the researchers suggest that learners whose L1 is not a tonal language might be able to establish new tonal categories in order to perceive tone contrasts. This is believed to work in a similar manner to how L2 learners are theorised to establish phonetic categories for segmental contrasts (e.g., Flege, 1995). In a follow-up study, Wang, Jongman, and Sereno (2003) further showed that training suprasegmental perceptual skills can transfer to the production domain just as it has been found to be the case for segmental HVPT training.

Using identification tasks in an HVPT regime, Brawerman-Albini et al. (2017) trained Brazilian Portuguese speakers to identify and produce preantepenultimate stress, which occurs in English (e.g., *category)* but almost never in Brazilian Portuguese. In five relatively short sessions (76 stimuli per session), they showed that the trained group (N = 30) performed significantly better than the control group (N = 10). Additionally, the experimental participants' newly gained skills transferred to unfamiliar words and new speakers, and a delayed post-test revealed that the gains could still be measured two months after the initial post-test. As such, this study provides encouraging signs that HVPT can be used as a viable tool to teach L2 word stress. It would be interesting to see a similar study conducted on sentence stress, although this might be challenging, as sentences stress is theoretically more complex than word stress.

An important caveat to HVPT is that training benefits might be learner-dependent. In a study of American students' ability to learn to distinguish tonal contrast, Perrachione et al. (2011) showed that while a HVPT paradigm worked well for some of their participants, it was detrimental to others. In fact, some of the participants showed larger gains from a low variability training regime (positive findings from low variability training have also been reported by Wong (2012, 2014)). Hence, Perrachione et al. (2011) conclude that '[t]aking individual differences in speech- or language learning aptitude into consideration allows for the development of one or more training paradigm designs that will maximally benefit all learners' (p. 471). While this is ultimately the best possible scenario, it will probably be a while before schools and language academies will be able to offer this kind of test and treatment. Thus, there can be little doubt that more research into the effects and applicability of HVPT in language teaching is most certainly desirable, as in its current form, it is unlikely to be of much use in most EFL classrooms (Saito and Lyster, 2012).

1.5.4 CAPT outside the Classroom

There are many reasons why pronunciation training outside the classroom makes sense. Tanner and Landon (2009) claim that:

> [f]or language instructors who do not feel comfortable teaching pronunciation or who cannot fit it into their curriculum, self-directed, computer-assisted cued pronunciation readings can provide an effective way to help students improve their ability to perceive, predict, and produce prosodic features outside of class. (pp.61-62)

As seen above, using HVPT, students now have the possibility to work on segmental features as well as suprasegmental features without the presence of a language tutor. Furthermore, in monolingual settings, learners do not always feel the consequences of mispronouncing words because the listeners they work with share the erroneous pronunciation features; something which could lead to fossilization (Eskenazi, 1999b; Jenkins, 2000). As described in 1.3.2, learners with the same L1 may even deliberately stray from the standard pronunciation of a word in order to make themselves understood or simply to avoid breaking the socio-linguistic norm in the classroom (Walker, 2005). With the right technology, such as automatic speech recognition, learners can practise the parts of language they struggle with the most.

As opposed to the monolingual classroom, in a multilingual classroom, learners' difficulties with pronunciation may vary to such an extent that the teacher may find it difficult to meet every learner's need. In some cases, the disparity between learners may be so extreme that pronunciation instruction might not even be appropriate. This is particularly true in cases where the teacher has not received the proper training to provide adequate feedback (Fraser, 2001). In such a scenario too, CAPT could be an ideal solution because the teacher can specify areas of each student's pronunciation that need improving as part of the students' homework.

Commercial products for pronunciation teaching

Since the advent of multi-media and the CD-ROM, the language learning industry has experienced a significant boom. Whereas language learning outside the classroom used to be restricted to listening to and repeating after a cassette tape, as is common in the Pimsleur Approach, learners are now able to engage much more with the materials. As prime examples of these products are the likes of *Rosetta Stone* and *Tell Me More*, which became hugely popular from the mid-1990s. What most of these new interactive products had in common – and this is by and large still true of today's products – was that they promised the learner that their product would have them talk like a native speaker of the language being learned. This promise was to a large degree based on the relatively new developments in ASR. As seen above, however, in many cases the companies promised more than their technology could vouch for.

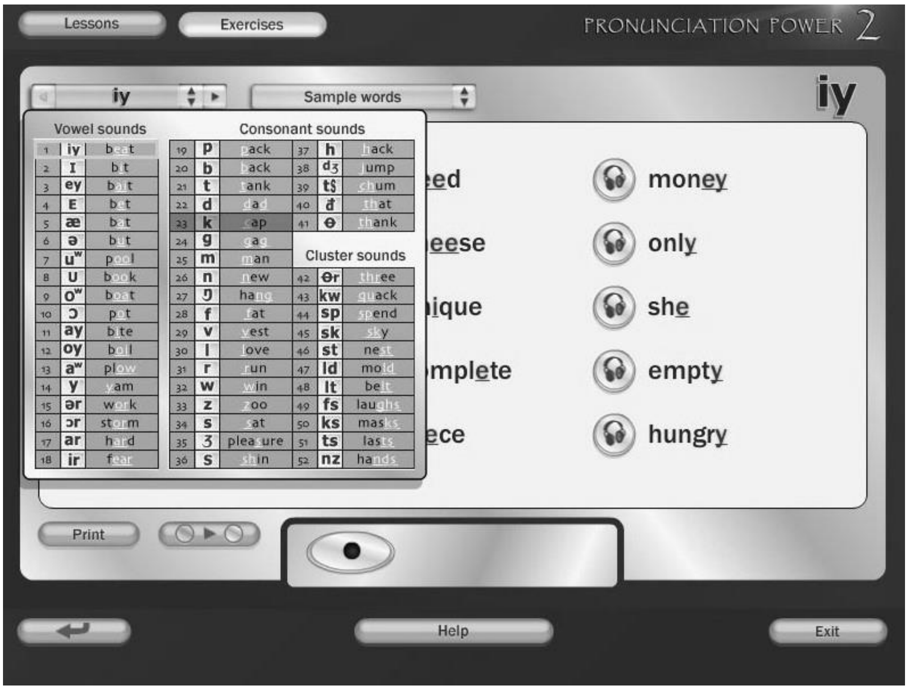
It is not only the technology behind these products that has come under scrutiny. Many researchers have criticised commercial language learning packages for thinking too much about design and too little about language learning theory (Heil et al., 2016; Murray and Barnes, 1998; Neri et al., 2002; Pennington, 1999) leading to fancy looking products that sell well but may not help the learners as much as they promise. Moreover, the focus of these language learning solutions is usually limited to vowels and consonants. As Derwing and Rossiter (2002) state '[c]omputer software packages designed for pronunciation training material, in particular, although replete with attractive graphics, are largely based on segmental approaches' (p. 157). This approach runs contrary to the current belief that a global approach to pronunciation teaching yields better results (Derwing, Munro, and Wiebe, 1998).

Nevertheless, it is not impossible that working with commercial software can lead to improvements in terms of language level and even in terms of pronunciation specifically. As a case in point, Elimat and AbuSeileek (2014) used *Tell Me More*'s online tool *Performance English* to improve the pronunciation of three groups of 9-year-old Jordanian school children (N = 38). Compared to the control group (N = 16), who received normal classroom-based instruction, the ASR groups performed significantly better on the post-test for all tested aspects (individual words, sentences, and dialogues). As part of the training, the teacher assisted with troubleshooting and answered questions from the students regarding the materials, so it does not tell us much about ASR technological for autonomous use. However, it does illustrate that using ASR-based application can be useful and, indeed, that commercial products can be used successfully.

In a similar vein, Seferoglu (2005) tested the effects of the off-the-shelf product *Pronunciation Power* on a group of English language teacher trainees. *Pronunciation Power* affords a range exercises and examples, and also includes very rudimentary phonetics explained using both stills and animations as Figure 1.16 exemplifies.

Figure 1.16.

Example of an exercise from *Pronunciation Power.*



The exercises range from practice with individual sounds, such as minimal pair drills and discrimination exercises, to whole-sentence practice. The researcher conducting the study and a colleague rated the students' speech productions on the use of both segmental and suprasegmental features produced during 10-minute presentations (both pre- and post-test). She found that students who used the software (N = 20) improved more than students who received regular classroom-based pronunciation teaching (N = 20) after just three weeks of practice. However, Liu (2008) tried to replicate the study, but failed to find any significant improvement in her participants after six weeks. One aspect lacking from this study is an inquiry into the students' experience of using the software, which is unfortunate as the students trained with the software during class time only. It is thus difficult to say whether the students would have used it, if they were working in an out-of-class context.

Considering other results of commercial language learning software studies such as Nielson (2011), who found that the majority of participants in her study failed to complete their training using *Rosetta Stone*, this is a very important question, as CAPT will only be effective if used.

Somewhat contradictory to this claim, Hincks (2002a) reported some rather peculiar findings in her study of adult engineers using *Talk To Me* by Auralog to learn English. Apart from taking part in the study, the students also followed a 10-week technical English course and had regular meetings with a pronunciation tutor. In spite of this, she did not report any statistically significant improvements in her participants' pronunciation. Conversely, counter to what one might have expected, Hincks reported that the students who improved the most were the ones who had used the software the least. It should be added that this coincided with the students having the strongest accent at the beginning of the study, so the findings could show that the software has a greater potential for learners who struggle with pronunciation, or that these students benefited more from meeting with the pronunciation tutor.

Mobile-assisted language learning

Most of the applications described above can be used either on a computer or mobile device. In the latter case, applications and the affordances they provide are referred to collectively as Mobile Assisted language Learning (MALL). It can be seen as a branch of CALL, but as the name implies, has the unique characteristic that learning takes place through portable devices (Kukulska-Humle and Shield, 2008). When MALL was in its infancy, it experienced significant teething problems due to inadequacies in the underlying technology, but despite these issues, student perceptions regarding the new learning opportunities were generally positive (Chinnery, 2006; Thornton and Houser, 2005).

The earliest mobile applications were limited to dictionaries, flash card games, and basic phrase books (Godwin-Jones, 2011) – something which led Burston (2014) to call for more effort to be put into the use/design of mobile applications as he pointed out that 'as more recent and innovative implementations attest, substantial curricular integration of MALL is in fact possible' (p. 115).

As noted above, with the development of more powerful mobile devices, it is now possible for language learners to access high-quality L2 input and language learning materials any time of day. Furthermore, as mobile plans offer much larger amounts of data than they used to, and as free Wi-Fi networks have become standard in educational institutions and cafes alike, students no longer have to worry about incurring costs by taking advantage of this increased access. Thus, students can watch videos, listen to podcasts, or use other web applications at their convenience, as long as they are in possession of a smart phone. In addition, should they wish to engage with native speakers of their L2, they can do so equally easily on social media (Mitchell, 2012).

The access to mobile devices may also have the added benefit that teachers and tutors can reach their students whenever they wish and vice versa – something which can potentially increase motivation and lead to better engagement with the curriculum (Dunlap and Lowenthal, 2009). While this might at first seem like a new and exciting idea, one has to consider long term whether teachers and tutors should be expected to be available outside of school hours (Grosseck and Holotescu, 2008). In fact, implementing technology that increases the time tutors are available to their students seems a disfavour to educators, as technology should help free up time rather than take it away. After all, even if students cannot receive help from their teachers late in the day, they will still be able to consult their peers through Facebook, WhatsApp, Twitter or similar platforms.

Similarly to the other technologies mentioned above, MALL applications have been used in a wide range of studies. In a study of Turkish ESL learners, Saran, Seferoglu, and Cagiltay (2009) compared three types of instruction: handouts, websites and multimedia messages (MMS), and found that the MMS group used the materials more, and saw larger gains in their pronunciation than the other two groups. Interesting as this may be, this type of messages may already seem a bit dated and could thus alienate some learners. Luckily, the language learning software companies mentioned above have all realised that a great deal of learning now happens outside traditional educational settings, and have created mobile applications in response to this development, just as smaller independent app developers have tried their luck, as a quick search on *Google*'s or *Apple*'s app stores will show.

An increasingly important player in the market is *Duolingo*, an American language-learning website (duolingo.com) and mobile app. *Duolingo* offers both a free version as well as a premium version, which can be accessed for a monthly fee. Similar to the subscription-only products like *Rosetta Stone*, it consists of individual units that require the learner to complete various tasks. Figure 1.17 shows part of the user interface in *Duolingo*. As can be seen, the learner is taken through a step-wise progression of units that are increasing in difficulty as the learner improves.

Figure 1.17.

Screenshot showing part of the *Duolingo* interface.

It also uses ASR to help the learner with their pronunciation. Several studies have looked at the potential for the use of *Duolingo* both in-class and as self-study, but very few have actually measured whether students improve as a result of using the application, which is surprising considering its popularity and, with that, its potential. One of the few studies that have focused on the effectiveness of *Duolingo* for L2 speech learning is Martinelli (2016). In the five case studies she did, she found that although working with the software for three weeks helped her participants improve the features under investigation (Italian geminates and voice onset time), these improvements did not translate to improved comprehensibility as judged by NS raters. However, considering the fact that the intervention period was only three weeks, it might have been a little optimistic on the researcher's part to hope for improvements in comprehensibility.

Another dedicated language learning platform which should be mentioned is *Busuu* (busuu.com). Similar to *Rosetta Stone* and *Duolingo, Busuu* comprises both a website and mobile app, and also offers either a limited free version or a premium subscription version. Additionally, *Busuu* taps into some of the features offered by social networks in the sense that it allows learners to find other people to practise with. As an extra, useful feature, *Busuu* provides learners with the opportunity to book live lessons.

Only a handful of studies have investigated *Busuu* for language learning. However, these studies have generally reported good results for both effectiveness and user evaluations. For example, Kétyi (2013) asked 59 Hungarian learners of German to rate the premium version of the app after a week's use. A total 79% of the learners rated the app either 'good' or 'very good'. However, 92% of the learners also stated that they would not want to pay for the premium version after the end of the free trial. The main complaints from the participants were that the exercises were either too easy or too repetitive. Drawing on the experience from this pilot study, Kétyi (2015) investigated the effectiveness of the app as well as the users' evaluation. He found that the experimental group using the app made significantly bigger improvements than the control group, which did not use the app. As in the pilot study, users reported that they had found the app useful. A potentially important finding was made by Liu et al. (2013), who looked at the use of *Busuu* in the classroom. The authors found that more than half of their participants reported to also having used the app outside the classroom. It goes without saying that if teachers can implement technology into their teaching which learners are happy to use autonomously, both teachers and learners win.

Looking again outside the field of materials developed by larger corporations, Foote and McDonough (2017) saw the comprehensibility and fluency of their 16 participants improve after doing shadowing exercises using iPods for eight weeks; although their degree of accentedness remained unchanged. In this study, the learning device in this case were iPods, but it is important to note that the materials themselves could be used with other mobile devices too. The participants even had the option to access the materials on YouTube if they preferred to do so. The materials for the shadowing exercises were all taken from popular TV shows, and there is no reason why students should not be able to use any spoken text for practising shadowing. As mentioned above, the results obtained in this study were positive, but it is worth pointing out that most of the participants were enrolled in an ESL course and all were exposed to native speakers of English on a daily basis, which may have influenced the results. Even so, shadowing is certainly a type of practice that needs more investigating due to its affordability and convenience, as it requires very limited instructions on the part of the teacher.

As a final application to be mentioned here is Twitter. Unlike the applications mentioned so far, Twitter was not developed as a language learning tool, but rather as a microblog for users to give their views on whatever was on their mind. Twitter has received a fair deal of attention in the field of CALL in recent years, and for good reason, as it shares an array of advantages with other popular social networking sites (SNSs). It is free, popular, and easy to access from any device. Moreover, with its character limit, it forces its users to be concise. As a platform where people tend to share – often unfiltered – opinions, it can certainly be used to start a conversation in the classroom and help keeping students engaged (Junco, Heiberger, and Loken, 2010). It has even been suggested that using Twitter in the classroom can help students improve higher-order skills such as critical thinking (Hattem and Lomicka, 2016). At the same time, however, teachers should be cautious about what type of threads they include in their lessons due to the content some users post.

Most studies on the use of Twitter in education have largely ignored the platform as a tool for learning pronunciation. Rather, studies have focused on using Twitter for developing writing skills, community building, and interaction with native speakers (Hattem and Lomicka, 2016). However, it has been argued that Twitter is also well suited as an instrument for improving L2 learners' pronunciation through explicit instruction (e.g., Fouz-González and Mompeán, 2012). That this is indeed the case was further indicated by (Mompeán and Fouz-González, 2016) in which 16 learners of English received daily tweets showing them how to pronounce difficult English words. The participants showed a good rate of engagement in the sense that they confirmed reading the tweets, and they appeared to have benefited from the instruction, as the mean post-test scores increased by 74.9%. The authors do acknowledge that further studies into the effectiveness of Twitter for pronunciation instruction are needed as their study did not include a control group.

Chatbots and Virtual Tutors

A chatbot is a kind of computer software that the user can interact with using text or speech. The computer's ability to 'understand' speech has been made possible by the development in automatic speech recognition (Levy, 2009). They have been suggested as a solution for students who find it difficult to have access to native speaker interlocutors (Fryer and Carpenter, 2006). Considering the number of websites dedicated to language learning nowadays, it is difficult to see how finding native speakers could pose a problem to the keen language learner, but some might prefer the non-judgmental chatbot to an actual human being.

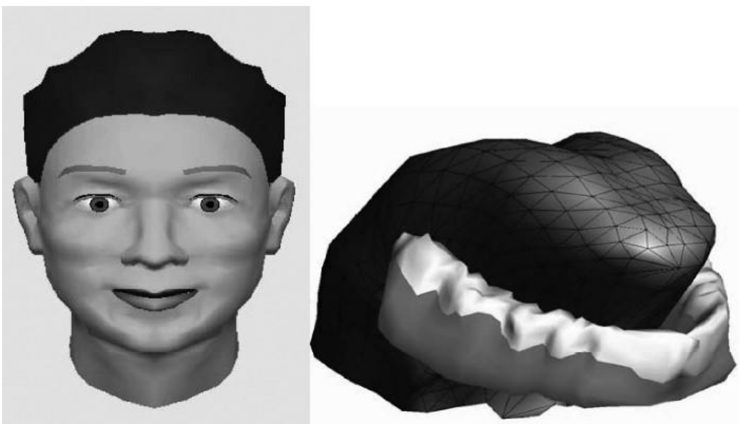
The ideal chatbot should be able to mimic a real conversation to the extent that the learner cannot tell if they are chatting to a computer or a person. However, although the technology has been around for quite some time now, chatbots still have a long way to go before performing to a satisfactory level (Gallacher, Thompson, and Howarth, 2018). In terms of efficacy, though, for some pronunciation features, chatbots may prove to be as useful as human instructors (Cardoso, 2018). A key issue in the usage of chatbots for language learning is the software's ability to understand learner language; that is to say, language that is in some way at odds with linguistic or phonological rules. For pronunciation training specifically, the main problem is creating speech recognition software that understands learners' mispronunciations.

An extension of the chatbot is the virtual tutor, which is a system the learner can interact with, but which is also able to correct errors. Whereas Gick et al. (2008) worked directly with their participants in order to help them interpret the visual feedback and improve their pronunciation, Engwall (2008) attempted to let a virtual tutor provide the feedback to make it even more accessible to the learner. His setup was, however, somewhat incomplete from a technological point of view as the ASR was replaced by a trained phonetician who told the computer what feedback to provide.

There can be little doubt that virtual tutors offer an array of potential advantages. For example, 3D virtual talking heads can show learners where in the oral tract a speech sound is articulated – something which may be close to impossible for a human instructor (Wik and Hjalmarsson, 2009).

Figure. 1.18.

(ARTUR) Talking face with underlying 3D structure of the tongue and lower jaw.



Note: Model developed by Engwall et al. (2006).

A couple of projects have investigated the use of these virtual tutors such as the BALDI project (Massaro, 2003; Massaro and Light, 2003) and ARTUR (Engwall et al., 2006). The BALDI project was initially developed to aid people with hearing difficulties, but have since been tested for L2 learning as well (Massaro and Light, 2003). Interestingly, despite the originally intended use of the visuals, Wik and Engwall (2008) state that using a visual model of the vocal tract that allows learners to see tongue movement works better for pronunciation training than for supporting people with hearing loss.

So far, work on 3D talking heads seems to suggests that this type of visual support can indeed benefit the learner, at least with certain speech sounds (see Dey, 2012; Massaro et al., 2008). In some cases, however, it is not clear whether 2D or 3D visuals are to be preferred. As a case in point, Kröger, Graf-Bortscheller, and Lowit (2008) examined children's ability to mimic speech sounds produced by 2D and 3D articulatory models and found that 2D models might be easier to use, although their results were not statistically significant. At any rate, the studies conducted in this area show – perhaps unsurprisingly – that learners need to be trained to use images of internal articulators in order to benefit from training (Wik and Engwall, 2008). What might be surprising is that talking heads have been used as part of a CAPT method for teaching stress (Hincks, 2002b). Although stress cannot be observed directly by viewing the vocal organs, Hincks instructed the participants in her study in how to make their virtual tutor nod when producing a stressed syllable. This was only a small part of how her participants worked with the materials, and it is difficult to say how much of a difference this addition made. The study shows, however, that there is no reason why virtual tutors cannot be used in the teaching of suprasegmental aspects of an L2.

As a final note on virtual tutors, it has been argued that students who use animated agents tend to spend more time with the material due to the 'persona effect' (Lester et al., 1997), a supposed positive effect of having an animated agent as part of the software's interface. This effect has been questioned, though, as van Mulken, André, and Müller (1998) found the effect to apply only to certain aspects. Even though their experimental participants reported to have enjoyed the course more than the control group, who did not use a virtual instructor, in terms of learning, they did not find any difference between the two groups. Miksatko, Kipp, and Kipp (2012) found no effect at all. It might be worth noticing that the average age of the participants in the study by Lester et al. (1997) was 12 years, whereas in the other two studies, the average age was 28 and 26 years, respectively. Also, the studies investigated the persona effect, but not directly in relation to second language acquisition. This suggests that there is a possibility that the persona effect is age-dependent, although this needs further investigation to be confirmed. It is also apparent that more research is needed on the persona effect in SLA. So far, all that can be said is that the idea that virtual tutors are somehow more appealing to learners is most certainly an intriguing one.

To sum up, although chatbots and virtual tutors offer great potential in terms of innovative teaching methods and materials, it seems clear that much research and development is still needed before these tools can be used as a stand-alone solution for pronunciation teaching.

Web 2.0 tools

There can be little doubt that the constant advances made in technology give teachers new avenues to explore when trying to improve students' pronunciation. Much has happened since the time when a cassette recorder was the only tool available for practising pronunciation, and the use of technology could provide teachers with what they need to teach pronunciation more effectively in the classroom. As an added bonus, the development of web 2.0 tools has provided a host of options for students to learn languages outside the classroom, which as a result, creates the opportunity to introduce blended learning i.e., a learning form that combines classroom teaching and online learning (Ducate, Lomicka, and Lord, 2012; Johnson and Marsh, 2014). As many teachers find that they lack sufficient time to focus extensively on pronunciation in their lessons, these tools could prove highly beneficial for both students and teachers.

The distinction between web 1.0 and 2.0 is usually made on the basis that whereas web 1.0 technology only offers static websites containing information which can be read or downloaded by the user, web 2.0 is seen as interactive in the sense that web content can be accessed and created or adapted by the users (O'Reilly, 2005). Web 2.0 technologies thus offer users the possibility to contribute to a lesser or larger extent as seen with free online encyclopaedias, created and edited by users around the world, such as Wikipedia (wikipedia.org).

In the past decade or so, researchers have started to look at the possible applications of web 2.0 tools in pronunciation teaching. This seems but a timely change as pronunciation has received far less attention than other areas of language learning as shown by Wang and Vásquez (2012). In their review of 43 studies using web 2.0 tools, they identified just one that focused on pronunciation, namely Lord (2008).

Podcasts

One of the most popular applications currently being investigated for its L2 learning potential is podcasts (Ducate and Lomicka, 2009; Lord, 2008). The word 'podcast' is a portmanteau of the words 'iPod' and 'broadcast', and is in its essence an audio-file that can be shared online and sent regularly to subscribers (Levy, 2009; Rosell-Aguilar, 2007). Through podcasts, learners can gain access to vast amounts of native and non-native speaker input (Warschauer and Liaw, 2011). What is more, because of the sheer number of podcasts available, chances are that learners will be able to find podcasts exploring a topic they find interesting and want to learn more about. Learners who are learning on their own might run the risk of engaging with materials that are too advanced for their level and, as a result, lose motivation. However, many websites list podcasts according to learner levels, so in most cases, this should not be an issue. Thus, podcasts can be suggested as extra-curricular activities or even be used as part of the lessons, the latter of which is likely to boost motivation among students and increase learning (O'Bryan and Hegelheimer, 2007).

While it could be argued that podcasts are little more than easily available radio shows, they are included here because of the way they have been used in teaching and research. Rather than just being used for listening activities, podcasts have been used for a range of purposes such as speaking activities (Ducate and Lomicka, 2009; Lord, 2008) and specific pronunciation practice (Fouz-González, 2019). Additionally, because podcasts are distributed through really simple syndication (RSS), they have an element of automaticity that normal online radio broadcasts lack (Rosell-Aguilar, 2009). Finally, the fact that students can quite easily produce their own podcasts means that they can contribute content, which is one of the defining features of web 2.0 (O'Reilly, 2005).

As a case in point, Lord (2008), had her students produce podcasts in calls throughout a semester of Spanish. Each podcast focused on a different element of Spanish pronunciation. Unfortunately, due to the way the study was designed, it did not yield any information about whether the participants improved as a result of working with the podcasts. However, the study did show that the participants' attitudes toward the use of this technology improved as a result of the study.

A study that did look at the effectiveness of podcasts was Ducate and Lomicka (2009). This study explored whether podcasts could be used to improve the comprehensibility and intelligibility of L1 English learners of French and German respectively. Throughout a semester (16 weeks), American undergraduate students worked with various types of podcasts, but did not receive direct pronunciation instruction. The study did not find any significant improvement in terms of comprehensibility, but the authors argue that this could be due to a ceiling effect, as most of the participants turned out to be perfectly intelligible at the beginning of the course. With regard to autonomous learning of pronunciation outside the classroom, they suggest that:

> such independent study does not seem sufficient. If teachers hope that students' pronunciation will improve as a result of outside practice with CDs, MP3s, or podcasts, it may require more focused and consistent pronunciation practice in class or meetings outside of class with an NS in addition to the assigned tasks. (p. 77)

This may be an overly negative assessment, though, as podcasts can be used for shadowing exercises, which have been shown to improve some aspects of oral proficiency. It might not be that particular attention has to be given to the materials in class or with a private tutor, but rather that the focus or expectations of the podcasts need changing. Alternatively, teachers could provide their students with specific guides as to how to use podcasts for autonomous learning as done by Foote and McDonough (2017).

Social networking sites

Social networking sites (SNSs) have become hugely popular in recent years with sites like Facebook (facebook.com) and Weibo (weibo.com) having hundreds of millions of users logging in every month. With such a degree of popularity, it only makes sense to ask how these networks can be used for educational purposes. Different approaches as to how to make use of the SNSs can be taken: the SNS can be used as part of the lesson in the classroom (e.g., Blattner and Fiori, 2011; Blattner and Lomicka, 2012) or the learners can be asked to use the SNS in their own time (e.g., Ota, 2011). Finally, it would be possible to use a combination of the two (Brick, 2011). Brick further suggests making available some of the SNS features through the learning management system Moodle in order to control how much time students spend on learning language. As this is strictly speaking not a social network, it will not be considered here, but the idea seems to be worth exploring.

The first approach (i.e., using SNSs in the classroom) is convenient in many ways since it allows the teacher to better control how much time students spend on the tasks and to help students if they get stuck with a particular question. However, having the learners use the technology in their own time seems much more appealing as it frees up valuable class time for other purposes. Furthermore, even if students struggle with the tasks, being on a social network allows them to get help from their classmates.

In a perception and usability study focusing on language learning websites such as Babbel (babbel.com) and LiveMocha (livemochas.com), Stevenson and Liu (2010) surveyed 164 users to explore what made the learners use this particular type of online tool and what they liked and disliked about them. Interestingly, 25 users mentioned that the chat page and discussions boards were the least useful elements on these sites. This is somewhat surprising given that these features are the ones that provide the sites with social dimension. However, the surveyed users were complaining that other members used the sites as more of a dating site or as a place to establish new relationships, both of which they found inappropriate. This could be seen as another advantage that teacher-administered social network groups can have, as these would be limited to the more secure environment of a closed group. This view was repeated in the usability section of the study in which the authors had five learners test the sites and give their opinions regarding pros and cons. The testers further reported that they would like some sort of linguistic scaffolding to help them communicate with other users, as talking to a stranger without sufficient help could make them feel uncomfortable. Once again, this is something a teacher would be able to provide as part of the task-setting. Once the learners feel more secure about their linguistic ability, they can explore other groups in which they can interact with native speakers or advanced learners of their L2.

While social media offer a large number of potential benefits and possibilities for language teachers and learners, they also present some challenges. At a very basic level, it has been questioned whether the students necessarily know how to use the technology (McBride, 2009), just as privacy issues have been mentioned as a potential stumbling block (Wang et al., 2012). While these challenges for the most part can be overcome by the teacher giving thorough instructions about user settings and use of the technology, this could lead to valuable classroom time being wasted on explaining to the students how to engage with the materials – something which would be rather counterproductive (Busà, 2008). It has even been argued that issues such as these have kept educators from trying to integrate social media into their teaching (Bowers-Campbell, 2008). However, this might be an overly pessimistic view, which seems to underestimate the average language learner's tech skills.

Certainly, the more popular applications require minimal instruction. Furthermore, once the students have learnt how to use the technology, it becomes an extension of the classroom that allows for additional learning (Johnson and Marsh, 2014). Finally, the sheer number of publications on SNSs as a learning tool shows that '[r]esearchers are moving from the question of whether or not to use SNSs in language learning to the question of which ones to use and how' (Mompeán and Fouz-González, 2016, p. 166).

Facebook

It is fairly uncontroversial that language learning happens through a combination of language exposure (input) and language use (output). As such, Facebook offers its users the chance to improve their language simply through engaging with others on the platform (Mitchell, 2012).

The claim that Facebook lends itself to educational purposes is supported by the wealth of Facebook pages and groups dedicated to various type of instruction, with more than 70 groups dedicated to English pronunciation alone (accurate as of September, 2021). It should be mentioned that several of these groups have only one member – the creator – however, 18 groups have over 100 members, and a few over 1000. One thing that can be said against these groups is that they require learners to be very persistent and actively ask for advice from other group members, as simply waiting for the right content to be posted will not get you far. As an example, the group Advanced English has more than 90 thousand members with new content uploaded daily. However, the content is often just individual words or expressions with no added context or encouragement to use the materials so unless the learner has an identic memory, chances are that most of what is uploaded will be forgotten by the end of the day.

One of the first large-scale (N = 909) attempts at judging the value of Facebook as an educational tool was Selwyn (2009). He looked at what type of posts appeared on students' Facebook walls (i.e., their person profile) and concluded that 'SNSs such as Facebook do not merit any particular laudation from educators, neither do they present any cause for moral panic' (p. 172). Whilst this statement leaves much to be desired in terms of embracing SNSs as a teaching tool, it is important to notice that only the wall feature of Facebook was looked at. Because today's Facebook has developed significantly from what it used to be, it is plausible that a more positive conclusion could be reached.

Despite the popularity of Facebook, very limited research has been published regarding its effectiveness as a teaching tool (AbuSa'aleek, 2015). Instead, many studies have chosen to focus primarily on student perceptions of Facebook (e.g., Eren, 2012; Kabilan, Ahmad, and Abidin, 2010), and only to a lesser extent on whether the students actually improve (Manca and Ranieri, 2013), and the ones that do generally focus on aspects of language learning other than pronunciation. As such, Barrot (2018) surveyed 41 academic articles focusing on Facebook as a language learning tool. Of these, only one study looked at oral productive skills (speaking). Thus, to obtain an idea of students' perception of Facebook in an educational context, it is necessary to look outside the field of L2 speech learning as such studies – to the author's knowledge – have not yet been published. The studies that have been conducted have generally made some encouraging findings for participants whose primary aim is to learn English, as well as for students learning English as part of their non-language related studies (but see Madge et al. (2009) for an opposing view).

In another large-scale study, Kabilan et al. (2010) surveyed 300 randomly chosen students at a university in Malaysia to investigate their views on Facebook's contribution to their language development. The students were not enrolled in a language programme, but nonetheless stated that they generally found that Facebook helped them improve their English and increased their motivation to engage with content in English. Focusing on students in Saudi Arabia, AbuSa'aleek (2015) replicated this study using a slightly modified version of Kabilan et al.'s questionnaire. His findings were to a large extent similar to that of the original study.

Yunus and Salehi (2012) carried out a study involving 43 of their students studying for a bachelor's degree in TESL to see how they felt about using Facebook as a tool to improve their composition skills. The authors pointed out that '[t]he findings showed that 'Facebook groups' is an effective tool in improving the students' writing skills, especially in the brainstorming of ideas before the actual writing' (p. 87). As their study only focused on students' perceptions of using Facebook groups as part of the writing process, this may be slightly too optimistic. However, their participants did report almost unanimously (97%) that they felt encouraged when other students liked their posts. In a similar vein, Ghani (2015) interviewed 35 students on a technical communication course after introducing Facebook as a communication tool as part of the teaching, and concluded that Facebook has real potential for both writing and discussion activities. Karal, Kokoc, and Kacir (2017) specifically explored the use of Facebook groups as a tool to improve the written language (L1) of Turkish teenagers and found – as was the case in the SLA studies – that their students responded well to using Facebook for educational purposes. The study thus found that the participants improved their writing skills, and that teachers-student communication improved as well. Although this last study does not contribute to our knowledge about the use of Facebook in the area of SLA, it does support the view that Facebook seems like a popular tool in terms of improving writing skills.

Although Facebook has been shown to increase student motivation and student involvement in English written exercises, it is important to keep in mind that most of the studies discussed above have been short-term studies. Some studies have made the important point that the good feedback received in the studies of student perceptions could be a result of the novelty effect (Blattner and Lomicka, 2012). This was also indicated by Buga et al. (2014), who reported that the involvement of their students initially improved, but then declined after the first month. This may show that the novelty effect is something one should always look out for when studying the effects of new technologies. If the novelty indeed is an issue in these cases, it could indicate that Facebook is better suited for short courses than as an actual course management site, although more research is needed to establish this. This, however, should not be seen as a reason not to use Facebook or similar technologies. It simply means that they should be used with caution and only be used for what they are good for.

A rather unusual – and yet intriguing – way of increasing student motivation through Facebook was tested by Saylag (2013), who found that sharing personal information about herself (Teachers' Self-Disclosure) in a Facebook group dedicated to the class made her students feel more closely connected to her. Although the researcher found that her students were very positive about the teaching method, she does acknowledge that this way of teaching is a balancing act requiring the teacher to judge carefully what sort of information should be shared. This echoes Mazer, Murphy, and Simonds (2007), who found that while students generally appreciated their teachers being friendly, teachers did run the risk of being perceived as less professional.

Rather than sharing personal information on Facebook, it has been suggested that teachers post videos of themselves producing various speech sounds on their private Facebook page (wall) for students to watch (Eckhart, 2009). While the intentions behind this idea are good, it seems more appropriate to post this type of material in a group. Also, it must be kept in mind that many speech sounds cannot be properly shown without animations.

On the downside, some studies have pointed to potential issues when using Facebook as a learning tool. Arzu (2014) found that some of her participants completely ignored the tasks posted in the group. This is of course unfortunate, but not far off what can happen in the classroom, where 100% student participation is rare. Similarly, Shih (2011), in spite of the overall positive results of her study, reported that her students sometimes failed to complete their writing assignments on Facebook, because they were too engrossed in other online activities.

As argued above, using new technology in or outside the classroom comes with the potential drawback that teachers need to show students how to use it. However, this drawback becomes much less severe when using popular tools such as Facebook because most students will already know how the technology is used. Furthermore, what often characterises popular applications is that they, for the most part, have a very intuitive interface, which should make it easier and quicker for instructors to show learners how to navigate the technology.

YouTube

As is the case with Facebook, YouTube makes available large quantities of input that learners can use to help them learn an L2. In fact, many of the videos on Facebook could be considered YouTube content as they are, in many cases, shown on this platform. Although users can communicate with one another in the comments section below the videos, user interaction is somewhat limited compared to Facebook, with its easy-to-use text and video chat.

Even so, YouTube has been shown to be a useful tool for incidental vocabulary learning (Arndt and Woore, 2018). This is not surprising as any web activity is likely to lead to at least some incidental vocabulary learning. Indeed, Arndt and Woore (2018) found no statistical difference between participants who watched videos and participants who accessed the same input through blog posts. Furthermore, the fact that users can comment on most videos can lead to negotiation of meaning between users (Benson, 2015), which can lead to language development (Lightbown and Spada, 2006).

It seems uncontroversial to suggest that YouTube can also be a resource for cases of more direct learning. YouTube is a platform where instructors upload their own lessons so learners can find information on most aspects of the English language from basic grammar and pronunciation to more advanced semantics and pragmatics. One of the strengths of YouTube is undoubtedly that it allows for a great deal of learner autonomy (Watkins and Wilkins, 2011). For independent learners, especially the more advanced ones, YouTube can be a very useful tool because of the vast amounts of language varieties and speaking styles – ranging from formal speeches to slang – that can be accessed. With regard to language variety, Smith (2011) argues that YouTube has the advantage that it makes available several possible language models, which helps the learner choose the accent they would like to learn, and further argues that the teacher should help the learner pursue this accent by finding more online material the learner can work with. While this might seem like a sensible idea on the surface, one cannot help but wonder how much this would affect teachers' preparation time if every single student in a class wants to focus on a different model.

A potential downside to YouTube – which probably applies to most SNSs when it comes to language learning – is that, since anyone can upload videos to YouTube, there is no guarantee that the quality of the content is adequate. This should not be an issue if YouTube is used as part of an online course or in a blended learning context, as the teacher would be able to guide students to the most appropriate material. However, contrary to the suggestions by Smith (2011) above, in this case the teacher should be able to control what content is used, despite the fact that this might decrease the level of learner autonomy to some degree. Hafner and Miller (2011) see no issue with a drop in learner autonomy at all as they argue that learner autonomy should be defined much broader than simply as independent learning. In their view, learner autonomy can be fostered in the classroom as long as the approach is student-centred. Indeed, their study focusing on STEM students in Hong Kong taking a module on English for science and technology showed that working with YouTube can increase learners' motivation and encourage learner autonomy. It is interesting to note that in this study, YouTube was used in a different way from many other studies because the students themselves created and uploaded the content to a YouTube channel. In other studies, on the other hand, participants were generally asked to absorb instructional content created by others. This is, in many ways, similar to studies done on podcasts where students, rather than simply listening to content, created their own content and shared it with peers (e.g., Fouz-González, 2019). These studies show the importance of thinking innovatively when using technology in an educational context.

Unlike the research carried out on Facebook as a learning tool, which has so far only focused on learner perceptions, the effectiveness of YouTube as a learning tool has been the focus of a few studies. For example, using YouTube to deliver content to EFL teacher trainees at a university in Turkey, Alwehaibi (2015) showed that YouTube can be an effective tool for blended learning. His experimental participants (N = 51) watched videos dealing with the topic of the upcoming lesson and had its main points pointed out to them in class, before watching the video a second time and discussing the content. A control group (N = 45) also consisting of EFL teacher trainees attended lectures on the same topics as this was the standard mode of delivery at the university. The post-test results showed that while both groups improved their knowledge about the course content, the experimental group performed better than the control group. It is important to keep in mind that it is most likely not YouTube as such that leads to learning, but the way the learners engage with the material. In order to show that the platform itself had an effect, one would have to compare the effectiveness of YouTube to, for example, Vimeo or a similar video sharing platform. To my knowledge, this had not yet been done.

In a study exploring YouTube as a language learning tool for Turkish EFL trainee teachers, Hismanoglu (2012) looked at the effectiveness of using video lessons freely available on YouTube in terms of improving the students' command of English word stress. The performance of the experimental group was compared to that of a control group, which received classroom-based lessons. Post-test results showed that the experimental group outperformed the control group in all seven categories tested, but the difference only reached statistical significance in three of the seven categories. While this could be taken to suggest, at first glance, that video lessons are better than traditional classroom teaching, it should be pointed out that the content of the classroom lessons differed in some potentially important respects from the video content. In both contexts, the students were asked to repeat examples after the tutor, but the video lessons additionally provided the students with more detailed information about what a stressed syllable is and worked with syllable counting. What is more, in the videos, many of the words taught were presented in context rather than as individual lexical items, thus giving the experimental group a distinct advantage, as context seems to positively affect the learning of pronunciation (Hirata,

2004).

In summary, as shown in the discussion above, it becomes evident that while there are many promising and interesting technologies available today that can support teachers in many respects, most of those tools still vary significantly with regard to their applicability for language teachers. Teachers in a monolingual classroom have a distinct advantage because many of the tools and applications available target specific speech sounds. These applications may still work in a multilingual classroom, but they will require a bit more preparation on the part of the teacher, because they will have to look into what types of speech sounds their students struggle with.

Although many of the applications currently available have shown promise, there is still a need for implementation of better language pedagogy; especially when it comes to the development of language learning software. A major problem which has yet to be solved in many cases is that visual feedback tools are often very difficult to use. This means that teachers would have to give up precious instruction time to show students how to interpret the feedback they receive. It is important for both teachers and learners to realise that the effectiveness of a piece of software or an app cannot be judged simply on how it looks. This is made very clear from the studies on HVPT, which have proven effective for both segmental and suprasegmental features, despite the very basic interface learners often work with. This type of training is in stark contrast to many ASR-supported systems which look flashy, but often fail to deliver due to poor quality of either the recogniser itself, the system's ability to provide feedback, or in some cases both. That said, ASR is still a hugely interesting and promising part of the CAPT, and undoubtedly something that will attract many researchers in years to come.

Looking outside the field of dedicated language learning products, web 2.0

tools of various kinds need to be investigated further. These tools allow language learners to receive large quantities of input in both text and speech, just as they give learners the opportunity to produce language – be it written or spoken. A possible issue with these is that the content is often individual expressions or rules that do little to engage the learner. However, this issue can be solved if a teacher is involved to add structure to the learning experience.

Chapter 2 – Word and Sentence Stress in English

This section will introduce the characteristics of English word and sentence stress patterns as well as discuss some of the theoretical approaches to the two phenomena. Although it is beyond the scope of this project to give an exhaustive account of all different theories in the field, a certain degree of familiarity with the differences between English and Spanish will prove helpful in order to predict what aspects of the English language could potentially prove difficult for L1 Spanish learners of English. Therefore, references to the literature on stress in Spanish will be made where appropriate.

As will be mentioned shortly, the term 'stress' has been defined in many different ways in the literature. Here, however, a simple distinction will be made between 'word stress' (also referred to as 'lexical stress') and 'sentence stress' (also referred to as 'prosodic stress'). Word stress will be used to describe a feature that makes a syllable more prominent than others in a given lexical item in citation form (Gutiérrez-Díez, 2005) regardless of how the prominence is achieved. Sentence stress can involve both 'rhythmic stress' and 'nuclear stress', and its domain is connected speech (Mompeán, 2014). According to Goodwin (2014), the term sentence stress only refers to the rhythmic stress of a sentence whereas 'tonicity' (see 2.3.1) should be used to refer to the identification of the most prominent syllable in an utterance. However, as this thesis does not deal with rhythmic stress in any great detail, the term sentence stress and tonicity will be used interchangeably. Although this may sound like an over-simplification, the view is consistent with how it is treated by Cutler (2005), who states that 'stress is the accentuation of syllables within a word, or of words within sentences' (p. 264). Furthermore, this use of the terms is also in line with many teaching materials, and thus makes sense in a context where teaching of pronunciation to learners of English is the focus.

Finally, as for the term ‘primary stress’, this is often used interchangeably with nuclear stress – especially among American authors. However, it is also common to use it in relation to word stress. Hence, if used here, it will refer to word stress and, thus, to the most prominent syllable within a word.

2.1. Stress Fundamentals

Of all the terms in phonetics and phonology, ‘stress’ is probably one of the most ambiguous and most debated (Ladd, 1983) as it is a term that has been applied to different concepts in phonetics and linguistics. Sweet (1906), for example, equated stress with loudness. Other authors have described stress as prominence achieved through intensity and duration (or intensity and fundamental frequency (F0)), and stress has even been used to refer the ‘potential for accent on utterance’ (Cruttenden, 2014, p. 25). This is aligned with the description provided by Bolinger (1986), who says that:

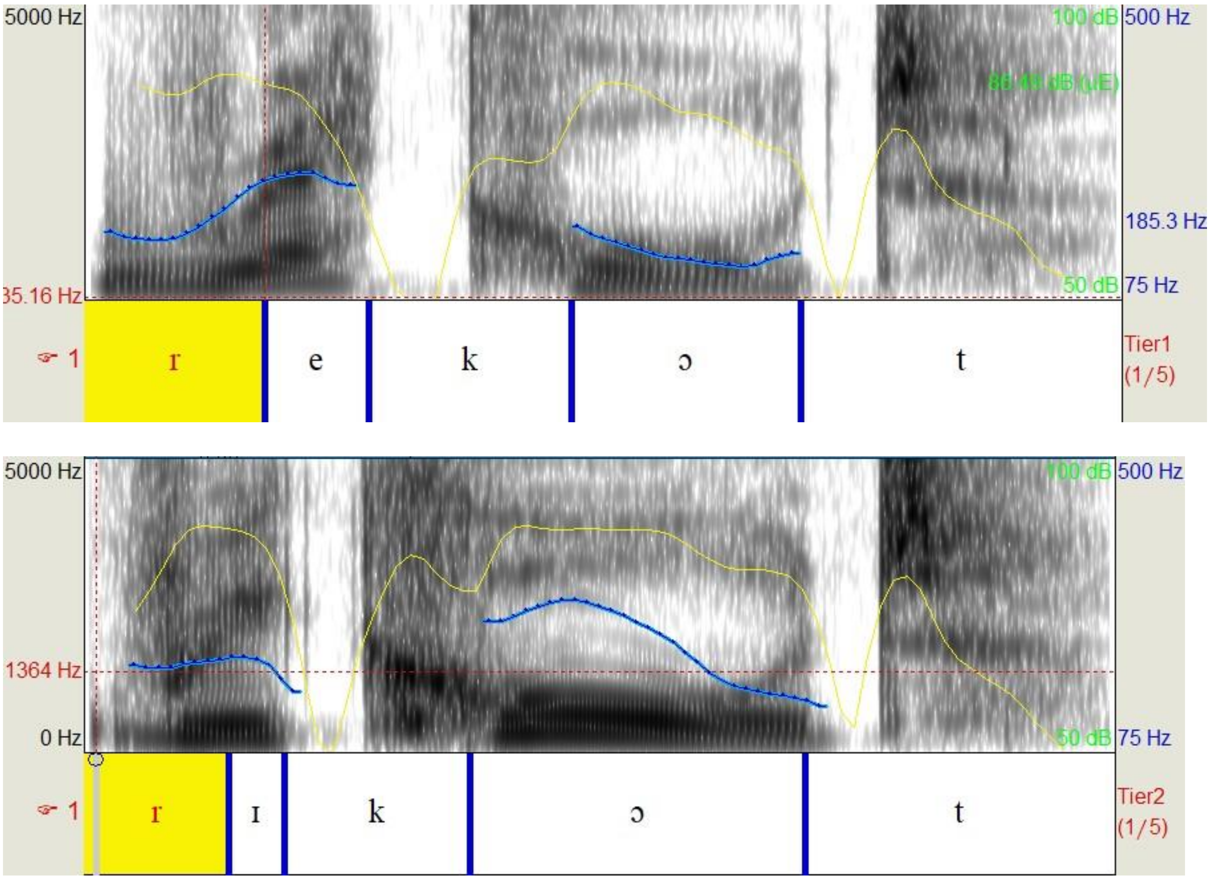
> it is misleading to think of the stressed syllable of a word as something that is regularly more emphatic than the other syllables. Rather, that syllable is the one that will get the special emphasis whenever the word is emphasized. (p.14)

Hence, Bolinger says that in the word that is emphasised in a sentence, the stressed syllable should be referred to as having accent. The remaining words have stress as per their dictionary definition.

When talking about stress, one has to consider both what is produced by the speaker as well as what is perceived by the listener, as these – although similar – are not identical (Roach, 2009). From a production point of view, stress can be said to be realised through an increase of air expelled when producing a stressed syllable (Ladefoged and Johnson, 2010). However, this might be an over-simplification as, according to several researchers, stress seems to be produced through a combination of intensity, duration, and fundamental frequency, and vowel quality (García Lecumberri, 1995). A change in F0 is brought about by a change in vibration in the vocal cords with faster movement resulting in a higher pitch and slower movement resulting in a lower pitch (Ladefoged and Johnson, 2010). Duration is rarely defined, possibly because it is fairly intuitive.

From a perception point of view, intensity, duration, and F0 are perceived by the listener as loudness, length, and pitch respectively (Cutler, 2005). There has been a great deal of debate among phoneticians about what acoustic cues are the most important for the perception of stress in English – and to some extent in other languages – but there is still no definitive answer to what the most reliable cue to stress perception is (Hincks, 2012). Fry (1958) concluded that pitch was the most important cue, with duration also playing a part, whereas intensity – much against what many might expect – plays only a minor role if any. In fact, Bolinger (1958) showed that increasing intensity could have the opposite effect in that syllables modified to have increased intensity were judged as less stressed than unaltered ones. Interestingly, studies have shown that the change in F0 does not have to be an increase to make a syllable be perceived as stressed. Morton and Jassem (1965) found a positive effect of lowering F0 (although this effect was lesser than that of raising F0). The three cues (intensity, F0, and duration) are illustrated in Figure 2.1 below. The figure shows two spectrograms of *record* (n.) and *record* (v.) produced by a native speaker of English. If comparing the first syllable in the noun (top) to the first syllable in the verb (bottom), it can be seen that this has a higher intensity across the syllable (although the peak appears to be similar to the verb form). Furthermore, the variation in F0 is more pronounced in the first syllable of the noun than in the verb. Finally, the vowel has a slightly longer duration.

Figure 2.1.
Spectrograms depicting the production of the noun *record* (top) and the verb *record* (bottom) by a native speaker.



Note: The use of /t/ in the final syllable is probably not standard, but it is also not important for the analysis provided here. The recordings were both taken from oxfordlearnersdictionaries.com.

Based on his findings, Fry also suggested that vowel quality plays a part in the perception of stress in English, as full vowels are more likely to be perceived as stressed than reduced ones. This can, in some cases, present a problem for L1 speakers of Spanish learning English because Spanish does not make use of vowel reduction (Hualde, 2007). Hence, L1 Spanish speakers may accidentally mis-stress a word because they fail to reduce a vowel correctly (Flege and Bohn, 1989). It is now considered uncontroversial that, in English, stressed vowels are full vowels, but that full vowels need not be stressed.

Disregarding the case of vowel quality, the same properties of speech are in use when talking about Spanish stress perception. Investigating research findings on stress perception in Spanish, which suggested that only F0 is important for signalling stress in Spanish, Llisterri et al. (2003) tested the perception of 30 native speakers of Spanish to see what cues played the biggest part in perception. Using resynthesised tokens of both Spanish words and non-words, the authors found that while no single cue is responsible for the perception of stress, the combination of F0 and duration as well as F0 and intensity both contributed to stress perception – as did a combination of all three.

These findings were echoed by Ortega-Llebaria, Gu, and Fan (2013) who stated that research has now provided:

> a cumulative body of evidence that a cluster of non-independent acoustic correlates, such as duration and intensity, cue word stress [...], and that these correlates are in a trading relationship with the complex f0 patterns that constitute the intonation of a sentence. (p. 187)

Thus, what has become clear from the paragraphs above is that stress cannot be described as a single structural feature that applies across the board in all languages. Rather, stress is a feature that is produced and perceived through a variety of cues that differ between languages (Beckman and Edwards, 1994).

2.2. Word Stress

Languages can be categorised according to how they use stress lexically. Some languages do not make use of lexical stress at all. French is often categorised as one such language, although some debate remains as to whether it should in fact be categorised as having fixed stress (Armstrong, 1932). There can be little doubt, however, that the use of stress in French is markedly different from, for example, English and Spanish. Later in this chapter, references will be made to 'tone languages'. These are languages that change in pitch pattern to differentiate lexical meaning (Ashby and Maidment, 2005). It should be added that languages that use both stress and tone do exist, but they are generally not very well researched, and 'continue to pose descriptive and theoretical challenges' (Michael, 2010, p. 57).

In languages that do have lexical stress, the stress is either fixed on one syllable or varies freely. Examples of the former are Hungarian, where the stress falls on the first syllable of a word; Polish, where the penultimate syllable receives the stress, and languages like Turkish that stress the last syllable of a word (Ashby and Maidment, 2005). There are of course exceptions, and there may be morphological features that play a role in where a word is stressed, but the vast majority of words in these languages follow the given stress patterns. In variables stress languages, we find languages such as English and Spanish. In these languages, stress can be used contrastively. This means that stress alone can distinguish lexical meaning. Table 2.1 below presents a few examples of this from Spanish and English.

Table 2.1

Examples of variable lexical stress in English and Spanish.

English		Spanish	
Word	*Transcription*	*Word*	*Transcription*
insight (n.)	/ˈɪnsaɪt/	limpio (adj.)	ˈlimpjo/
incite (v.)	/ɪnˈsaɪt/	limpió (v.)	limˈpjo/
refund (n.)	/ˈriːfʌnd/	camino (n)	/kaˈmino/
refund (v.)	/rɪˈfʌnd/	caminó (v.)	/kamiˈno/

It has been argued by several authors (e.g., Cruttenden, 2014) that the stress pattern is not necessarily the only thing that changes in English, as many word pairs also display a change in vowel quality in the sense that the unstressed vowel is reduced. This can be seen in the case of *refund* in Table 2.1. As such, Xu (2019) talks about 'near minimal noun-verb pairs' (p. 322). However, as can be seen from the *insight* – *incite* word pair, this change is not always a given, at least not in citation form.

The extent to which non-native speakers are able to perceive and produce stress has been investigated in the field of psycholinguistics. As is the case with segmental features of speech, the mental representations of stress in a speaker's L1 can heavily influence how the speaker perceives and produces stress in an L2. One model claims that speakers who do not use stress in their first language will experience 'stress deafness' (Peperkamp and Dupoux, 2002). Another model takes a purely phonological approach and posits that language typology can be made to make predictions regarding both L2 speech perception and speech production (Altman, 2006). These models will both be described in more detail in 2.2.2 below.

2.2.1. Learning L2 Stress Perception and Production

Word stress plays an important role in word identification and consequently in speech processing in English. In terms of word identification, Cutler and Clifton (1984) found that 'mis-stressing a word hinders its recognition just as would mispronunciation of part of its segmental structure' (p. 194). Interestingly, their study further found that the participants appeared to be less troubled by the stress moving from right to left compared to the stress shifting in the other direction. That is, their participants were generally happy to accept a construction such as *to* followed by a word with a strong-weak stress pattern (e.g., *to CONduct*) as grammatical. This is clearly something that has relevance for the ESL classroom, as learners need to be aware of the potential issues they will encounter in terms of intelligibility loss if their word stress is not correct. This view is also shared by Dalton and Seidlhofer (1994), who argue that lexical stress is easier to teach than some other suprasegmental features (e.g., tone) and can play a more important role than some phonemes in terms of contribution to the general intelligibility of speech. Additionally, it is possible that an improvement in the use of lexical stress could positively influence the learner's speech rhythm.

As regards speech processing, it is interesting to consider the fact that native speakers of English are not necessarily as proficient at making use of word stress as speakers of other variable stress (or free-stress languages). Cutler and Pasveer (2006) argue that this can, in part, be explained by the heavy reliance on vowel reduction in English. The authors compared possible instances of embedding (a word within a word such as *sea* inside *secret*) between Spanish, Dutch, German, and English. Their study found that whereas English speakers benefit only to a very limited extent from considering stress when making a decision on lexical items, speakers of the other languages significantly reduce the number of possible words in their mental lexicon when parsing a sentence.

In addition to its importance with regard to word recognition, vowel reduction plays an important part in giving English its characteristic rhythm. Part of the reason is that many short function words like *of*, *and*, *has*, and *have* are reduced in natural speech. Because of the importance of reduced vowel sounds in English, Gómez Lacabex and García Lecumberri (2010) looked at two ways to improve L1 Spanish speakers' production of reduced vowels in English. One group focused primarily on perception training, whereas the other group focused primarily on production training. After three sessions of approximately 30 minutes, the participants took a post-test in which they were asked to read sentences out loud as well as imitate words that were presented to them through headphones. Their data showed that in the imitation task both groups improved, and although the perception-based group performed better, there was no statistically significant difference between the two groups. For the read-out-loud task, however, only the group receiving perception-based training showed statistically significant improvement. These results underline the importance of exposing learners to quality input.

Interestingly, NSs of English may not necessarily outperform NNSs of English when it comes to identifying stress in English. García Lecumberri (2006) investigated the ability of native speakers and non-native speakers to identify stress in simple words and compounds in two conditions: with a citation form stress pattern and with the stress pattern shifted. Although the NSs performed slightly better with regard to stress shift in simple words, there was no statistically significant difference between the two groups when looking at stress identification in compounds. It must be added that the L1 Spanish participants were students of English Studies, which will most likely have given them an advantage over the average Spanish EFL learner.

This idea is supported by the fact that experience – as measured by length of stay in an NS environment – has been associated with increased accuracy in English (L2) lexical stress production (Trofimovich & Baker, 2006). These authors state that:

> learning appears to be driven by linguistic experience and is likely to depend on the particular segmental or suprasegmental aspect being studied, suggesting that L2 speech-learning theories (e.g., Flege, 1995) can be extended to account for the processing and learning of both L2 segmentals and suprasegmentals. (p. 26)

What is particularly important in this regard is that both García Lecumberri (2006) and Trofimovich and Baker (2006) show that at least some suprasegmental features can be learnt through experience, and the possibility of learning these exists in both an ESL and an EFL context.

As one might expect, it is not only learners of English who may struggle to learn L2 stress patterns. Studying the perception of Spanish lexical stress by native speakers of English, Romanelli, Menegotto, and Smyth (2015) compared two groups of adult English speakers learning Spanish as an L2 with a group of L1 Spanish speakers. The training of one group of learners focused on the identification of Spanish lexical stress, while the other group was trained to identify contrasts between Spanish consonant sounds. After three weeks of training (10 minutes, three times a week), the group that had received training on lexical stress perceived stress contrasts in Spanish at equal levels to native Spanish speakers. These results are extremely encouraging for English learners of Spanish considering the relatively short training period. However, it should be added that training specifically targeting suprasegmentals may be necessary, as simply following a general language course might not be sufficient. This was shown by Romanelli et al. (2015), who tested the ability of L1 English speakers to perceive Spanish word stress. Upon completing a 90-hour Spanish course stretching over three weeks, and despite performing as well as native speakers with regard to penultimate stress, the participants still struggled to perceive final stress in Spanish.

Studies have also been carried out on speakers whose L1 does not use lexical stress. Carpenter (2015) thus focused on the effects of phonetic training on L1 French speakers' ability to perceive lexical stress. The training was done using a fading technique and exaggerated nonsense stimuli. Four groups were involved in the study, namely a trained and an untrained group of L1 French speakers as well as a trained and an untrained group of English speakers. The two L1 English groups were only included to control for effects of exposure to the stimuli and showed no statistically significant improvement between tests. Thus, of the four groups involved in the study, only the trained group of L1 French speakers improved after training.

A similar training study was carried out on Taiwanese speakers by Ou (2011). This study is interesting because, although the participants improved in one of the two trained categories (lexical stress with a rising tone), the performance of participants in the experimental group actually deteriorated in the other category (recognising stressed syllables with falling intonation). The author states that the reason for this is likely to be that the participants taught themselves to rely too heavily on vowel duration, thus classifying words like *import* (n.) /'import/ as stressed on the second syllable. This study inadvertently shows the importance of having a teacher involved in the process of learning L2 phonology.

2.2.2. Stress Perception and Production Models

Considering the existing models on segmental speech perception and production, it is not unreasonable to assume that the properties of a learner's L1 may influence his or her ability to learn suprasegmental properties of an L2. Comparatively less work has gone into developing models for the acquisition of L2 suprasegmentals, but two models are worth mentioning. These are the Stress Deafness Model (SDM) and the Stress Typology Model (STM).

The Stress Deafness Model

As mentioned, much work has gone into investigating L2 learners' ability to perceive foreign speech sounds. Research in the field has generally shown that if an L2 uses a phonemic distinction of similar sounds that is not found in the learner's L1, this distinction will be very difficult for the learn to perceive - and often produce (Flege, 1995). Interestingly, there seems to exist a similar issue with regard to contrastive lexical stress in the sense that L1 speakers of languages that do not use stress phonemically (e.g., French) may struggle to learn this aspect of languages such as English, Spanish, and Dutch.

Through a series of studies, Dupoux and colleagues proposed the Stress Deafness Model (SDM) to account for the problems some speakers have when trying to perceive lexical stress in an L2. In one study, Dupoux et al. (1997) compared the ability of L1 Spanish and L1 French speakers to perceive stress differences in nonsense words produced by Dutch speakers. In an ABX discrimination task (see 1.5.2), Spanish speakers significantly outperformed the French speakers, thus suggesting that L1 French speakers are unable to perceive stress differences at a lexical level. However, in another experiment (part of the same study), the authors simplified the task by using an AX discrimination task as well as reducing the phonetic variation by using only one speaker, and found that the L1 French speakers perceived stress differences quite accurately (an error rate of 3.2%). The authors suggested that rather than perceptual shortcomings in L1 French speakers, their results could show that L1 French speakers have not got an abstract meta-linguistic representation of lexical stress to help them store these differences in short-term memory. This hypothesis was discarded, however, after a follow-up study a decade later by Dupoux et al. (2008). The researchers compared the ability of L1 French learners of Spanish who had been specifically taught about stress to L1 speakers of French with no experience of Spanish learning and found that they performed equally poorly in the tasks. Regarding their earlier findings, they argued that the good results on the simplified perception tasks were 'due to the fact that in these experiments, participants could use an acoustic representation rather than an abstract phonological one' (p. 19). It can be theorised that, because French does make use of some prosodic features (e.g., to mark prosodic boundaries), if the task presented to the L2 learners is sufficiently simple, they might be able to rely on the acoustics in the signal to help them in the perception tasks. However, when the task is made more complicated either by using ABX tasks instead of AX tasks or by using more than one talker to deliver the items, this strategy no longer works, and the participants' stress deafness is revealed.

Most of the studies on stress deafness have been conducted using L1 French speakers as participants. However, Peperkamp, Vendelin, and Dupoux (2010) included L1 speakers of Polish, Finnish, and Hungarian as well because these three languages have predictable word stress, and thus use word stress in a manner different from English and Spanish. They found that speakers of Hungarian and Finnish displayed a similar degree of stress deafness as the French participants, but surprisingly, the Polish participants fared better, although still significantly worse than L1 speakers of Spanish. Since there are more exceptions to the predictable word stress in Polish than in the other languages in the study, the authors theorised L1 Polish speakers have a slight advantage over speakers of Finnish and French, because they occasionally encounter words with a non-standard word pattern. Thus, the SDM predicts a hierarchy of learners' problems as shown in Table 2.2.

Table 2.2.
Four-level hierarchy of difficulty in stress perception predicted by the SDM.

Class	Language	Condition	Prediction
I	French and Finish	Not exposed to variable lexical stress.	Severely challenged in their ability to perceive word stress.
II	Fijian	Exposure to rule-governed stress.	Low ability to perceive word stress.
III	Hungarian	Occasionally exposed to variable lexical stress.	Mediocre ability to perceive words stress.
IV	Polish	Frequently exposed to variable lexical stress.	Good ability to perceive word stress.

It thus seems as though speakers of languages with predictable word stress are at a distinct disadvantage when learning English compared to speakers whose L1 has unpredictable word stress. However, it may be the case that perceiving word stress is a skill that can be improved even for speakers of L1s with predictable word stress as shown in 2.2.1.

It is worth noticing that not all studies have been able to confirm the predictions of the SDM. As a case in point, Choi, Tong, and Samuel (2019) found that L1 speakers of Cantonese, a tone language, actually outperformed native speakers of English in a test of stress perception in English word pairs. Comparing the test scores of 30 L1 English speakers to those of 30 L1 Cantonese speakers, the authors discovered that by tapping into their experience with perceiving tone in Cantonese, the NNSs were able to achieve better results than native speakers of English for both non-words and real English words. The Cantonese speakers' advantage disappeared when F0 was removed as a cue, leading the authors to conclude that it was indeed the tonal aspect of lexical stress that the NNSs utilised. These findings are rather extraordinary, and it would be interesting to see them replicated in other studies – possibly with speakers of other tone languages. This is because the suggestion that speakers of tone languages should be able to utilise their predisposition to hear tone to identify stress would entail that speakers of English can use their ability to perceive prosodic changes in English to distinguish tone contours in Mandarin (Archibald, 1997). However, this does not seem to be the case.

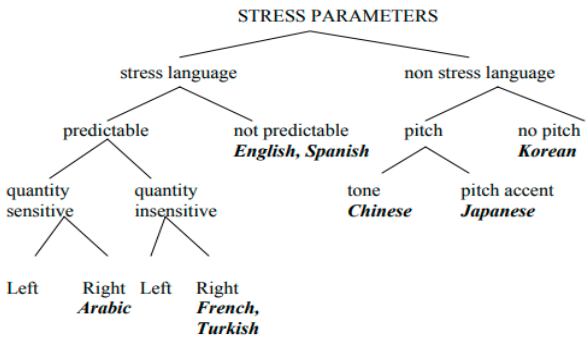
As a final note, it might be worth revisiting Ortega-Llebaria, Gu and Fan (2013). As stated above, they believe that stress is cued by various acoustic correlates – as do most other researchers in the field. However, where Ortega-Llebaria et al. differ is that they argue that the problem of perception is one of lower-order processing. As such, while the SDM would predict English speakers to have good perceptual skills in L2 Spanish (which is not the case), the authors argue that although L1 English speakers are used to utilizing the same correlates as L1 Spanish speakers, they use them slightly differently. This leads to what the authors refer to as context-sensitive stress deafness.

Stress typology model (STM)

Whereas the SDM only deals with stress perception, the Stress Typology Model attempts to predict the ability of language learners to both perceive and produce lexical stress in the L2. As can be seen in Figure 2.2, languages are categorised depending on whether certain parameters are present or not. An important point in the STM is that only a positive setting for a parameter may impede perception in an L2.

Figure 2.2.

STM: Stress parameters believed to influence stress perception and production taken from Altmann (2006).



In other words, French speakers are predicted to struggle with English lexical stress because French is categorised as a '+predictable stress' language. By contrast, English and Spanish share the same parameters meaning that lexical stress should cause few issues. Recalling once more the results obtained by Choi, Tong, and Samuel (2019), it is interesting to note that Cantonese has no impeding parameters, thus encouraging the prediction that L1 speakers of Cantonese should perform well on tests of perception of lexical stress in English. Similar results were obtained by Altmann (2006), who showed that speakers of the non-stress languages Chinese, Korean, and Japanese performed to the same level as native speakers of English on English perception tests.

With regard to stress production, the results obtained by Altmann (2006) are somewhat surprising. The Korean, Japanese, and Chinese speakers, who had performed to native-like levels in the perception tasks, were significantly outperformed by the speakers of stress languages. Thus, Altmann hypothesises that whereas the parameter '+stress language' seems to impede the perception of lexical stress in the L2, for production, this parameter could be rather important. Additionally, Altmann's results suggest that although speakers of Spanish are used to unpredictable word stress, they did not agree with the native English speakers with regard to word stress in many of the tested constructions. Rather unexpectedly, it was the speakers of stress languages with predictable word stress who performed most closely to the English baseline group.

The STM has not received a great deal of attention in the literature, but one of the studies that did set out to test its predictions was Brawerman-Albani and Becker (2014). More specifically, they tested the ability of L1 Brazilian Portuguese speakers to perceive and produce English words with preantepenultimate stress as this is a type of stress rarely encountered in the participants' L1. With Brazilian Portuguese being a stress language with unpredictable stress, like Spanish, the STM predicts that the participants would have a decent ability to perceive the English words i.e., a *YES* for the parameter 'stress language', but a *NO* for the parameter 'predictable stress'. In terms of production, the model predicts some potential issues as the L1 Spanish speakers in Altmann (2006) showed very little agreement with the English baseline group compared to speakers of stress languages with predictable word stress. According to the data obtained by Brawerman-Albani and Becker, the predictions were, for the most part, met. In the perception test, the participants scored 85.1% correct on average, which is very good compared to non-native speakers in other studies. However, in the production task, the participants only scored 28.1%. While some issues had been predicted, this is still a rather low score. In their study of L1 English speakers' ability to perceive Spanish word stress, Romanelli et al. (2015) also tested the predictions of the STM. Once again, the findings were discouraging as the authors stated 'stress seems to be a major problem for English speakers learning Spanish, in contrast to the predictions made by the STM' (p. 39). Hence, their findings contradict the predictions of the STM, since the L1 English speakers were expected to be able to learn to perceive stress in Spanish at the same level as the L1 Spanish speakers due to the fact that the parameter settings for the two languages are the same (see Figure 2.2 above).

The parameter-based explanation given by Altmann is somehow a little too vague to be entirely satisfactory. One reason is that it is not obvious or logical that one parameter can lead to great perception and poor production at the same time. In addition, it has been pointed out that the binary branching used in the model might be an overly simple way to depict the phonological structure of languages. Kijak (2009), for example, argues that the parameter 'predictable stress' is imprecise because English has somewhat predictable word stress in the sense that some morphemes attract stress while others may shift the stress in a word to a predictable syllable. Similarly, as mentioned above, there are languages with so-called predictable stress which has several unpredictable exceptions.

In addition to these arguments, some studies have found that Spanish speakers may display no stress deafness at all (Peperkamp, Vendelin, and Dupoux, 2010). Additionally, Hualde and Kim (2015) found that, contrary to what the STM predicts, L1 Korean speakers achieved rather low scores on their perception tests. The researchers tested both two- and three-syllable words with stress in all possible positions, and discovered that the Korean participants showed a very strong tendency to rate words as stressed on the penultimate syllable; so much so that they scored below chance on the tokens with stress on the last syllable. These findings clearly contradict the prediction that a *YES* for the 'stress language' parameter does somehow mean that L1 speakers of these languages struggle with lexical stress and vice versa. Thus, it is probably safe to say that more research is still needed in the field of stress perception and production models. These models should take into account both language experience and language background among other individual characteristics.

In summary, although there are models available that predict stress perception and production, and despite the fact that these models have proven accurate in several studies, there is still no model that gives a satisfactory account for L2 perception and production when it comes to stress. Learners who have no experience with word stress may display stress deafness, but it is possible that speakers of tonal languages may be able to offset this disadvantage by tapping into their ability to use tone. Learners who do have experience with word stress, but only in fixed positions, seem to be challenged by L2s with variable word stress patterns. Finally, even speakers of languages with variable stress may still find word stress in an L2 challenging as the cues used to signal stress differ between languages, and may not be transferrable.

2.2.3. Stress in Individual Words

Most people who have looked up a word in a good dictionary will probably have noticed that stress in words of more than one syllable is often indicated using an apostrophe-like mark often referred to as a 'vertical stroke' (IPA, 1999, p. 174) e.g., ˈ*garlic.* Most dictionaries operate with three levels of stress, namely 'primary', 'secondary' and 'no stress'. Roach (2009) suggests that a tertiary stress level can also be identified (e.g., with full vowels), but concedes that in most cases, this will only make things unnecessarily complicated.

As can be seen from the previous sentence, the number of stress levels has indeed been a topic for debate in the field of linguistics. Chomsky and Halle (1968), for example, claimed that there are at least five – and possibly more – stress levels in English. They further stated that 'there is little reason to suppose that the perceived stress contour must represent some physical property of the utterance in a point-by-point fashion' (p. 25). However, according to Ladefoged and Johnson (2010), the assertion that there might be that many levels of stress in English is not in compliance with phonetic facts. These authors thus suggest limiting the number of stress levels to two, namely primary stress and secondary stress, as 'no stress' is not regarded as a level of stress. They further argue that the difference between primary and secondary stress is not actually a matter of stress but one of pitch change. It is easy to see how the two accounts are bound to be conflicting as Chomsky and Halle describe the perceived reality of the listener, whereas Ladefoged and Johnson are concerned with the production by the speaker. Chomsky and Halle do not deny the fact that there is a phonetic reality to stress contours, but they argue that the physical reality is different from the perceived reality.

Taking once again the perspective of a teaching context, using three levels of stress (primary, secondary, and no stress) seems to make a great deal of sense, despite it possibly being at odds with the phonetic facts as stated by Ladefoged and Johnson and simultaneously falling short of the theoretical multi-level description put forward by Chomsky and Halle. This is because the three-level description of word stress allows the teacher to draw students' attention to the fact that stress in English is indeed variable (as described in 2.2), and that the stress pattern can change the mean of a word.

Although not to the same extent, the number of stress levels in Spanish has also been a topic for debate for some time. Hualde (2007) argues that Spanish has both primary and secondary stress, but that secondary stress in Spanish differs from secondary stress in English in that, in Spanish, 'secondary stress is a purely post-lexical or phrase-level phenomenon' (p. 80). However, it is generally accepted that Spanish operates with two stress levels. Regardless of the number of stress levels that academics argue about, in a teaching context, it is possible to work with the same system used in dictionaries i.e., one that uses primary stress, secondary stress and no stress.

Despite the fact that Spanish and English belong to the same category of languages in terms of the use of word stress (i.e., variable stress languages), it is safe to say that stress patterns in Spanish are much more rigid than in English. A part of the explanation is that English has been heavily influenced by both Germanic and Romance languages throughout its history (Cruttenden, 1997). This would not necessarily be a problem if a word's etymology would lead the learner to a target-like pronunciation. Unfortunately, this is not always the case as many words derived from Latin follow a Germanic stress pattern (Mott, 1996). What has brought about this change is not easy to say, but there seems to be a tendency for native speakers of English to prefer stress on the first syllable of a word – at least in two-syllable words (Vitevitch et al., 1997) – a tendency which is also found in other Germanic languages.

The stress pattern in individual words is, in some cases, determined by the presence of a suffix. As a general rule, suffixes of Latin or Greek origin (e.g., -tion, -ic, -eous) are said to be stress-shifting because they move the word stress to the syllable just before the suffix e.g., *tranSPORT* (v.) vs *transporTAtion* (n.); suffixes of French origin (e.g., -aire, -ee, -ette) are said to be stress-bearing as they take the stress on to themselves e.g., *billionAIRE*, whereas suffixes of Germanic origin (e.g., -dom, -er, -ly), on the other hand, do not tend to affect the stress pattern and can be said to be stress-neutral e.g., *FORTnight* (n.) vs *FORTnightly* (adv.) (Yavas, 2011).

Outside the realm of suffixes, it is worth noting that word stress may be dialect dependent. For example, the pronunciation of words such as *research, oregano,* and *perfume* may vary depending on whether they are uttered by a speaker of American or British English. To add to the confusion, even within these broad varieties of English, varying stress patterns can sometimes be applied to one lexical item: e.g., *CONtribute* or conTRIbute, *KIlometre* or *kiLOmetre* (Mompeán, 2010). This variation does not seem to be particularly rule-governed, and in most cases, the best a learner can do is to learn these by heart.

2.2.4. Stress in Compounds

Part of this thesis focuses on the teaching of compound nouns. A compound noun is a lexical item that consists of two or more elements or constituents e.g., *credit card*, *well-being*, *blackboard*, *take-off*, *credit card bill*. As can be seen from these examples, compound nouns (or indeed any compound) can be spelt in one of three ways: in one word, with a hyphen, or in two words. Unfortunately, the spelling provides learners with limited information about the stress pattern of the compound. What can further be seen from the examples above is that the elements do not have to be nouns themselves. Although compounds consisting of two nouns (NN) are the most frequent, many other combinations are possible. Some examples are provided in Table 2.3.

Table 2.3.
Examples of different types of English compound nouns.

Item	Type	Item	Type
bacon sandwich	N. + N.	bitter-sweet	Adj. + Adj.
compact disk	Adj. + N.	farfetched	Adv. + Participle
sky-high	N. + Adj.	runner-up	N. (Agent) + Adv.

There has been some debate about the stress assignment in compound nouns in English. While there is general agreement among researchers that the vast majority of compound nouns in English have single stress, there seems to be less certainty about how to treat compound nouns with double stress. For example, Quirk et al. (1985) argue that 'in some cases we may be in doubt as to whether we should regard sequences with this stress pattern as compounds or as free syntactic phrases' (pp.

1592-1593).

Having seen the amount of debate there has been over the nature of stress, it should come as no surprise that both the stress pattern and even the nature of compound nouns have been questioned. Based on the concepts of generative phonology, Chomsky and Halle (1968) formulated the Compound Rule which, in its essence, states that compounds of nouns, verbs, and adjectives are stressed on the first element. As is evident from the materials used in this study, which only deals with nouns, it is not difficult to come up with exceptions to this rule, and Chomsky and Halle do acknowledge that there are exceptions to the rule 'of various sorts' (p. 156). Later, using the term Compound Stress Rule (CSR), Liberman and Prince (1977) elaborated on Chomsky and Halle's Compound Rule by showing that its predictions can also be applied (and even improved) in metrical phonology – a post-generative theory of stress popular in the 1970s and 1980s. However, Giegerich (2009) flatly states that 'the CSR accepted in the literature for the past fifty or so years is wrong' (p. 14), and argues that many word combinations that are often taken for compound nouns are in fact noun phrases. Recall that this was also what Quirk et al. (1985) alluded to.

For a long time, it has thus been argued that an NN structure with double stress, rather than being a noun, should instead be interpreted as noun phrases (Bloomfield, 1933). However, as pointed out by Giegerich (2009), this claim is difficult to support when considering that words like *Madison Street* and *Madison Avenue* are very similar in terms of syntactic and semantic qualities, and yet have different stress patterns. As mentioned, this is not to say that Giegerich necessarily agrees with the common interpretation of NN structures in English, as he argues that structures such as *steel bridge* should be analysed as phrases rather than nouns. Comparing items such as *steel bridge* and *wooden bridge* from a syntactic point of view, he points out that constructions such as *a wooden bridge and a steel one* are perfectly grammatical whereas **a clock-maker and a watch-one* are ungrammatical. He thus suggests that many NN combinations are, in fact, noun phrases where the first element is an adjectival modifier of the head, thus distinguishing between compounds nouns, which are said to be complement-head constructions and NPs, which are said to be attribute-head constructions.

In this analysis, constructions such as *wooden bridge* and *steel bridge* should be stressed the same, and the reason behind this shift in stress pattern between the two semantically very similar words is not immediately obvious. However, it has been suggested that an explanation could be to do with the lexicalisation of the compound *steel bridge* (Plag et al., 2008). Exactly what is behind the process of lexicalisation is not entirely clear either, but some have suggested that the process could be influenced by word frequency (Mompeán, 2014; Plag et al., 2008). This would indeed explain why many chefs seem to pronounce various oils using single stress when, according to the general rule, they should have double stress, as they can be categorised as foods or ingredients. However, this observation probably needs verifying experimentally.

The somewhat philosophical ponderings by both Bloomfield and Giegerich are, in part, based on the role stress plays in deciding whether a compound should be considered a lexical item or a phrase. For example, ˈ*hot dog* is a term used for a sausage often served in a bun, whereas *hot* ˈ*dog* is a noun phrase that would describe a canine that is warmer than usual. Similarly, a ˈ*blackboard* is used in classrooms whereas a *black* ˈ*board* could be any type of board with the colour black. In writing, it is easy to pick out the difference between the two as one is spelt in one word and the other in two. In speech, however, the listener only has the stress pattern to rely on, and admittedly, the context, which can clear up many problems in some situations.

The analysis above leaves teachers with two options: either they omit NN combinations with materials in them from their lessons, or they ignore the deep analysis of NN structures in English and focus on the fact that including materials in compounds will help their students correctly assign stress to a large number of lexical items in English. Thus, this project adopts the approach taken by Plag (2006) in that no clear distinction between compound nouns and NPs is made. This also makes sense because this approach is in accordance with the majority of published pronunciation materials (Hancock, 2003; Hewings, 2007).

Another feature of compounds that should be mentioned is that of stress shift, the tendency to shift the primary stress in simple or compound words to a previous syllable with secondary in connected speech for rhythmic reasons. Stress shift happens when the last syllable of the first element is stressed in its citation form and the second element is stressed on the first syllable. A classic example of stress shift from the literature is that of *Chinese restaurant* (e.g., Henderson, 2010). In its citation form, *Chinese* is primarily stressed on the last syllable, and also carries secondary stress on the first syllable, i.e., /ˌtʃaɪˈniːz/. However, because *restaurant* has primary stress on the first syllable, the primary stress in *Chinese* is shifted onto the first syllable of the word so the stressed syllables do not clash. The urge for native speakers of English to avoid a clash of stressed syllables is so strong that it may even change the pronunciation of names as in *Fidel ˈCastro* as opposed to the Spanish pronunciation of the name *Fiˈdel ˈCastro* (Ortiz-Lira, 1998).

As with stress in individual words, there appear to be some differences between dialects. For example, in British English, compounds with the word *cake* as the last element have single stress, whereas in American English they have double stress. As a general rule, compounds spelt in one word are said to have single stress.

However, there are exceptions to this rule, such as *marshmallow* - pronounced /ˌmɑː(r)ʃˈmæləʊ/.

Similar to English, compound nouns in Spanish can consist of various word classes. Whereas N+N type compounds are the most frequent in English, in Spanish, although it does occur, this construction is often replaced by N + *de* + N (e.g., *abuso de droga*; 'drug abuse'). Another difference between English and Spanish compounds relates to their stress patterns. Hualde (2007) explains that the reason for this is that English has several stress levels and allow vowels to be either full or reduced. This, he says, means that the elements in a compound can be described in terms of a subordination relationship (i.e., which has the most stress). In Spanish, on the other hand, because these features are not present, the relationship between the elements in the compound changes. Hence, in Spanish, the elements in a compound can either retain their stress level or the stress on the first element can be deleted. This is illustrated in Figure 2.3. In the word *pine tree*, the first element has stronger stress (xx) than the second element (x). In *gold medal,* on the other hand, the relationship is the other way round, with the second element displaying the highest level of stress. When looking at the Spanish compounds, the two elements either both retain their stress (*perro lobo*) or the stress on the first element is deleted (*tocadiscos*). It must be added that Hualde concedes that 'the topic is obviously open to experimental verification' (2007, p. 67).

Figure 2.3.

Stress levels in English and Spanish compound nouns.

```
English:     x                 x
             x   x         x   x
           pine tree vs.  gold medal

Spanish:
             x    x            x
           perro lobo  vs. toca discos
```

Note: Figure is taken from Hualde (2007, p. 67).

As can be seen, the one stress pattern that is not used in Spanish is the most frequent one in English, namely single stress, in which the first element is stressed, but the second element is not. For this reason, it can be theorised that L1 speakers of Spanish may struggle with this particular type of compound.

2.3. Sentence Stress

As would be expected, sentence stress, also known as prosodic stress, refers to where stress is placed in an utterance. The point can be made that the term ‘sentence stress’ is somewhat illogical because it refers to a phenomenon in connected speech, yet the term ‘sentence’ refers to a grammatical construct. However, for the sake of simplicity, the words ‘sentence’ and ‘utterance’ will be used as synonyms.

To iterate a point made at the beginning of the chapter, non-nuclear rhythmic stress will be disregarded in this thesis as the focus is nuclear stress placement. As will be described in section 2.4, English sentences consist of stressed and unstressed syllables with some degree of equal intervals between them, but one syllable always stands out among the rest as more prominent in terms of pitch, duration, and intensity. This syllable is known as the tonic syllable. In some languages, placing the tonic syllable is rather straightforward, as is the case in French, which places the stress on the last word of the phrase. The same is, to a large extent, the case in Spanish, in which the last word is also stressed. In English, however, the nature of sentence stress is far more complex because the main stress in the sentence can be placed in various places depending on what the speaker wants to draw attention to (Cruttenden, 1997). It follows naturally from the varied sentence stress patterns of English, that nuclear stress can be final or non-final, but predicting where the stress goes is not necessarily straightforward. The sections below will outline some of the debates and issues in the field of English and – to some extent – Spanish sentence stress.

2.3.1. The three Ts

Sentence stress is part of what in the literature is often referred to as 'the three Ts', which encompass Tonality, Tone, and Tonicity (Halliday, 1967). Tonality refers to how speakers divide an utterance into smaller units. These are referred to as intonation units (IUs), thought groups, or tone groups, and can be seen as the punctuation of spoken language (Goodwin, 2014). There is a certain degree of predictability in the way IUs are structured. In English, IUs can have several stressed words knowns as rhythmic stress (Mompeán, 2014), but one – the nucleus – will stand out as more prominent than the others. At this point, it might be worth having a brief look at the theorised structure of intonation units in English. An IU can be said to consist of a pre-head, a head, a nucleus and a tail (Wells, 2006) as shown in example 2.A.

(2.A)	[He's]	[presenting a paper at a]	[CONference],	[I believe]
	Pre-head	**head**	**nucleus**	**tail**

Whereas the pre-head consists solely of unstressed syllables, the head may contain one or more stressed syllables and starts at the first rhythmically stressed syllable in the intonation unit and ends just before the nucleus. The nucleus, as mentioned, is the most prominent syllable in the intonation unit, and is generally the last rhythmically stressed syllable in the intonation unit, as in the example above. As will be discussed below, the nucleus is often placed on the last lexical item (e.g., *Peter bought his wife a DRESS)*, although exceptions do exist as (2.A) shows. In some cases, an intonation unit may additionally have a tail, which often consists of unstressed syllables, although post-nuclear rhythmical stress is possible (Wells, 2006).

Tonicity, which will be the focus of this subsection, deals with the placement of the nucleus, also known as the tonic syllable (Halliday, 1967). Nucleus placement has been a topic for debate for a few decades now, and so far, conflicting views still exist. Exactly where the nucleus falls in an English sentence is not entirely predictable, although some general rules do apply. This will be the focus of the next subsection.

Finally tone describes how speakers can use the tonal pattern on the nucleus to signal different meanings in terms of mood and attitude. For example, in the sentence *let's go,* the speaker can sound either neutral or impatient depending on the tone applied to the nucleus.

2.3.2. Nucleus Placement

Native speakers of English rarely get nucleus placement wrong. Even so, exactly where the nucleus goes is not as straightforward as one might think. The nature and the placement of the nucleus within an intonation unit have been debated heavily over the years (Bolinger, 1985; Gussenhoven, 1983; Ladd, 1980). The intention here is not to resolve this debate, but rather to shed some light on the various points of view in the literature. To begin with, only nucleus placement in a neutral or unmarked context will be considered. Even the meaning of 'unmarked' has been debated, but here the approach taken by Cruttenden (1990) will be followed as he states that 'this is a debate in which I do not wish to get embroiled here' (p. 9). Instead, it will suffice to say that unmarked tonicity will refer to cases where no particular emphasis is used.

In a neutral context, there seems to be general agreement between several approaches to nucleus placement. In the generative approach, as described by Chomsky and Halle (1968), nucleus assignment is said to be cyclical. This basically means that an utterance undergoes several cycles of nucleus assignment before it is produced by the speaker (see 2.B – 2.D). Each cycle modifies the structure according to pre-set rules so that, in theory, neutral tonicity is entirely predictable. For example, in the sentence s*trong men lift weight*s each word is first assigned a nucleus (2.B). In the example below, this is symbolised with the number '1' in accordance with the notation used by Chomsky and Halle (1968). In the next step, each constituent is assigned a nucleus, which is the rightmost item (2.C). This process involves downgrading the elements that no longer have a nucleus. Then the whole intonation unit gets a nucleus; again, the rightmost item, and again downgrading the other elements (2.D).

```
           1              1            1             1
(2.B) [Strong]        [men]       [lift]        [weights]

           2     1      2    1
(2.C) [Strong men]  [lift weights]

           3     2   3    1
(2.D) [Strong men lift weights]
```

In a metrical account (Liberman and Prince, 1977), right branches are said to be strong when dealing with items above word level. This entails that the stress pattern in the example sentence above would be the same, at least in terms of nucleus placement. Thus, neither of the two accounts seem to contradict the Last Lexical Item rule (LLI) as expressed by Halliday (1967), which states that unless some form of markedness is involved, the nucleus of a sentence falls on the last lexical item of an intonation unit as pointed out above (2.3.1). Here, a lexical item is either a noun, main verb, adjective, or adverb. These word classes are often referred to as 'content words' and are seen in opposition to 'function words' (e.g., prepositions, pronouns, and determiners). In this respect, Wells (2006) argues that the distinction between 'word' and 'lexical item' is paramount as using the latter enables the inclusion of compounds at the end of the intonation unit, e.g., *John lives on Water Street*, where the nucleus is on *Water*, despite it not being sentence-final.

Even though the LLI accounts for the nucleus placement in the vast majority of English sentences, there are some exceptions. Cruttenden (1997) shows that in yes/no questions, the nucleus may be placed on a word outside the four categories mentioned above as shown in (2.E):

(2.E) Are you going OUT tonight?

Additionally, Wells (2006) adds that some function words tend to attract the nucleus, particularly when they are sentence-final. The word *too* is one such word as shown in (2.F):

(2.F) Mary likes ice cream, and Peter likes ice cream TOO.

In these cases, however, *too* may sometimes have its own intonation unit, which means that in the preceding clause, the nucleus falls on the last lexical item (*ice cream*) as predicted by the LLI.

Another scenario in which the LLI seems inaccurate is in cases where the last lexical item is a content word, but the nucleus falls elsewhere as illustrated in (2.G).

(2.G) They get on like a HOUSE on fire.

Additionally, Wells (2006) shows that native speakers of English tend to prefer to place the nucleus on a noun, even if other lexical items are available nearer the end of the intonation unit as shown in (2.H).

(2.H) Which BOOK did you choose?

It is thus clear that there are at least some exceptions to the LLI. Ortiz-Lira (1995) identified nine cases which he believed could be exceptions from the LLI. These are presented in Table 2.4 below. The first three refer to the presence of a final adverbial, vocative, or given information after the nuclear stress. The fourth item is 'reporting clauses', which are used to describe someone saying something (e.g., *'Where did you go', Peter asked Paul*). The fifth item on the list is 'objects of general reference', which refer to so-called 'empty' words – these will be described in more detail below.

Table 2.4.

Problematic areas of non-final tonicity identified by Ortiz-Lira (1995).

Note: Items in italics were selected for Study 2.

1 Final adverbial	**6** Post-modifying infinitive clause
2 Final vocative	**7** Restrictive relative clause in definite NP
3 Final given information	**8** Softening phrases
4 Reporting clauses	**9** *Non-pronominal subject+intransitive predicate*
5 *Objects of general reference*	

The case of 'post-modifying infinitive clauses' is the sixth item in the table. These refer to infinitive clauses that come after the nucleus, such as *I have lots of work to do*. Next on the list is 'restrictive relative clauses' which, as the name implies, refer to relative clauses that restrict or define something. For example, in the sentence *I live across the street from the pub (that) we used to go to*, the relative clause *(that) we used to go to* does not contain a nucleus. The penultimate item is softening phrases. These can be used to slightly modify an utterance as in the following example: *He'll be here on Friday, I think*. The ninth and final item is sentences with a non-pronominal subject+intransitive predicate. These will also be described in more detail below as they were selected for the study.

Of these exceptions, two were selected for this study because Ortiz-Lira (1995) found that they proved very difficult for the L1 Spanish speakers (a third exception, namely 'contrastive stress' was also selected). The first of these is number five from the list above, namely 'objects of general reference'. These have also been referred to as 'empty words' by Wells (2006, p. 150) because they lack the semantic weight one would expect. Among these are nouns such as *guy, places, stuff* and *things*, pronouns such as *someone*, as well as adverbs such as *somewhere.* For example, if one compares the two sentences (2.I-2.J), (2.I) I know JOHN.

(2.J) I KNOW someone.

it can be seen that whereas *John* in (2.G) is a specific person, *someone* in (2.H) is much less specific and could refer to anyone. As can be seen in (2.K) and (2.L), which are the Spanish equivalents of 2.7 and 2.8, this is different in Spanish.

(2.K) Conozco a JUAN.

(2.L) Conozco a ALguien.

Here, the nucleus is kept on the last item regardless of its semantic weight (Hualde, 2005). It must be added that pronouns in English are not necessarily empty words as the sentences in (2.M-2.N) show.

(2.M) The keys must be SOMEwhere!

(2.N) If you're bored, we can go for a DRIVE somewhere.

In 2.M, *somewhere* refers to someplace specific, i.e., the place where the keys are. In 2.N on the other hand, *somewhere* is much less specific and could refer to any place accessible by car.

The other exception from Ortiz-Lira (1995) selected for this study was number nine from Table 2.4, namely 'non-pronominal subject+intransitive predicate'. These constructions are also known as 'event' sentences. Event sentences have been described as structures that often (but not always) describe disappearance or misfortune (Cruttenden, 1997). The following is one of the examples used by Cruttenden (1997).

(2.O) Watch out! That CHIMney's falling down.

Wells (2006) uses a somewhat broader description of event sentences as he says that these are sentences that describe events, and which have an intransitive verb, as shown in (2.P).

(2.P) The PHONE is ringing.

He further points out that, in these sentences, the nucleus falls on the noun, which, as mentioned above, is a common feature in English nucleus assignment. It should be mentioned that a distinction has been made between event sentences and sentences that carry a notion of something unexpected (Estebas-Vilaplana, 2015). In this view, event sentences have a relatively narrow pitch range and could include something fairly mundane as

(2.Q) The KETtle is boiling.

whereas sentences with broader use of pitch range should be interpreted as surprising in one way or another. Thus, Cruttenden's example above would not strictly be an event sentence as it clearly conveys an element of surprise. However, for the purpose of this thesis, the approach followed by Cruttenden (1997) and Wells (2006) will be used.

The third case of non-final nucleus placements selected for this thesis is contrastive stress. Contrastive stress can be sentence-final, but it is easy to think of examples in which this is not the case.

As with almost every other term in linguistics, the nature of contrastive stress has been a topic for debate. Some authors (e.g., Celik, 2004), distinguish between 'emphatic stress' and 'contrastive stress'. In this view, the speaker can emphasise part of their speech as in (2.R),

(2.R) Yesterday's lecture was VERY boring.

or they can create a contrast to something that has been either implicitly or explicitly stated:

(2.S) THESE bowls are white.

In 2.S, the bowls that the speaker is talking about are referred to in opposition to some other bowls of a different colour, hence the term 'contrastive stress'.

Other authors (e.g., Couper-Kuhlen, 1986), however, argue that contrastive and emphatic stress is one and the same thing. Bolinger (1961) disagrees with the terminology in general as he makes the distinction between 'contrastive accent' and 'contrastive stress'. He argues that the latter should be used for within-word stress patterns, whereas the former is better suited for sentence-level analysis due to the influence pitch change has on the production and perception of accent and stress.

Cruttenden (1997) simply states that the term 'contrastive' is one that intonation researchers cannot escape using, but which cannot easily be given an exact definition. Similarly, using Bolinger (1961) as a basis for her analysis, Schmerling (1976) concludes that 'the notion of "contrastive stress", like the notion of "normal stress", is not a particularly useful one; contrastive stress does not seem to be a clearly definable entity unto itself' (p. 66).

Despite the disagreement among academics, in teaching, contrastive stress can, in most cases, be equated with highlighting. Hence, learners need to know that they can guide the listener's attention to a specific part of their utterance by use of stress. However, although this is a relatively simple rule, it is not necessarily easy to apply in practice.

Whereas finding exceptions for the LLI in English is easy, in Spanish, the last lexical item rule is far more accurate (Hualde, 2007), although with the peculiarity that any word (not just content words) at the end of an intonation unit can contain the nucleus (Gutiérrez-Díez, 2012). That said, there are cases in Spanish where the nucleus can be placed earlier in the intonation unit. The examples below are taken from Hualde (2007).

(2.T) PEdro trajo el libro.

(2.U) El libro lo trajo PEdro.

In (2.T), special emphasis is placed on *Pedro* to signify that it was him and not, say, *Maria* or *Pablo* who performed the action. In (2.U), on the other hand, *Pedro* is in a sentence-final position and thus becomes the default focus of the sentence. This seems to be the preferred option of the two in Spanish, as Gutiérrez-Díez (2005) states that 'rather than using intonation to establish information focus, several syntactic devices are preferred for the same purpose, such as word order, changes in the theme-rheme structure and the use of cleft or pseudo-cleft sentences' (p. 132).

There is some debate about the extent to which L1 speakers of Spanish use this shift in nucleus when they speak Spanish. García Lecumberri (1995) argues that nucleus displacement is used to varying degrees depending on the construction of the sentence, but her analysis has been disputed by Gutiérrez-Díez (2005, 2008).

Regardless of how frequently nucleus displacement is used in Spanish, it is safe to say that it is much less frequent than in English, and there is little debate about the fact that this way of highlighting information can be problematic for L1 speakers of Spanish learning English.

2.4 Stress-timed and Syllable-timed Languages

As a final mention, the concepts 'stress timing' and 'syllable timing' will be included, as this distinction helps to account for some of the differences between English and Spanish sentence stress patterns. In its essence, stress-timed languages are said to equal amounts of time between each syllable. This is also known as isochrony (Cauldwell, 1996). A consequence of isochrony is that syllables have to vary in length. For example, the distance between the stressed syllables is supposedly equal in the two sentences *the horses ate the grass* and *the horses might have eaten the grass*. This is despite the latter having more syllables between the stressed words. English is often said to be a stress-timed language, although not all varieties of English follow this pattern (Ashby and Maidment, 2005). Syllable-timed languages, on the other hand, are languages where each syllable takes roughly the same time to pronounce. Spanish and French are often used as examples of syllable-timed languages. It should be obvious that languages that have varying syllable lengths and languages that have equal syllable lengths will sound different in terms of rhythm.

With regard to the notion of 'stress-timing', research has not been able to validate the claim of isochrony. On the contrary, speech data seems to suggest that not only does isochrony not apply to English, but that the whole concept of stress-timed vs. syllable-timed distinction is faulty (Cauldwell, 1996; Hardison, 2002; Nolan and Jeon, 2014).

Although very few, if any, linguists believe in the strong version of this concept, it can still be prudent to maintain a dichotomy between stress-timed and syllable-timed languages for language teaching purposes. One of the reasons is that although research has been unable to show how languages differ in terms of stress-timing, the mere fact that there is a difference is something learners can hold on to (Marks, 2007). Furthermore, it gives teachers something to work with as it allows them to direct their students' attention to these differences and raise their awareness of subtle, yet important, aspects of the L2 (Schmidt, 1990).

Based on current research data, the notion of stress-timed and syllable-timed languages is difficult to uphold. However, it is very plausible that languages fit on a continuum somewhere between two extremes (Trofimovich and Baker, 2006). Speakers of Germanic languages such as Dutch, German, and Danish seem to have relatively few issues learning this aspect of English as their L1 resembles English in terms of stress-timing (Swan and Smith, 2001). However, for speakers of Romance languages such as French, Italian, and Spanish, the situation is very different. Here, the differences in how sentence stress is applied can have significant consequences for their perception and production of English.

The current chapter has presented an overview of word and sentence stress in English. The next chapter looks at how the issues engendered by the differences between English and Spanish might be solved and how social media can help both learners and teachers in the process. It will also present the research questions used to guide the studies described in Chapter 4.

Chapter 3 – Objectives and Research Questions

As discussed in Chapter 1, learning pronunciation presents a range of challenges for both learners and teachers. The studies in this thesis aim to explore the extent to which the social media platforms Facebook (Study 1) and YouTube (Study 2) can be used to overcome some of these challenges. This is done by investigating the effectiveness of pronunciation courses delivered on these platforms. The focus is on aspects of English suprasegmental phonology that are believed to be difficult for L1 speakers of Spanish (see 2.2.4 and 2.3.2).

More specifically, Study 1 presents a study of a four-week pronunciation course teaching three stress rules for English compound nouns. These are English place names (e.g., *Victoria Street*), compounds of materials or ingredients (e.g., l*emon cake*), and compounds consisting of objects and agents (e.g., *skyscraper*). In a similar vein, Study 2 presents a study on a four-week pronunciation course aiming to teach learners three pronunciation rules relating to sentences stress. These are objects of general reference, event sentences, and contrasts. In addition, both studies investigate the participants' experience with and evaluation of the pronunciation course. This is done for the simple reason that if the participants dislike using social media for language learning, this will have considerable implications for the recommendations that can be made regardless of how much the participant improves their pronunciation.

Given the key role Facebook and YouTube play in the studies, a brief overview of the two platforms is provided below. Following this overview, the specific research questions of the two studies are presented before the final rationale for the two studies is provided.

3.1. Using Facebook and YouTube in Pronunciation Teaching

Even though Facebook and YouTube are both well-known to most people, a short introduction to the two social media platforms and their use in pronunciation teaching is provided here.

3.1.1. Facebook in Pronunciation Teaching

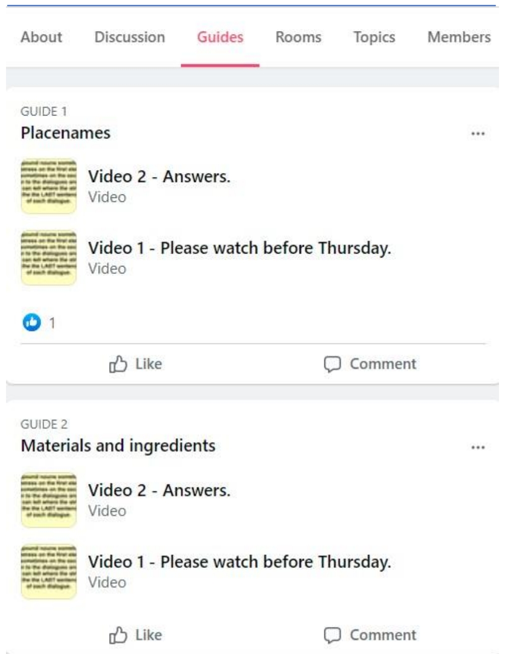
Since its inception in 2006, Facebook has gone from strength to strength in terms of user numbers. In 2012, it reached 1 billion users and in 2017 it reached 2 billion (forbes.com). In its essence, Facebook is a platform that allows people to create a personal profile or a page about a topic they find important, share content on their profile (their 'wall') or in groups, and connect with others for free. However, due to its many functionalities and intuitive interface, it lends itself to educational purposes. The site is compatible with all commonly used file formats such as .doc, .pdf, .img and .mp4 which facilitates great task variation in a language course. As Blattner and Fiori (2011) state, 'it is not technology itself that promotes learning, but rather the teaching methods employed' (p. 25), but given the many functionalities that Facebook has to offer, it must be considered as a very interesting learning tool indeed.

Users can create and join groups that match their interest, which is a feature that could be particularly useful when delivering course content (Blattner and Fiori, 2009), as learners can get notifications about new posts in their groups as soon as they are uploaded (in this way Facebook is at least on par with podcasts). The creator of the group is also the administrator, and can control who is invited to join the group, and who can view its content. This ensures a high degree of security for the students as well as the tutor. Facebook groups have been explored in academic contexts, but the results have been mixed (see section 1.5.4).

As a fairly recent innovation (April, 2018), Facebook has added the feature 'social learning groups'. These are interest groups like most others, but the difference is that the administrator can organise content into individual units, which makes it easy for a teacher to structure a course, and easy for students to follow. A screenshot from one of the groups in Study 1 is shown in Figure 3.1.

Figure 3.1.

Screenshot showing part of the layout of the Facebook pronunciation course.



One of the features that teachers might find helpful is that units can be prepared ahead of time and scheduled to upload at a pre-determined time. This means that teachers can design entire courses and reuse these in following semesters.

Bosch (2009) stated that 'if one considers the large numbers of students on Facebook often actively participating in discussions and groups, it cannot be ignored as a potential educational tool' (p. 190). However, despite its enormous popularity, Facebook has only received limited attention as a teaching tool, and more research into this area is needed – in particular when it comes to pronunciation teaching

(Wang and Vásquez, 2012). As mentioned in 1.5.1, this study reviewed 43 studies of language learning through web 2.0 tools, and only listed one that focused on pronunciation.

The studies that have been carried out on Facebook in a broader contexts have shown that students generally respond favourably to learning through Facebook. For example, Wang et al. (2012), attempted to use Facebook as a learning management system (LMS) in the same way commercial systems such as Edmodo or Blackboard are used. Despite some concerns from a minority of the 28 participants regarding privacy issues, they concluded that 'The Facebook group has the potential to be used as an LMS' (p. 437). This study disposes of some of the privacy issues that Wang et al. (2012) reported by using Facebook for uploading content only. Students had the option to leave comments in the group, but were not required to do so. Other studies on Facebook in language learning have shown that using Facebook made students more confident and increased their motivation in relation to L2 writing (Kho and Chuah, 2012; Promnitz-Hayashi, 2011; Yunus and Salehi, 2012). Hence, it would be very interesting to see if Facebook has a similar effect on students in terms of pronunciation training.

3.1.2. YouTube in Pronunciation Teaching

The use of YouTube for educational purposes has been the focus of several studies in recent years. Some of these have focused on user perceptions (Balbay and Kilis, 2017; Fleck et al., 2014; Kelsen, 2009; Seilstad, 2012), while others have investigated the effectiveness of YouTube as a language learning tool (see 1.5.3). YouTube has been suggested as a tool that that is well-suited to classroom use (Alwehaibi, 2015; Watkins and Wilkins, 2011). As mentioned in section 1.5.3, Hismanoglu (2012) found that using YouTube in classroom teaching can indeed engender positive changes in

EFL learners' oral proficiency. However, this does not mean that the platform would not be well-suited for autonomous use as well (Benson, 2015). This is particularly true with regard to Hismanoglu's study, in which the experimental group received additional instruction from the teacher on top of what they learnt from the videos. Thus, there is still a need for studies that show the effectiveness of YouTube in pronunciation teaching in a context where the learners have a much higher degree of autonomy.

There is reason to believe that if students already have a habit of using YouTube (which is likely due to its current popularity), doing exercises related to language learning will become more natural and perhaps seem less strenuous. This very idea was proposed in relation to using podcasts for language teaching by Thorne and Payne (2005), who stated that '[t]he opportunity presented by podcasting is that it leverages habituated behavior: many students already own portable mp3 players and routinely download content that they listen to during downtime or transition time between activities' (p. 386).

Finally, it has been suggested that there is a correlation between improved perception and improved production (Bradlow et al, 1997; Flege, 1995). With this in mind, looking at how the audio aspect of YouTube can be used for pronunciation learning seems appealing. Although the training used in the pronunciation courses described in this thesis is nothing like actual perceptual training, the input could help improve the learners' perception and hence their production. It should be emphasised that the link between perception and production is only correlational, and learners generally only experience relatively small gains in production from perceptual training alone (Sakai and Moorman, 2018). However, this issue can be mitigated on YouTube as the platform allows for tasks and/or instructions to be added in the video description, thereby providing learners with a bi-modal learning experience. It thus seems clear that YouTube affords features for pronunciation learning which have significant potential, but which have not been thoroughly explored.

3.2. Research Questions and Hypotheses

In order to carry out the investigations described above, the research questions below were formulated. Each research question is accompanied by a hypothesis based on the literature in the field and a potentially likely outcome based on the results and insights provided by various studies.

Study 1

RQ.1: Can the L1 Spanish learners of English state the three stress rules for English compound nouns practised in a four-week pronunciation course delivered through Facebook?

H1) The participants will be able to state the three rules they worked with. Considering the simplicity of the rules, this should not cause too many problems, although it should be added that the participants will not be told that learning the rules is part of the test.

RQ2: Can the L1 Spanish learners of English apply what they have learnt to aural input?

H2) The students will apply the rules to aural input by identifying the type of word they hear and selecting the correct option (single stress or double stress).

RQ.3: Can the learners apply what they have learnt to their speech production (i.e., produce single-stressed compound nouns correctly)?

H3) Although single stress is not used in Spanish, the participants are expected to be able to produce single-stressed compounds with the correct stress pattern.

RQ.4: Can learning be transferred to compounds with a similar structure?

H4) Once the learners know the pronunciation rules, they should be able to apply them to any compound of a similar type.

RQ.5: What are the students' perceptions of using Facebook for language learning?

H5) Based on previous research on technology and language learning, the participants are expected to respond positively to the participation to the pronunciation course and be favourable to using Facebook for language learning in general.

<u>Study 2:</u>

RQ.6: Can the L1 Spanish learners of English state three sentence stress rules for non-final tonicity following a four-week pronunciation delivered through YouTube?

H6) The participants will be able to state the rules following the course. This is predicted even though these rules are considered slightly more difficult.

RQ.7: Can L1 Spanish learners of English apply the sentence stress rules for non-final sentence stress to aural input?

H7) The participants will able to apply the rules they have learnt, and thus choose the correct option in the perception tests.

RQ.8: Can the participants apply the rules they have learnt to speech production in a read-out-loud task?

H8) The participants will be able to identify the sentence they are reading and apply the correct tonicity.

RQ.9: Can the participants apply contrastive stress correctly in a picture description task?

H9) The participants will be able to apply contrastive stress successfully. Although this a more challenging task than the read-out-loud task, it is made easier by not being timed.

RQ.10: Can learning be transferred to sentences with similar stress patterns?

H10) Learning will transfer to both aural input and speech production.

RQ.11: What are the students' perceptions of using YouTube for language learning?

H11) As in H5, based on previous studies on students' perceptions of online learning, the students are expected to rate the course itself as well as using YouTube for language learning favourably.

3.3. Rationale

This final section presents the rationale behind the two studies. Three areas are looked at in more detail. First, the section 'methodological considerations' discusses the thoughts behind the pronunciation course itself. Next, the section 'linguistic considerations' discusses why word and sentence stress were chosen as topics for the two studies. Finally, the section 'technological considerations' discusses what makes Facebook and YouTube potentially effective tools for pronunciation teaching. Together, these three areas should provide a solid foundation for the studies presented in Chapter 4.

3.3.1. Methodological Considerations

Pronunciation instruction can be approached in a variety of ways. As mentioned in

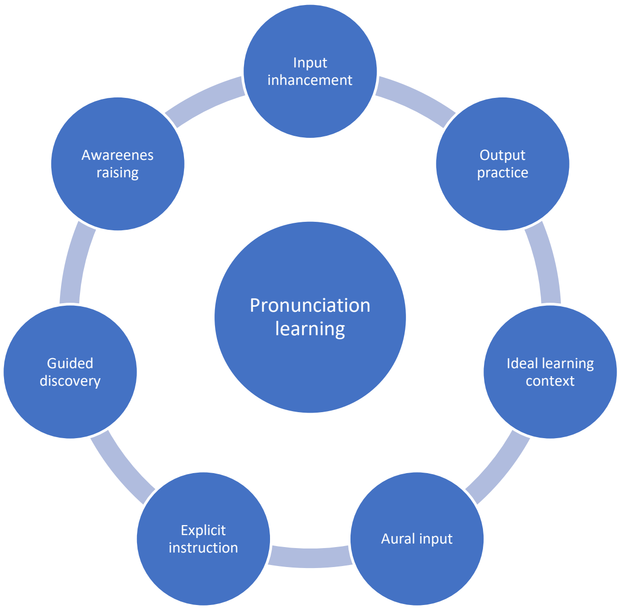
1.1, telling students what contrasts to focus on and then have them drill these repeatedly was the method used in the Audio-lingual Method. Other approaches have prioritised improving the learners' meta-linguistic knowledge. It has also been suggested that improving learners' perception through HVPT could be an effective way to improve L2 speech production.

When deciding how to do a pronunciation course through social media, it is important to consider at least two things. First of all, it is important to analyse what features the online tools offer. For example, there is little point in recording videos if the platform used does not allow videos to be played. Secondly, it is important to know what type of instruction is likely to be effective as well as appealing to the learners. The reason the latter is important is that if learners are working with the materials autonomously, these must engage the learners to ensure that they do, in fact, put in the required work.

Hence, in the creation of the courses, a range of potential factors were identified which were hoped to make the course effective. Many of these have already been described. These include explicit instruction (Saito, 2012), ideal learning conditions (Neri et al., 2002), output practice (Swain, 1985), and aural input (Krashen, 1983). The full set of factors that were identified are depicted in Figure 3.2. The concepts not yet touched on will be explained below.

Figure 3.2.

Graphic representation of the elements believed to contribute to the participants' learning experience.



Two of the concepts that have not yet been explained are 'input enhancement' and 'awareness raising'. Looking first at input enhancement, this concept has been thoroughly described by Sharwood Smith (1993) and Sharwood Smith and Truscott (2014). As the name indicates, input enhancement works by altering the input provided to the learner to make it better suited for learning. Input enhancement can be achieved in three ways, namely through increased salience, modification, or elaboration (Chapellle, 2003) as shown in Table 3.1.

Table 3.1.

Examples of different types of input enhancement.

Input enhancement	**Example**
Salience	Underlining Italicising Colour coding
Modification	Adding images Adding translations
Elaboration	Restructuring original text using more accessible grammar while maintaining original meaning

Increased salience is probably what most people associate with input enhancement. Examples of increased salience could be underlining text or adding colour to certain words in order to make them stand out. Salience can be increased for any type of input, be that visual or aural input. As a case in point, Lu, Wang, and De Silva (2012) used exaggerated stress cues to improve the perception of English sentence stress for participants of a variety of L1s. Modification on the other hand, refers to some sort of additional feature. This could be a translation of a phrase or an image to help learners understand a new word. Finally, elaboration refers to the addition of text to help explain the original input. This could be by adding to the original text synonyms that the learner is more likely to know in order to make difficult vocabulary more accessible.

The primary reason for using input enhancement is that it is theorised to lead to awareness raising. That is, by using enhanced input, teachers can raise learners' awareness of what they are supposed to be focusing on. This goes back to Schmidt (1990), who claimed that learners have to notice the gaps in their linguistic knowledge before they are able to fill them.

The theories of input enhancement and awareness raising are not without their critics. For example, from Krashen's point of view input enhancement and awareness raising are not useful at all because raising learners' awareness will only improve their learnt knowledge but not their acquired knowledge. Al-Hejin (2004) argues that awareness is not necessary for learning, but states that it facilitates it. In their review of 15 years of research on textual enhancement, Han, Park, and Combs (2008) concluded that the research on the matter has yielded mixed results. Despite these issues, research using these techniques in combination with other learning tools have yielded positive results. As a case in point, Fouz-González (2017) used input enhancement in combination with explicit instruction delivered through Twitter to help L1 Spanish speakers improve their pronunciation of a range of problematic words. Hence, there is reason to believe that employing this strategy as part of pronunciation course on other social media platforms is indeed worthwhile.

The final element included in Figure 3.2 is the term 'Guided discovery'. Also referred to as 'inductive learning', this teaching tool allows the learner to work with examples of a language feature in order to figure the rules out for themselves. According to some, this can make learning more effective. In the current studies, the students submitted their answers to the guided discovery through Google Forms (see 4.4.2), but this was primarily for the sake of making data collection more convenient. There is no reason a teacher could not receive the answers on Facebook, via email, on a wiki or some similar technology.

In sum, a wide range of issues have been addressed to ensure that the course is effective. Some relate to the conditions in which the learners can practise with the materials, while others relate how the course content is presented, and how the students actively work with the materials.

3.3.2. Linguistic Considerations

Because the aim of the two studies is to explore whether the two social media platforms Facebook and YouTube can be used to aid teachers in the challenging task of teaching pronunciation, any topic could in, theory, be used. However, word and sentence stress are good topics for several reasons.

First of all, pronunciation is a vast area of research with many possible angles of approach depending on the learners' L1. Much research has looked at how to help speakers of Japanese perceive and produce the distinction between English /r/ and /l/ (Akahane-Yamada et al., 1996; Gick et al., 2008; Yamada and Tohkura, 1992, among others). However, as the work for this thesis was carried out in Spain, it made sense to choose an area of focus that applies to L1 Spanish speakers. The above contrast (i.e., /r/-/l/) is not particularly difficult for Spanish learners. In contrast, previous work on Spanish speakers has shown that English vowels and other consonants have received considerable attention (Aliaga-García and Mora, 2009; Aliaga-García, 2009; Cebrian and Carlet, 2014; Iverson and Evans, 2009; Gómez Lacabex, García Lecumberri and Cooke, 2008; Thomson, 2016). Suprasegmental phonology has received comparatively less attention. As described in 1.4.3, there is an ongoing debate about whether segmental or suprasegmental errors influence a listener's perception more. It is clear that the misapplication of both aspects is detrimental to comprehension, intelligibility, and accentedness, but there is no agreement as to what should be prioritised in pronunciation teaching, although some researchers (e.g., Derwing, Munro, and Wiebe, 1998) recommend focusing on both. Hence, the need for additional research in the area of suprasegmental phonology is greater than for segmental phonology.

In support of focusing on suprasegmentals, Eskenazi (2009) stated that 'prosody is the backbone of speech, providing the structure that links the individual sounds to one another and to the linguistic substrate' (p. 837). An important part of guaranteeing prosodic speech that is intelligible is making sure the stress in one's utterances is placed where necessary.

As mentioned above, word and sentence stress are two topics which both seem to cause L1 speakers of Spanish problems because of the differences in how the two are applied in English and Spanish. Starting with word stress, from a purely theoretical point of view, one would predict that word stress could be an area of difficulty for L1 speakers of Spanish based on the models presented in Chapter 2 (the SDM and, in particular, the STM). This has been further emphasised by Ortiz-Lira (1998), who states that '[s]tress in English compounds, with its varying position, is a real stumbling-block for Spanish-speaking learners due mainly to the rather fixed occurrence word stress has in Spanish' (Ortiz-Lira, 1998, p. 28). Furthermore, as seen in 2.2.3, Spanish compound nouns are never stressed on the first element only as is the case for the majority of English compound nouns.

With regard to sentence stress, there are also theoretical grounds to believe that this area of English pronunciation could prove challenging for L1 speakers of Spanish. As described in 2.3.2, the vast majority of utterances in Spanish have the nucleus on the last word, whereas English nucleus placement is far more varied. However, studies with L1 speakers of Mandarin have shown that at least some aspects of sentence stress are learnable, so one would expect L1 Spanish speakers to respond well to instruction in this area. Also, García Lecumberri (1995) found that L1 Spanish speakers are, to some extent, able to move the nucleus on a sentence and deemphasise information in the tail of an intonation unit. Both of these are important for producing the three sentence types selected for Study 2.

Findings from research that has dealt with English and Spanish sentence stress suggest that there are particular areas of English stress placement that seem to cause trouble for L1 speakers of Spanish, and that marked tonicity in particular 'would have to be specifically and explicitly taught if students are to avoid or overcome the interference stemming from the contrastive nature of tonicity in English and Spanish' (Gutiérrez-Díez, 2005, p. 145). While contrastive stress is the only type of sentence stress of the three taught in Study 2 that is, strictly speaking, marked, the two other sentence types will most likely need to be taught specifically as well simply because they are not sentence-final.

It is worth reiterating that Spanish can make use of nucleus displacement. That means that this kind of stress is not a completely new concept, but rather a concept which the learners need to learn how to use in a different way (and/or more frequently) when speaking English. In this regard, input enhancement and explicit instruction seem like an ideal combination for facilitating this learning.

In addition to this, as shown in Chapter 2, word stress and sentence stress are both thought to play a part when it comes to avoiding communication breakdowns. For example, Hahn (1999) showed that misplaced sentence stress makes processing times longer for native speakers of English, which means that misplaced sentence stress makes the speaker more difficult to understand. Similar findings have been made for word stress. For example, Cutler and Clifton (1984) found that shifting stress in words can be detrimental to listeners. Their study also found that the situation is aggravated by a reduced vowel being produced as a full vowel. Considering that Spanish is a syllable-timed language, in which vowels are generally produced as full vowels, the importance of producing word stress correctly becomes clear. Specifically, regarding the stress in compounds, Dickerson (2004) states:

> [l]earners whose listening skills are tuned to recognizing constructions by their stress patterns and whose speaking skills enable them to render these constructions with appropriate rhythm-promoting stress

> alternations will greatly enhance their aural comprehension and oral comprehensibility. (p. 94)

Although the effects of the training will not be tested with regard to communication breakdowns, from an ethical point of view, it is important that the participants have at least a theoretical opportunity to benefit from their participation in the studies.

As a final reason to focus on these topics can be mentioned that focus on word and sentence stress can be applied to both ESL, EFL, and ELF. Although the latter sees word stress as a 'non-core' feature of English language teaching, its proponents do acknowledge the usefulness of teaching word stress as it provides the basis for teaching sentence stress, which is regarded as a core feature (Walker, 2010, Jenkins 2000, 2002). Thus, regardless of the approach a teacher follows with regard to teaching, the results of this thesis should be of some value.

Despite presenting learners with challenges, there is still sufficient evidence to conclude that both word and sentence stress can be successfully taught to L2 learners of English at CEFR B1 level (see 4.3 for a description of the participants in this thesis). This is important because a topic that has not been explored at all could yield a situation in which one would be unable to conclude whether a lack of improvement was down to method (i.e., the use of social media) or topic (i.e., the content being too difficult).

Hence, to sum up: for the purpose of this thesis, the areas that will be included in the instruction will be three types of compound nouns (place names; materials and ingredients; and objects and agents) taught through Facebook, and three types of marked tonicity (event sentences; objects of general reference; and contrastive focus) taught through YouTube.

3.3.3. Technological Considerations

From the earlier discussion, it is clear that L2 pronunciation teaching and L2 pronunciation learning are two highly complex topics to deal with because of the wide range of factors involved. In spite of this, research has clearly shown that learning new aspects of pronunciation in an L2 can be accomplished by most learners regardless of age. This learning can be facilitated – and in some cases enhanced – by technology. To some extent, it can be argued that whether the technology used improves on traditional learning or simply replaces it is not crucial. However, it is paramount that the technology is fit for purpose. As seen in Chapter 1, researchers suggested using computers for pronunciation training even before the turn of the millennium. In addition, Hincks (2003) – among others – argues that pronunciation is an ideal area of focus for out-of-class practice.

The first reason for investigating social media for pronunciation training is purely academic. There is currently a dearth of studies carried out outside the rigid context of the laboratory setting in the field (Lord, 2010; Wang and Munro, 2004). This study will contribute to the slowly emerging pool of studies that try to rectify this. The relative scarcity of studies on the effectiveness of social media in pronunciation teaching is somewhat of a conundrum as their popularity is so great that they are used as 'a primary means of communication for many students' (Mompeán, Ashby, and Fraser, 2011, p. 98). Hence, it seems only logical to explore how these tools can be used for pronunciation teaching.

Another reason, which was discussed in Chapter 1, is the issue of time constraints teachers face when teaching pronunciation. By using technology to take parts of the pronunciation syllabus out of the classroom, teachers will be less pressed for time. Given the right technology, this can be done without compromising the benefits that classroom teaching has to offer. This is because, as can be seen from

Table 3.2, technology-based and classroom-based instruction share many of the same features. It can further be gathered from the table that technology-based instruction adds features that are not available to students in the classroom.

Table 3.2.
Features of technology-based and classroom-based pronunciation instruction.

	T.E. Input	*NS input*	*Explicit instruction*	*Output practice*	*Unlimited access*	*Ideal context*	*Immediate Feedback*
Tech-based	Y	Y	Y	Y	Y	Y	**(Y)**
Classroom-based	Y	Y	Y	**(Y)**	**N**	**N**	Y

As a note on the table above, it should be pointed out that feedback in technology-based pronunciation teaching is not necessarily immediate. In cases where students work with ASR or any of the tools described in 1.5.1, technology will be immediate. However, in the present studies, for example, the feedback was delayed. It should be noticed, though, that the types of feedback differ in that the feedback described in 1.5.1 generally focused on the stimulation of the learners' sensory systems (e.g. visual or audio-visual feedback). In the studies presented in this thesis, on the other hand, the feedback was used to improve the students' meta-linguistic knowledge and raise their awareness of the features taught on the course.

With regard to output practice in the classroom, although it is most certainly possible, it takes place in conditions that some learners may find intimidating, which could lead to a lesser effect. Hence, allowing learners to study autonomously grants them a range of advantages in the sense that they can study when they have time, as well as study in an environment they find most comfortable. Being able to work in a comfortable space without added time pressure should lower the learners' affective filter and help them absorb the content of the course more effectively.

A very important feature which is lacking from classroom-based instruction is the unlimited access. If a student misses a lesson in the classroom (and attendance in some language schools is sometimes poor), the pronunciation instruction of that lesson will most likely be lost. If the instruction is available online, however, the students can access it when they have time and as many times as they want.

One feature that makes both Facebook and YouTube very interesting focal points for pronunciation teaching studies is the possibility of uploading videos as well as text. This means that it should be possible to use input enhancement (Sharwood Smith, 1993) in order to raise learners' awareness of the differences between their L1 and the target language. This, in turn, should help them notice the areas in their own L2 in which they may need to improve (Schmidt, 1990). Additionally, the two platforms allow for explicit instruction, be that in a video or text format. Because the learners will be encouraged to practice the dialogues presented to them, they will also have the benefit of production practice, which is believed to be an important part of learning to speak an L1 (Swain, 1985).

As described in Chapter 1, when using technology the learners can request to hear the input an infinite number of times without having to worry about how a teacher might react. Allowing the learners to hear the dialogues repeatedly could also be theorised to help improve their perception – which may lead to better production.

It must, of course, be added that the studies will not show that the social media platforms can be employed effectively for the teaching of all aspects of pronunciation. However, they are expected to show whether there is any point in exploring these platforms for other areas of pronunciation teaching.

Chapter 4 - Methods

This chapter details how the two studies for this thesis were conducted. The two studies followed a mixed-methods design. While the potential improvements in terms of word and sentence stress placement were analysed quantitatively, the students' experience of following the courses were analysed qualitatively as well as quantitatively. As both studies focus on the application of social media in English pronunciation teaching, they naturally share several features. These will be described together to avoid unnecessary repetition. Features that are unique to the individual study, such as stimuli and course design, will be treated separately. Thus, the recruitment process, the participants, and the instruments used are described first. After that, the specifics of Study 1 are dealt with. Finally, the specifics of Study 2 are described. However, as a first focal point, the somewhat uncharacteristic nature of the study, brought about by the Covid-19 pandemic, is described.

4.1 Overview

Study 1, an investigation of the acquisition of word stress placement in compound nouns, was carried out using a pre-test/post-test design. Between the pre-test and the post-test, the experimental group followed a four-week pronunciation course administered via the social media platform Facebook.

Study 1 and Study 2 were run a week apart from each other because testing all the groups in just one week proved unfeasible. Thus, the Facebook group started in mid-February of 2020 whereas the participants for the YouTube study started the week after as shown in Table 4.1. The control group was scheduled to be tested in mid-March, but due to the national lockdown an insufficient number of participants were available. This led to the control group doing their pre-test in November of 2020 instead, with a post-test the following month.

Table 4.1.

Tests and training conducted for Study 1 in 2020.

Part of Study						N
Pre-test (FB_1) Feb. 2020	Training (4 weeks)		Post-test (FB_1) March 2020			11
	Pre-test (FB_2) Feb. 2020	Training (4 weeks)		Post-test (FB_2) March 2020		2
		Pre-test (FB_3) Nov. 2020	Training (4 weeks)		Post-test (FB_3) Dec. 2020	23
		Pre-test (CTRL) Nov. 2020	NO training (4 weeks)		Post-test (CTRL) Dec. 2020	33

Note: N is the number of participants who completed the post-test.

As can be seen in the table above, after the initial round of testing and training, insufficient data had been collected for Study 1 (N = 11) due to extraordinary participant attrition. In an attempt to boost the numbers, a second course was run using participants who had initially signed up too late for the first study. However, this course also suffered participant attrition so even after two courses had been run, the amount of data obtained was still considered to be too small (N = 13). Consequently, it was decided to repeat the process a third time with participants doing a course on Facebook in the autumn semester of 2020. This was done a week apart from the tests of the control group.

Study 2, an investigation of the acquisition of non-final sentence stress followed essentially the same design as Study 1. The main difference between the two studies was that only one round of pre- and post-tests was needed. In studies conducted outside a global pandemic, it is probably most common to see the experimental and control group being tested more or less around the same time.

Unfortunately, this turned out not to be possible, which meant that about nine months elapsed from the time the experimental group took their pre-test to the time the control group took theirs (see Table 4.2 below). Even so, as was the case with Study 1, this study can be said to have followed a pre-test/post-test design with the experimental group taking part in four weeks of training delivered through the social media platform YouTube between the tests. The courses are described in 4.6.2 and 4.7.2.

Table 4.2.
Tests and training conducted for study 2 in 2020.

Part of Study				**N**
Pre-test (YT) Feb. 2020	Training (4 weeks)	Post-test (YT) March 2020		24
	Pre-test (CTRL) Nov. 2020	NO training (4 weeks)	Post-test (CTRL) Dec. 2020	33

Note: N is the number of participants who completed the post-test.

4.2 Recruitment

Initially, the idea was to recruit primarily from the language academies in Murcia. The reason for this was that a general problem in applied linguistics research, as well as in research in other fields, is that the participants used are almost always university students – be they undergraduate or postgraduate (Andringa and Godfroid, 2020). The issue here is that research often only yields findings that are immediately applicable to that particular subset of the wider population. With the relatively large number of private English academies in Murcia, it was expected that enough participants could be recruited for both studies. However, the recruitment process turned out to be much more complicated than expected.

Recruitment was initiated in September 2019 with the course expected to start a month later. However, after contacting close to 20 academies, only eight volunteers came forth, and the data collection date had to be postponed until February of 2020. Thus, a new round of recruitment was attempted. This time, the recruitment specifically targeted students enrolled at the University of Murcia's Campus de la Merced, which is a university campus in the city centre dedicated primarily to arts and humanities. The recruitment process is illustrated in Table 4.3 below.

Table 4.3.
Description of the recruitment process for both studies.

	Time of recruitment		
	September, 2019	*February, 2020*	*September, 2020*
Number of participants recruited	8	103	60
Target	Private academies. Flyers distributed on Campus de la Merced.	1st and 4th-year students on the English Studies degree at UMU.	2nd-year students on the English Studies degree at UMU.

Note: The numbers do not add up to the 144 in the text above because some participants were recruited twice.

Although the participants were not the first choice in terms of demographics, working with university students does come with the advantage of having a relatively homogenous population and, thus, fewer potential confounds.

The vast majority of participants were from various matriculation levels of the English Studies degree, which is a four-year degree taught in English. At the start of the degree, most students are at least at B1 level according to the Common European Framework Reference (CEFR); this is a commonly used framework in language testing in Europe. The relatively few participants who were not pursuing this degree had to undergo a brief level test, which consisted of a 15-minute interview with the researcher. This was only a test of their oral communicative abilities as these were the focus of the study. As for the students enrolled on the English Studies degree, it was assumed that their English level would be adequate for the purpose of understanding the instructions given on the course as well as working with the content of the course.

Although 103 students signed up, only 68 students took the pre-test. Furthermore, when the first wave of the Covid-19 pandemic hit in February/March 2020, and the university transitioned to an online mode of teaching, significant participant attrition followed. This meant that insufficient data had been collected after the first courses had run. For this reason, a final round of recruitment was done in September of 2020, which specifically targeted 2nd-year students of English Studies at the university. This yielded an additional 60 participants. Many of these participants had already participated in one of the two courses (as 1st-year students in the previous semester), but none of the participants did the same course twice.

Thus, across the three rounds of recruitment, a total of 144 participants signed up to participate in the study. Of these potential participants, 132 took the pre-test, and 123 completed at least a part of the post-test. Unfortunately, due to technical difficulties, some of the students were unable to complete the perception test. For example, one of the participants did not own a PC or a laptop at the time they took their post-test.

As described above, the vast majority of the dropouts were from the first round of testing. To this, it should be added that due to poor audio quality on some recordings as well as other technical difficulties (see chapter 6), only a total of 87 participants across the three groups provided data that could be used for the analysis. Included in the 87 participants were 32 control participants who provided data for both studies.

Although the vast majority were native speakers of Spanish, eight of the recruited participants reported to speak either Spanish as an L2 or have two L1s with Spanish being one of them. These participants were allowed to participate on an equal footing with the other participants, but their production and perception data was excluded from the data analysis. The one exception to this was an early bilingual Arabic/Spanish speaker. The data provided by this speaker was compared to the data obtained from the participants with just one L1 and was found not to differ significantly. The only difference was that this participant described the photos in the picture description task from right to left instead of left to right. Their speech samples were studied carefully and although there were very few traces of a Spanish accent in terms of segmental production, the suprasegmental features of the participant's speech did display some Spanish influences in terms of sentence stress.

Despite the fact that it would have been easier to accept L1 Spanish speakers only in the study, it was decided to allow anyone who signed up to take part. The reason for this was that many of the participants received academic credits for their participation, so it seemed unfair to deny any NNSs this opportunity to partake based on their L1.

4.3 Participants

This subsection describes the participants that took part in this study. In addition, although not strictly speaking participants, the judges who volunteered to rate the speech samples are also described.

4.3.1. Experimental and Control Participants

As mentioned above, a total of 123 total number of participants recruited completed the post-test. Because of the way the recruitment took place, 64 of the participants took part in the studies twice, but great care was taken to make sure that no one participated as an experimental participant in Study 1 in both the spring and the autumn semester.

Reward

The manner in which the participants were rewarded for their participation varied across and within groups. This was because the participants were recruited from a variety of courses at the University of Murcia as well as outside the university. Thus, in some cases, the reward for completing the study was five euros. In others, the participants were rewarded with course credits; this was by far the majority of participants. The number of credits they received depended on how much time they had spent on the course so that participants who only took the pre-test received fewer credits than participants who took both tests and completed all the exercises on the course. The exact number of credits were at the course tutors' discretion.

Background questionnaire

Before the participants started the study, they filled out a background questionnaire. The questionnaire was sent to the students immediately after they had taken the pre-test in order to avoid pre-test questionnaires from participants who failed to take the pre-test. The point of the pre-test questionnaire was to gather information about the participants' familiarity with technology and language learning as well as to obtain information regarding L1, educational background, and other basic demographic data. The background questionnaire can be found in Appendix I. The data reported here only covers the participants who completed the post-test (two participants who completed both the pre- and post-test failed to submit the pre-test questionnaires despite repeated reminders).

Much as one would expect, the demographic data gathered from the questionnaire shows that the participants formed a rather homogenous group. For example, in terms of age, the participants ranged between 18 and 29 years (M = 19.4). The main differences appeared in terms of gender as there was a significant majority of female participants in the study with only nine male participants completing the post-test. More male participants were recruited initially, but there seemed to be a much larger tendency for males than females not to complete the study. A few participants stood out by reporting only three to six years of formal English education. With the overall range being 3-20 years (M = 12.8), this was significantly below the rest of the participants. In spite of this apparent disadvantage, these participants achieved scores comparable to the other participants on both the perception and the production tests. Some of the data from the background questionnaire is summarised in Table 4.4.

Table 4.4.
Demographic data for the participants divided into groups.

	Facebook (N=36)	YouTube (N=22)	Control (N=33)
Gender (Female; Male)	33; 3	19; 3	30; 3
Age (mean; range)	19.9; (18-29)	18.7; (18-29)	19.8; (18-24)
Years of learning English (mean; range)	12.7; (3-20)	11.9; (5-17)	14.1; (3-20)
Importance of sounding like an NS (mean; range)*	5.7; (1-7)	5.9; (3-7)	5.7; (4-7)
Frequent users of the platform (at least once a week)	3	22	N/A

Note: N is the number of people who both returned the pre-test questionnaire and completed the post-test.

* Likert type question where, on a scale from 1 to 7, the participant was asked to what extend they agreed that sounding like a native speaker was important to them.

As mentioned above, participants with any L1 could participate, but only native speakers of Spanish and one early bilingual were included in the analysis. With regard to knowledge of additional languages, only two participants reported no knowledge of languages other than Spanish and English. The vast majority also spoke some French or German. It is possible that the two participants who reported not to speak any additional languages did, in fact, know some French or German as it is mandatory for the students to take either French or German classes as part of their degree. The participants' level of English was not tested directly, but instead, participants were asked to give a self-assessment of their level. Only one participant reported a level below CEFR B2. This participant reported a level of CEFR B1. However, based on the brief conversations the researcher had with this participant during the test, this was probably a rather low evaluation of their actual level. Similarly, 26 participants claimed to have a CEFR C1 level of English, but in many cases, it seemed as though they had overrated their own level. The remaining participants judged themselves to be at CEFR B2 level. It should be mentioned that it is likely the participants' self-evaluation was influenced by the English courses they followed as part of their degree. For example, in the first year of their studies, the students take Lengua Inglesa I + II which covers the CEFR B2. Similarly, in the second year, the students take Lengua Inglesa III + IV which covers CEFR C1.

Looking at the type of English instruction the participants had had in the past, the responses from the participants were rather mixed. Five participants said the main focus of their pronunciation instruction had been intonation and stress. About a third of the participants reported never having focused on pronunciation – apart from in their 1st and 2nd-year university courses. These are courses that specifically target segmental and suprasegmental aspects of English pronunciation as well a course in English phonetics and phonology, which all students must take as part of the degree in English Studies. The remaining participants had experienced an equal mix of segmental and suprasegmental pronunciation instruction.

In terms of language use outside the classroom, only four participants reported speaking English for more than an hour a day (between one and two hours); the remainder all reported speaking English less than an hour a day. When it comes to time spent listening to English, however, nine participants reported spending between one and two hours a day, with the remainder again spending less than one hour.

The final questions in the pre-test questionnaire inquired into the participants' familiarity with the social media platforms Facebook and YouTube. The participants in Study 1 varied rather much in terms of how much time they spent on this type of social media. Hence, nine of the 36 participants reported that they were not on Facebook before the course started. While this means that these participants would not use the platform as part of their e-routine, they might still provide information about whether Facebook is effective as a language learning tool. Of the 27 participants who already had a Facebook account, 17 stated that they logged on less than once a week. The remaining participants were more avid users and reported logging on at least once a day.

Unlike the participants in Study 1, all participants in Study 2 reported using the targeted social media platform (in this case YouTube) at least a few times per week (four participants). The majority of participants reported using YouTube several times a day (18 participants). They mainly used the platform to listen to music and other forms of entertainment. Only one participant stated that they used YouTube to learn languages.

Groups

A few options were considered with regard to the use of control and comparison groups. One option was to compare computer-assisted teaching to classroom-based teaching. An example of this is Neri et al. (2008), who showed that the ASR-based CAPT system was just as effective as traditional pronunciation teaching. Similarly, other studies have compared CAPT to traditional classroom-based teaching and generally produced positive results (see Chapter 1). This is a particularly important factor when judging technology for classroom use as this technology would be competing with other activities the teacher may have planned. However, when looking at technology that is used outside the classroom, in the learners' own time, simply seeing an effect can be seen as positive. For this reason, as well as for issues pertaining to recruitment and data analysis, a group of learners taught in the classroom was not included. In hindsight, this was fortunate as the coronavirus pandemic would have rendered data collection impossible.

As a result of the considerations above, the participants were divided into three groups: Study 1 (N = 58), Study 2 (N = 36), and Control (N = 38); the numbers presented here represents the number of participants who took the pre-test.

As mentioned, the control participants served as controls for both studies. It has been pointed out that when including control participants in language learning studies, these should also undergo some sort of treatment. This is because it would be unethical to offer a potentially useful intervention to some students but not to others (Lord, 2010). In the studies described here, however, the control participants were not asked to do anything. Instead, all the materials were made available online after the course. This way, the participants could work with them if they wished to do so. Thus, the control group did not receive training as such but were given the chance to work with the same materials as the experimental participants at a later time.

It is important to mention that, since some of the participants in the control group had already completed either Study 1 or Study 2, there was a risk that the control group's performance would be artificially good at the time of the pre-test. As described in chapter 5, however, there was no significant difference between the experimental groups and the control groups at this test time. As can be seen from Table 4.4. below, there was very little difference between the participants within and between the groups.

4.3.2. Judges

The elicited production data was judged by two native speakers and one non-native speaker, namely the researcher. Although it seems commonplace to use speakers of Southern British English in EFL studies conducted in a European context, the two judges used here were from the North of England (Liverpool and Manchester). These two judges were chosen because the researcher knew them well and, thus, knew that they were both highly qualified and very intelligible individuals. The latter was important because the two judges were also chosen to record both the test materials as well as the training materials.

The number of judges used varies greatly across studies and may depend on both availability of judges and the focus of the study. Studies of comprehensibility, intelligibility and accentedness tend to use more than other types of studies. For example, Derwing, Munro, and Wiebe (1988) used 48 judges to rate comprehensibility and accentedness; Derwing, Munro, and Thomson (2007) used 33 judges to rate fluency and comprehensibility; Saito, Trofimovich, and Isaacs (2015) used 20 raters to judge comprehensibility and accentedness. There are, however, studies that do not follow this pattern as Derwing and Rossiter (2002) only used six judges to evaluate the fluency and accentedness of their participants, and Ducate and Lomicka (2009) used two judges to rate their participants' comprehensibility and accentedness. Studies on specific speech sounds also seem to exhibit a large degree of variation with regard to the number of judges used. Barriuso and Hayes-Harb (2018) used 26 judges, Thomson (2011) used five, and Thomson and Derwing (2016) used only two. In this last case, it was the authors themselves who served as judges. It is thus clear that it is difficult to find any real consensus on the number of judges required. As very few studies have focused on stress placement, it is similarly difficult to find commonalities in terms of the number of judges used. Regardless of the number of judges, it is always preferable to have a high degree of interrater reliability as disagreement among raters might indicate that the raters did not understand the task they were performing (or that the quality of the data did not allow for accurate rating). For the two studies presented here, a total of three judges was believed to be sufficient due to the rigorous rating procedure (see 4.6.4 and 4.7.4).

With regard to the linguistic background of the judges, studies have generally used either NSs only or NNSs only. There seems to be some disagreement about what works best with some arguing that NS and NNS judges rate speech samples the same (e.g., Yoshida, 2004). However, others have found that NNS are stricter in their evaluation of L2 English speech if they speak the same L1 as the talkers they are rating (Koet and van den Bergh, 2018). Little is known about how NNS rate L2 English when the speech samples are produced by talkers with a different L1 to the raters, however. It is also possible that other parameters are important in how speech samples are rated such as whether the rater has a background in teaching or indeed phonetics. Finally, what is being rated could also be important in the sense that some aspects of language such as fluency, for instance, might be easier for NNS judges, whereas specific speech sounds and vowel contrasts might prove more difficult.

Thus, in the present study, a combination of native and non-native raters was used with the native speaker raters being the same as the ones who recorded the test and training items and the non-native speaker being the researcher (a non-native speaker of both English and Spanish).

It has been pointed out that EFL/ESL teachers are not ideal as judges because they are used to listening to foreign accents (Munro, 2008). However, as this study did not focus on intelligibility or accentedness, this should not present a problem. On the contrary, the raters' knowledge of ESL jargon might have helped to prepare them for the rating tasks as they were already familiar with concepts such as word stress, sentences stress, contrasts etc.

4.4. Instruments

The main instruments used for the creation and execution of the studies for this thesis were:

- NCH VideoPad (https://www.nchsoftware.com),
  - Used to create the course materials
- TP (Rauber, Rato, Kluge, and Santos, 2012)
  - Used for the perception tests
- Zoom (https://zoom.us)
  - Used for online testing
- Google Drive,
  - Used for pre- and post-test questionnaires and to assess weekly participation
- Facebook (Study 1),
  - Used for course delivery
- YouTube (Study 2),
  - Used for course delivery
- Praat (https://www.fon.hum.uva.nl/praat/)
  - Used for some parts of the data rating procedure to settle disagreements between raters.

4.4.1. Instruments for Course Creation and Delivery

NCH VideoPad is a commercial tool for creating videos. It can be trialled for free for a limited time, but a subscription was purchased as the tool was needed for a longer period of time as well as to enable the full range of features that the software offers. Its interface is relatively intuitive, and it was chosen among a wide range of video production software due to the many features available and general usability.

Despite these advantages compared to other similar types of software, the process of creating the videos turned out to be quite time-consuming. The initial idea for the study was to use materials that were quick and easy to create due to the already mentioned tight schedules that many language teachers have. Unfortunately, this was not achieved to the intended extent as the video creation process was far more strenuous than anticipated. However, because the content – once created – can be reused by other teachers and on other platforms, the cumbersome creation process becomes less of a problem. It should also be mentioned that the researcher used this software for the first time for this thesis, so it is possible that with more experience, the process will become smoother.

Facebook was used as the content management system (CMS). As described in more detail in Chapter 1, some of its features, such as the wide range of compatible files as well as the recently added Social Learning Units, made this an optimal choice for this study.

Similarly, YouTube was used as the platform for delivery in Study 2. While this platform has less of a feel of a CMS about it, as there are fewer features available compared to Facebook, it still possessed all the required features for delivering the course. In fact, the fewer features could be argued to have been an advantage because having fewer features can make the interface less confusing. For example, a Facebook group has several places in which content could be located. With YouTube, on the other hand, the content is generally available immediately. One drawback, however, was that it is not as easy to communicate with participants on YouTube as it is on Facebook. This is due to how notifications differ between the two platforms. Whereas Facebook users are notified every time someone posts in a group they are a member of, YouTube users only get notified if someone replies directly to one of their comments. A way around this issue would be for the teacher to email their students directly or flag up any lack of YouTube participation in the next lesson. Sending a reminder email only had to be done once. This was a reminder about the video posted in the third week of the study, which had only been viewed six times by the time the feedback was uploaded.

4.4.2. Instruments for Data Collection and Analysis

Google Drive was used for its questionnaire feature and served several purposes. Firstly, the questionnaires were used both as part of the course proper in that they contained the exercises the learners had to do each week. Secondly, they allowed the participants to report how much time they spent on the exercises, thus serving as a tool to monitor student participation. Finally, Google Drive questionnaires were used to collect background information about the participants as well as for the purpose of studying the participants' perception of the course.

It has been argued that questionnaires are much better suited for quantitative research than for qualitative research because, as Dornyei (2010) explains 'no matter how creatively we formulate the items, those are unlikely to yield the kind of rich and sensitive description of events and participant perspectives that qualitative interpretations are grounded in' (p. 10). However, Dornyei does concede that 'open format items can provide a greater "richness" than fully quantitative data' (ibid. p. 36). As such, open-ended questions were kept to a minimum and only concerned the most vital information regarding students' perception of their experience with the course. As will be shown in the analysis, some of the participants provided some interesting perspectives which would have been lost had the open-ended questions not been included.

Using Google Forms has several advantages. Firstly, the data that the students submit is stored in an online spreadsheet making it easy to manage (Firth and Mesureur, 2010). This makes the initial data analysis much quicker. Secondly, because Google Forms can tell you how many people have completed the tasks, it allowed the researcher to see if it was necessary to post reminders about the tasks. Thirdly, having the students submit their answers in this manner meant that there was no peer judgement. It was hoped that this would make the students more comfortable about completing the tasks. Finally, apart from the uses already mentioned, Google Forms has the additional advantage that students cannot see each other's submissions, and thus they will not be tempted to alter their own responses to better reflect the 'norm' in the group.

TP is a fairly simple piece of software, which is most often used for training and testing L2 learners' ability to perceive segmental contrasts in a foreign language using an HVPT regime (see Chapter 1). As this was not an HVPT study, the software was instead used for testing and rating. The software was deemed ideal for this type of study because it is easy to use for both the researcher and the participants. Furthermore, features such as randomisation of test items, repetition of test items, and data collection were very important for the design and execution of this study. Although the simplicity of this software was in many ways an advantage, it was also its major drawback in the sense that preparing tests for participants and raters alike was very time consuming as each individual speech sample had to be imported separately.

The video conferencing software Zoom (zoom.us) was used for the tests that could not be conducted in person. Zoom has gained significantly in popularity over the course of the Covid-19 pandemic and has in many cases replaced Skype as the preferred video conferencing tool. Like Skype, it allows users to talk (with or without video) and share files. Another feature which many teachers find useful is the option to share one's screen with other people in the meeting. Overall, the quality of Zoom seems to be better than what you get with Skype with fewer users reporting issues with lags or crashes.

Finally, Praat was used for some parts of the data analysis. Praat is a piece of software originally designed for phonetic research, but it has found other uses as well. For example, several people have suggested it as a tool for teaching pronunciation due to its visual features such as pitch tracing (e.g., Le and Brook, 2011). In the studies presented here, the software was used to settle disagreements between raters regarding stress placement in words or phrases.

4.5. Procedure (Commonalities)

This section deals with aspects that the two studies have in common. Certain elements of the two studies such as how stimuli were recorded, how the testing and ratings were carried out as well as all aspects of the data analysis were very similar. Thus, it makes sense to describe these in one place and treat other aspects which are particular to each individual study in separate sections.

4.5.1. Recording of Stimuli

All the stimuli were recorded by the two native speakers of English described in section 4.5.1. Two types of recording equipment were used. For the test and training items, all recordings were made using the software Audacity (audacityteam.org) on a laptop and a Marantz MPM-1000U microphone. This type of microphone was chosen because it is rather sensitive and thus allows people recording to sit at a comfortable distance. This was deemed advantageous when recording the test and training materials where two people had to use the microphone at the same time. A considerable drawback to this type of microphone is that, due to its increased sensitivity, there must be no ambient noise while recording. Unfortunately, ideal conditions were difficult to come by and thus several sessions in different locations at the University of Murcia’s Campus de la Merced were required. One positive that was taken from this experience was the realisation that a different microphone was needed to record the participants. Thus, to record the speech samples in the face-to-face tests, a Samson C03 microphone was used.

There was a rather large exception to this, though, as the control participants and about half of the experimental participants never met face-to-face with the researcher, and thus had to record all their speech samples on their mobile phones. Finding suitable recording conditions can be difficult at the best of times if a dedicated recording studio or similar soundproof facility cannot be accessed. For this reason, the participants recording their speech samples remotely were sent written instructions on how to ensure good audio quality on the recordings. However, judging by the video chats, it appeared to the researcher as though most of the participants did the tests in their rooms regardless of how suitable that room was.

After recording, all the stimuli were normalised to -6db using Audacity, and noise reduction was applied as well. Despite the subpar recording conditions, the final product was generally of high quality.

For the pre- and post-test for this study, the stimuli used were a selection of compound nouns with either single or double stress. A total of 35 items were recorded for the two tests. Of these 35 items, 18 had in common that one of the rules taught on the course could be applied to them, whereas the remainder were filler compound nouns with double stress. In addition to these items, another 23 items were recorded to be used in the training phase. The test items were all recorded in one of two carrier sentences, namely, *The next word is...* and *Now I say…*. Following the recording, the carrier sentences were edited out using Audacity, and the test item was saved as an individual .wav file. The only editing that was performed on the dialogues for the pronunciation course was the aforementioned normalization and noise reduction.

4.5.2. Testing Procedure

For the most part, the pre-tests were carried out at the University of Murcia. The exceptions to this were the last group in Study 1 and the control group, who both took all the tests from home as the Covid-19 pandemic made meeting face-to-face impossible. The participants took part in both a production test and a perception test, with the former being administered first. This order was chosen as some of the test items appeared on both tests, so to prevent the participants from hearing test items before they had to produce them, this order was the only sensible one.

For the face-to-face tests, the participants were placed in front of a laptop with a connected microphone. Before the production test started, the participants were asked to make sure they were sitting comfortably and that the microphone did not interfere with their view of the computer screen. For Study 1, all the stimuli were presented on a laptop. The dialogues for Study 2 were provided on two sheets of paper. The initial plan had been to present the dialogues on the screen as well, but a brief pilot test showed that this could lead the test taker to scroll down using the mouse while reading, and thereby create unwanted noise in the acoustic signal picked up by the microphone.

Before the tests started, the participants received brief instructions about how to take the test and were allowed to ask questions for clarification. Before the production tests, they were also asked whether they wanted the researcher to listen to their productions or if they preferred not to have him hear them speak. In the latter case, the researcher played music on his headphones which drowned out the participants' speech but was inaudible to the test-taker. This was made much easier in the online tests because of the 'mute microphone' feature which is built into Zoom.

The online tests were run in a manner as close to the in-person tests as possible. The participants received a Zoom link to their university email account 10-15 minutes before the test started. This method was chosen rather than using a personal Zoom channel for all tests to avoid participants logging in before their designated test time. Changes to the test format for each study are detailed in 4.6.3. and 4.7.3 respectively.

4.5.3. Data Analysis

Because the two studies were quite similar in terms of both procedure and design, the data analyses used in the two studies were also similar. The various tests in both studies yielded both normal and non-normal data. This section is divided into five parts. The first subsection discusses the assumptions behind some of the statistical tests used. The next subsection discusses *t*-tests and their non-parametric equivalents. Third a discussion of the test called analysis of variance (ANOVA) is provided. The last two subsections described the use of post-hoc tests and effect sizes respectively.

4.5.3.1. Assumptions

The statistical tests included in the General Linear Model (GLM) all have very similar underlying assumptions (Nimon, 2012). As for the *t*-test, the first assumption is that the data has to be normally distributed. To ensure that this assumption was met, the Shapiro-Wilk test was used (Larson-Hall, 2016). Another assumption, but one that only applies to the independent samples *t*-test, is that the data has an equal variance around the mean. To ensure that this was the case Levene's test was used in cases where the normality assumption was met; however, when this was not the case, Bartlett's test was used as this test is less sensitive to violation of the normality assumption (Hatchavanich, 2014).

The dependent and independent samples *t*-test also differ in that the former assumes that the independent variable consists of two categorial, related groups whereas the latter assumes that the independent variable consists of two categorical, unrelated groups. An additional assumption the *t*-test is based on is that the dependent variable is measured on a continuous scale. Strictly speaking, since the participants' performance was measured on a scale with 1-point intervals (1 point per correct answer), the data is not continuous. This is because all scores will be integers. However, there seems to be a tendency to treat test scores as continuous even so. In addition to these three assumptions, it is also assumed that there are no significant outliers in the data. There is no exact definition of the term 'outlier' so it is generally up to the researcher to decide when a value should be labelled as such (although 1.5 x the interquartile range is often used). Finally, an assumption that only applies to the independent samples *t*-test is that the observations must be independent. This means that that a participant cannot be allowed to be in both groups.

4.5.3.2 *t*-tests

In the field of applied linguistics, as well as in other branches of academia, Student's *t*-test is frequently used to compare two independent samples. As just mentioned, it assumes that the analysed data is normally distributed and that it has a homogeneous variance around the mean. However, as these assumptions are not always met, as Delacre, Lakens, and Leys (2017) argue that Welch's test should be used instead by default as 'the Welch's *t*-test provides a better control of Type 1 error rates when the assumption of homogeneity of variance is not met, and it loses little robustness compared to Student's *t*-test when the assumptions are met' (p. 92). While it is important to acknowledge that Student's *t*-test can be biased if the data does not meet the aforementioned assumptions, the fact that the authors seem to suggest that Welch's test does lose, at least, some robustness when the assumptions are met could be seen as an argument against using Welch's test as a default. This is especially true considering how easy running statistical tests has become thanks to modern-day software. This is most certainly the case when it comes to basic assumptions such as those of normality and homogeneity of variance as most statistical software packages can include the Shapiro-Wilk test and Levene's test by the click of a button. Thus, rather than using Welch's test as a default, the underlying assumptions were first tested using the aforementioned tools to ensure that the *t*-test was appropriate. In cases where neither assumption was met, the Mann-Whitney *U* test was used to compare unpaired samples.

Here it should be pointed out that although Mann-Whitney *U* is a commonly accepted non-parametric test, Karch (2021) has very recently suggested that the Brunner-Munzel test be used instead. He argues that just as parametric tests have underlying assumptions, so do non-parametric ones. What Karch is referring to is the exchangeability assumption (essentially the assumption of equal variance), which he argues is seldom satisfied, as for this to happen, the sample distributions being compared must be identical. However, considering how frequently used the Mann-Whitney *U* test is, and given the fact that other authors have argued that the test can be used with slightly different populations (e.g., Hart, 2001), the test was nonetheless used in some instances. It also has to be said that if one delves into the statistics literature, most tests have been questioned to some extent. It is obviously beyond to scope of this project to decide who is right in the current debates on statistics, and thus, it seems safer to adhere to the traditions in the field of applied linguistics.

For this thesis, the independent samples *t*-test is primarily used to look for differences between the experimental group and the control group prior to further statistical tests. This is a commonly used approach, but it does suffer from the drawback that absence of evidence is not evidence of absence. What is meant by this is that even if the *t*-test fails to find a difference between two groups, an undetected difference could still be present.

There seems to be relatively less debate in the literature about how to treat paired samples. Thus, for paired samples, a paired samples *t*-test was used when the underlying assumptions were met, and the Wilcoxon Signed Rank test was used when this was not the case.

It should be mentioned that as an alternative to non-parametric tests for non-normal distributions, some researchers prefer to transform their data to make it fit a normal distribution. This can be done by applying any function (most often the square root, log transformation, or inverse transformation) to each individual data point. A key aspect to consider in this case is that even though the data set is mathematically the same after transformation, it has nonetheless been changed, thus rendering interpretation of the data more difficult. Such a transformation is something that can influence inferences on effect sizes. Consequently, rather than transforming the data, the aforementioned tests were used.

4.5.3.3 ANOVA

The ANOVA has been the most used statistical tool in L2 research in recent years

(Plonsky and Oswald, 2017). The use of the ANOVA has become so prevalent that some researchers have argued that it is possibly used too much. This is because it is sometimes used because of its convenience and not because it is appropriate (Plonsky and Oswald, 2017). One issue is that researchers fail to acknowledge the assumptions that the ANOVA is based on. This criticism could in some cases be extended to include other members of the GLM. Thus, in order to make sure that the ANOVA is used appropriately, it is important to consider its underlying assumptions, which are very similar to the assumptions of other GLM tests. For example, the ANOVA assumes that the data is normally distributed and the variance around the mean is equal. Furthermore, the values of the independent variable must be independent, meaning that scores in one level of the variable do not influence scores in another level. As the independent variables here are 'Treatment' and 'Time', this assumption is met as no participant was included in two groups and as each participant was only tested once at each test time.

It should be mentioned, however, that the ANOVA may be applicable in some cases even when all the assumptions are not met. For example, simulation studies have shown that the ANOVA has some degree of robustness even when the normality assumption is violated (Schmider et al., 2010). Although the authors of this article hinted that the ANOVA might be robust against violations of other assumptions, only the violation of the normality assumption will be accepted here. If other assumptions are violated, a non-parametric alternative will be used. Such an alternative is easy to find for the one-way ANOVA as the non-parametric Kruskal-Wallis test is commonly used in place of the ANOVA (but see Zimmerman, 1998 for a discussion of the accuracy of non-parametric tests when either the assumptions of normality or homogeneity of variance is violated).

Unfortunately, this is not so straightforward when it comes to the mixed measures ANOVA as there is currently no non-parametric equivalent to this test. Instead, groups are compared using the non-parametric Mann-Whitney *U* test, and the test times are compared using the non-parametric Wilcoxon Signed Rank test. This does have the aforementioned drawback that interactions cannot be detected, but as these tests are only used a small number of times, it should not present a major issue.

The criticisms mentioned above do not mean that the ANOVA should not be used; they only serve as a reminder to researchers that depending on the data at hand, other methods may be more appropriate. For example, if the output data is closer to a Poisson distribution than a normal distribution, it is likely that a mixed model approach would be a better choice. Likewise, if a study looks at several independent variables, multiple regression is likely to be a better option. What is important to reiterate is that one of the main problems with how the ANOVA is used is that some researchers tend to use multiple one-way ANOVAs when running a different test would be a better approach (Plonsky and Oswald, 2017).

As mentioned, the studies conducted for this thesis focused on two independent variables. For this reason, a two-way mixed measures ANOVA was chosen, as this test requires a between-subjects factor (Treatment) and a within-subjects factor (Time). One of the strengths of this test is that it can shows whether there is an interaction between the variables. In other words, the mixed measures ANOVO can show whether the two factors being analysed affect each other.

4.5.3.4. Post-hoc Tests

The ANOVA described above can compare the different factors in an analysis and state whether there is a significant difference somewhere in the data set. However, it cannot specify where that difference is. In order to find this specific information, a post-hoc test is required. This is, in essence, an additional test used to specify the pair-wise relationship between the factors in the analysis.

Because a significance test is never 100% accurate (i.e., most tests use a 95% level of accuracy – also referred to as an alpha-level of .05) an analysis can become very misleading when several pair-wise comparisons are made. This is due to the family-wise error rate. For example, in a one-way ANOVA with just three groups the family-wise error rate is 14.3% (if using *t*-tests as the post-hoc test). To counter this error rate, a range of post-hoc tests can be used. A very simple post-hoc test is the Bonferroni test which simply divides the alpha-level by the number of pair-wise comparisons. However, this test is often seen as too stringent (Streiner and Norman, 2011). An often-used alternative is the Tukey test, which, while still being conservative and thus effectively controlling for Type 1 errors (Larson-Hall, 2016), is less conservative than the Bonferroni test. Hence, for the purpose of the studies in this thesis, the Tukey test was the preferred post-hoc test. As a final note, it should be added that the same assumptions apply to the Tukey test that apply to the ANOVA. This means that for the Tukey test to be reliable, normally distributed data is assumed as well as equality variance. As will be shown in Chapter 5, the assumption of normality was violated in a few instances. For these cases, the value of the post-hoc Tukey test was compared against the stricter Bonferroni test to ensure that its value was not inflated due to this violation.

4.5.3.5. Effect Sizes

An effect size (ES) is a value that says something about the relationship between two variables or the effect of an intervention (Durlak, 2009). ESs are often reported using values such as Cohen's d (see below) or eta squared (η^2), but it can also be measured in more accessible terms such as percentage of participants that improved following an intervention.

One of the advantages often mentioned when talking about Cohen's d is that the value is expressed in standard deviations. This makes it relatively easy to compare results across studies. However, the reliance on standard deviations is also drawback since it means that Cohen's d is sensitive to violations of the normality assumptions mentioned above. In other words, if the data is significantly skewed, Cohen's d is likely to be inaccurate.

In ethical terms, the silver lining is that if the data is skewed, the variance is higher, which means that the standard deviation is higher. This will lead to a lower d-value because the pooled standard deviation is used as the denominator when calculating Cohen's d. Hence, although using Cohen's d could be inaccurate when working with non-normal data, the effect will not seem artificially large.

Even so, for the results in this thesis, it was decided to report both Cohen's d and the effect size Mean Absolute Deviation (M|D|). This approach was advocated by Gorard (2015), who, apart from discussing the normality issue, argued that M|D| is more intuitive and already used on other branches of science. MAD is very similar to Cohen's d, but instead of using the pooled SD in the denominator, it uses the mean absolute deviation, which is the average distance each data point is from the mean.

4.6. Study 1: Word Stress

The four subsections below describe the aspects of Study 1 that were unique to this study. The four aspects were the stimuli, the course design, the test design, and the rating procedure.

4.6.1. Stimuli

Stimuli were chosen that were considered appropriate for L1 Spanish learners of English at CEFR B2 level. Various resources were used to determine the CEFR-level of the words, for example, englishprofile.org (Cambridge University Press) and oxfordlearnersdictionaries.com (Oxford University Press). Lexical items of an unfamiliar register (e.g., *denim dungarees* etc.) were deliberately left out, whereas items expected to have a higher degree of familiarity (e.g., *chocolate cake* and *Hope Street*) were included (see Appendix II for a full list). Items that were not necessarily familiar to the participants but had a high degree of transparency such as *lemon cake* were also included. A combination of the two e.g., *lie detector* was also used. In this case, *lie* is classified as level CEFR B1, *detect* (v) is classified as CEFR C1, but as *detector* has a Spanish cognate, the compound was nonetheless included. As for the place names, specifically, these were, for the most part, compounds that can be found in popular culture or compounds where the first element was a name (e.g., *Edmund Street*). One exception was *Brook Street* which was included as the sound/spelling correspondence of *Brook* should not cause any issues for L1 Spanish speakers. A selection of sample words is provided in Table 4.5. A hyphen (-) indicates that the CEFR level could not be found.

Table 4.5.

Sample words from Study 1.

Item	CEFR Level	Item	CEFR Level
Orange juice	A1/A1	Brooklyn Bridge	-/A2
Taxi driver	A1/A1	Downing Street	-/ A1
Lemon cake	A2/A1	Oxford Road	-/A1
Ham sandwich	A2/A1	Painkiller	A2/B1
Skyscraper	A2	Drug dealer	B2/B2

Phonetic context was also taken into consideration. Due to the phonotactics of Spanish, L1 Spanish learners of English often struggle with some of the consonant clusters used in English. For example, Spanish speakers have a strong tendency to insert an epenthetic vowel before an initial consonant cluster consisting of /s/ plus another consonant in foreign words so that *school* is pronounced [es'ku:l] rather than [sku:l]. However, as the focus of the study was whether the participants stressed the correct element of the compound rather than whether they stressed the first or second syllable, words containing consonant clusters were found to be acceptable. Similarly, although it is not uncommon for Spanish learners of English to leave out consonants in certain contexts e.g., producing *instant* as ['instan] (Swan and Smith, 1987), words of this type were nevertheless included, as elision of this kind is unlikely to affect the stress pattern of the word.

Apart from the auditory stimuli that were presented in the pronunciation course, the participants also received one specific pronunciation rule each week. These rules were adapted from various textbooks such as *Pronunciation in Use* (Hancock, 2003) and *English Pronunciation for Speakers of Spanish* (Gómez González and Sánchez Roura, 2016). Table 4.4 shows the pronunciation rules that were presented to the participants in the study.

Table 4.6.
Pronunciation rules used in the Facebook pronunciation course.

Week 1	Rule 1	Place names: In English, place names with *Street* have single stress. Most other place names have double stress.
Week 2	Rule 2	Materials and Ingredients: Most compounds naming materials or ingredients have double stress. However, compounds with *juice* and *cake* have single stress.
Week 3	Rule 3	Objects and agents: When the first element is the object of the action mentioned in the second element, the compound has single stress.

Note: Each rule was followed by 10 examples using videos with audio scripts.

The first two rules (i.e., Rule 1 and Rule 2) were chosen because place names and materials/ingredients are commonly used topics both in everyday conversation and in English language teaching. Furthermore, they are fairly simple rules as they only have one or two exceptions to memorise. That said, the use of *most* in these two rules was a deliberate choice as there are other compound nouns within both rules which may be included in the exceptions. For example, many compounds ending in *bread*, *oil*, or *milk* often have single stress, but their use seems less fixed than the exceptions included in the rule. This is particularly true for *oil* and *milk* compounds. Rule 3 is less straightforward because it uses the thematic roles OBJECT and AGENT. However, for participants who are doing their degree in English, this should not be too far beyond their capabilities. Also, the learners of this rule had the option to ask questions in the Facebook group if they struggled to understand the rule. This rule was chosen because, like the other two, it covers many everyday objects, and has even fewer exceptions than the other two. For this reason, the word 'most' was excluded. According to Ortiz-Lira (1998), there are exceptions to this rule such as *stage manager* and *school governor*, but these are relatively few and concern mainly specialised vocabulary.

4.6.2. Course Design

As mentioned previously, the pronunciation course lasted four weeks and was divided into four modules. Each week a new pronunciation rule was presented. An exception to this was the fourth week, which was used to summarise the content of the previous three weeks. Thus, the researcher created a Facebook group using an account specifically for the study, that is, not his personal Facebook account. As there are theories about the benefits of teacher self-disclosure on Facebook in terms of student engagement (Mazer, Murphy, and Simonds, 2007; Saylag, 2013), this variable was taken out of the equation by making the account strictly professional. Once the account had been created, the researcher created a group where the materials were made available when appropriate. An explanation of how to interpret the word ‘stress’ was pinned to the top of the wall to remind the students what the focus of the study was.

A relatively recent tool on Facebook is ‘Social Learning Units’ (SLU). The SLU is a tool that has been developed to make instruction on Facebook easier. Course content can be divided into individual units making it possible to create a very structured course that is easy for the learner to follow. Therefore, materials for each individual week of the pronunciation course were contained in a single unit, and a new unit was uploaded on the Monday of each week (see Table 4.7.).

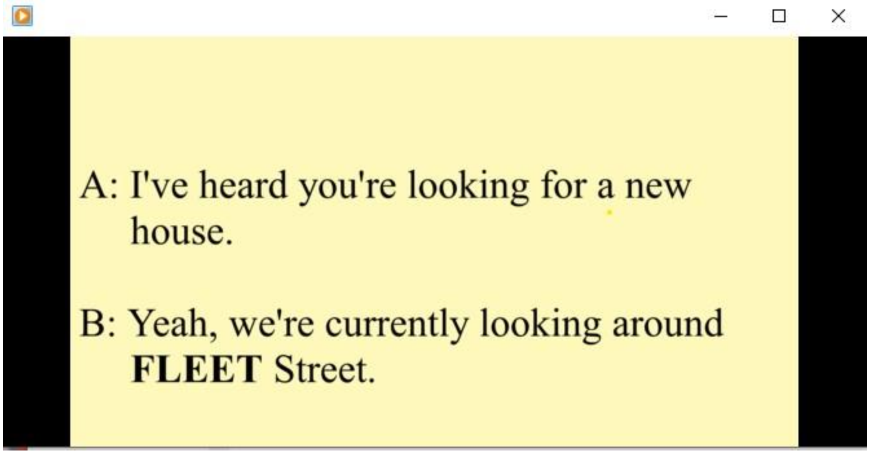
Table 4.7.
Structure of the four-week pronunciation course in Study 1 (Facebook).

Week 1	⇒	**Week 2**	⇒	**Week 3**	⇒	**Week 4**
(Place names)		(Materials/ingredients)		(OBJECT/AGENT)		(Recap)

The course materials were short videos created specifically for this course. Each video contained 10-13 short dialogues (audio script only) with a mixture of single- and double-stressed compounds. The task for the students was to identify the three compounds that were different from the rest and come up with a rule that explained why these were different. The dialogues used can be found in Appendix III.

The participants were encouraged to finish this task before the following Thursday, as this was the time when the answers were posted. The feedback thus consisted of a short explanation of the rule, which was posted in the Facebook group to improve the participants’ meta-linguistic knowledge of the topic (Ioup, 1995) along with a video repeating the dialogues uploaded. In the video posted with the feedback, orthographic features had been made more salient (Han, Park, and Combs, 2008) through the use of boldfaced text in capitals to help participants notice the difference in word stress. An example of this can be seen in Figure 4.1 below in which the first element of *Fleet Street* has been orthographically enhanced.

Figure 4.1.
Example of a short dialogue with enhanced orthographic features in Study 1.



For an ESL lesson, the students could be asked to post their answers in the group. However, personal experience with online courses suggested that students tend to copy and paste answers in a group without giving the questions much thought. This would work counter to the purpose of the exercise, namely that the students listen carefully to the input, thereby training their perception. Instead, the students were asked to use a link in the course module to a Google Forms page where they could submit their answers along with how much time they spent on the module.

With the answers, the participants were also encouraged to practise saying the words out loud by reading along with the audio. This was done because pronunciation practice should always be a component of pronunciation teaching (Fraser, 2001). It is important to note, however, that there was no fixed amount of time that the participants had to practice for. This was to make sure the study replicated a real-life learning context as much as possible (Foote and McDonough,

2017).

As mentioned above, each module included a task asking the participants to deduce a stress rule based on the input. This type of task should be particularly useful as learning seems to increase when the participants are told specifically what to focus on (Guion and Pederson, 2007). Although, on the surface, it may appear to be counter-intuitive to ask students to identify certain types of stress patterns when the assumption is that they struggle to do so, it was hoped that sufficient repetition of the items would help them hear the difference. As explained in Chapter 2, participants similar to the ones who took part in this study have been found to detect double stress in English compounds quite successfully (García Lecumberri, 2006).

4.6.3. Test Design

Perception

The participants underwent two tests: one receptive and one productive. The receptive test was administered through TP. As mentioned above, this software is most often used by researchers interested in segmental perception, but it was nonetheless deemed fit for purpose here. In the receptive test, the participants were asked to listen to and identify the stressed element in the compound nouns they heard. A very similar method was used by Archibald (1993), although Archibald's study focused on one-word lexical items and, therefore, had the participants choose the syllable they believed to be stressed rather than an element of a compound noun.

Although the TP software is usually used for perception studies, it is important to stress that the receptive test was not strictly intended as a perception test. Instead, the test was supposed to give the participants a chance to hear a compound noun, to decide whether it belonged to a category they had worked with during the course, and then to show their control of the rule by selecting the correct option – 'Single stress' or 'Double stress'. The TP interface can be seen in Figure 4.2 below. Apart from the two response options, the participants could replay each word twice, and also had the option to go back and change their most recent answer.

Figure 4.2.
Options presented to the participants in the perception test in Study 1.

Despite not being intended as a test of perception, the test will still be referred to as the 'perception test' for the sake of simplicity and because, as Chapter 5 will show, the training seemed to have an unexpected effect on the participants' perception of compound nouns.

Before the test started, the participants were given instructions as to how to take the test (see Appendix IV). They were also presented with four practice items and were invited to ask questions if anything was unclear. The tokens were presented in randomised order to avoid order effects. In the test, the participants were allowed to hear each token a maximum of three times.

The pre-test consisted of 34 compound nouns (see Table 4.8) of which 17 had single stress and the other 17 had double stress. The single-stressed items were all related to the pronunciation rules taught on the course and were thus either place names, ingredients/materials, or OBJECT/AGENT compounds (e.g., *bricklayer*). The double-stressed items were not all related to the rules since, prior to the tests, it was expected that the participants would not find these compounds problematic. In hindsight, and especially considering the results of the production test, it might have been worth using all relevant double-stressed items.

Table 4.8.
Test items divided into test and stress categories.

	Pre-test only	Pre- and post-test	Post-test only
Single stress	8	9	8
Double stress	9	8	9
Total items Pre / Post	34		34

The post-test also consisted of 34 words. A total of 17 words were part of the pre-test, and 17 words were new items. The latter were used to test whether the participants were able to use the rules learnt on the course to correctly predict/perceive the stress pattern of new novel compound nouns. None of the items in the post-test were used in the four-week pronunciation course. The post-test consisted of nine single-stressed words from the pre-test plus an additional eight single-stressed items. In addition to these, eight double-stressed items from the pre-test plus nine new double-stressed compounds were also part of the post-test. As TP collected all the perception data, it would have been easy to include all the pre-test items in the post-test, but it was decided to keep the test relatively short to avoid the participants getting bored with the activity, as this could affect the results.

Production

The production test was carried out by having the participants read out words from several PowerPoint slides. Because PowerPoint does not allow the user to randomise the slides, eight different PowerPoint presentations were created and used in random order. Each slide in the presentation contained a heading which served as a carrier sentence (either *Now I say…* or *The next word is…*) and one of the target compound nouns.

Figure 4.3.

Sample slide from the pre-test in Study 1.

It was decided to use two carrier sentences rather than one to avoid the participants getting too comfortable with the format of the test. The idea was that if the participants were forced to consider the carrier phrase as well as the target item, they would be more likely to produce natural-sounding tokens.

The ineffectiveness in eliciting natural speech is just one of the criticisms that have been levied against this type of test (Colantoni, Steele, and Escudero, 2015). Furthermore, this design has been argued to be inappropriate because what is tested is learning rather than acquisition (Krashen, 2013). However, from an information-processing point of view, testing whether the participants were able to learn the rules is vital because, in this view, declarative knowledge can turn into procedural knowledge (Segalowitz, 2003). In fact, due to the short treatment time, it can be argued that participants are unlikely to have turned what they learnt on the course into procedural knowledge. This means that a test that allows them to draw on their declarative knowledge is better suited for the purpose.

Other types of elicitation tasks were considered. For example, Thomson and Derwing (2016) elicited spontaneous versions of target items from their participants by showing them pictures of the nouns and asking the participants to make a sentence with the noun. While this worked well with their target words (simple nouns), a similar approach might prove difficult with compound nouns. This is because as a picture of a carrot cake, for instance, could be described by the respondent as either a cake, a fruit cake or a tray bake.

Trofimovich and Baker (2006) used a delayed sentence repetition task where the participants first listened to a question and a reply produced by NSs, then heard the question again as a prompt for their own production. This type of test has the drawback that participants may rely on their phonological short-term memory (Mora and Darcy, 2017), which is not the focus of this study. Thus, a read-out-loud task was chosen because it offers the participants the time to think about the words they produce and because it is very accurate in the sense that participants almost always produce the item they are prompted to.

Before the recording began, the students were allowed to practise with four dummy items which were not recorded. It should be mentioned that due to the Covid-19 pandemic, it was not possible to do the post-tests on the university premises. For this reason, the participants took the tests in their own homes. While this is not an ideal situation, it is not a scenario that is entirely unheard of either in the literature (e.g., García Lecumberri, 1995). The difference between the test design here and García Lecumberri (1995) is that, in the present study, only one of the tests was taken at home.

Due to the concern that the data set would be overwhelmingly large in terms of the subsequent data analysis, only a subset of the items in the pre-test were carried over to the post-test. This, of course, meant that not all items could be compared in the analysis. However, as it turned out that the filler words needed analysing too, this was probably the right decision.

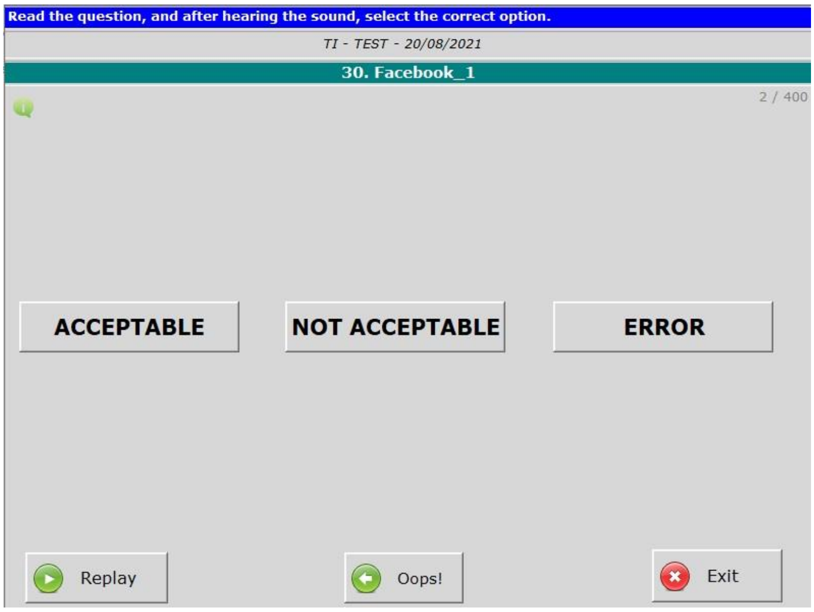
4.6.4. Rating Procedure

To prepare the speech samples for rating, each sample was edited using Audacity. Each individual compound noun recorded by the participants was normalised to -6dB and noise reduction was applied. Due to the background noise on some of the files, noise reduction occasionally had to be applied several times, but great care was taken to ensure that the quality of the speech was not affected significantly. This yielded 2748 individual files, which were loaded into the TP software described above (4.4.2) in order to create several tests for the judges to rate.

The tests presented one compound noun at the time and gave the judges two response options: 'Acceptable' and 'Not acceptable' as shown in Figure 4.4 below. Initially, the response options were 'Single stress' and 'Double stress', but to align the response options with those presented in Study 2 (see 4.7.4), 'Acceptable' and 'Not acceptable' were chosen.

Figure 4.4.

Options presented to the judges in the TP software.



Additionally, the judges had the option to report an error. This was added due to the large number of test items in each test. Hence, if an item of poor audio quality or an item containing unintelligible speech was presented, the judges could report that using this button. Apart from the response options, the raters were able to repeat the item twice and were also able to go back and change their most recent response.

To determine whether a speech sample was acceptable or not, the researcher first rated all the items twice. The ratings were compared, and disagreements were analysed in more detail using Praat, as this tool can give visual representations of each speech sample. In most cases, however, there was no need for a detailed analysis as it was immediately clear which of the two ratings was the correct one. Next, the adjusted ratings were compared to the ratings of the NS judges, who each rated a subset of the data. Any disagreements between the researcher and the NS judges were discussed and a final decision was made. In a few cases, Praat was once again used as the raters disagreed even after a third listening.

4.7. Study 2: Sentence Stress

Due to the many similarities between the two studies conducted for this thesis, some of the aspects of Study 2 will only be mentioned in passing in cases where an in-depth description has already been provided for Study 1.

4.7.1. Stimuli

As described in Chapter 2, there are considerable differences between English and Spanish stress placement. These differences contribute heavily to the difficulties many L1 Spanish learners have when perceiving and producing English. One major difference between the two languages is that whereas nucleus placement in Spanish is rather fixed, it shows a large degree of variation in English. Hence, three sentence types with non-final tonicity were chosen as the focus of this study. Two of these sentence types have been found to be particularly problematic, namely objects of general reference and event sentences (Ortiz-Lira, 1995). As explained in Chapter 2, objects of general reference are items with limited semantic weight referring to things or people (e.g., *someone, stuff, people*). Event sentences consist of a subject and an intransitive predicate (e.g., *The kettle is boiling*). The third sentence type included in this was contrasts. The terms contrast is fairly intuitive and refers to the highlighting of an item usually in opposition to another item which may be implicit or explicit. An example would be a sentence like *Henry BOUGHT the painting,* where *bought* emphasises that the painting was purchased and not obtained in some other way.

As in Study 1, the stimuli used for the tests and the stimuli used for the pronunciation course differed slightly from one another. The test stimuli for the speech production tests consisted of 27 short dialogues for the participants to read out. Eighteen dialogues were used for the pre-test. Of these 18 dialogues, nine were repeated in the post-test to which another nine new dialogues were added. Table 4.9

summarises the number of items used in each of the tests.

Table 4.9.
Type and number of items in each test in Study 2.

	Pre-test		Post-test	
	Perception	Production	Perception	Production
Objects of general reference	6	6	9	6
Event sentences	6	6	9	6
Contrasts	6	6	9	6
Neutral tonicity	6	0	6	0

Note: Picture description task not included

Sentences with stress on the final word were not included in the production test as it was expected that this would be the default stress pattern the participants would use. As shown in 4.A and 4.B below, although none of the sentences had a sentence-final nucleus, the number of words in the tail of each sentence differed significantly. This was done to make sure the participants would not spot a pattern in how the sentences were produced.

(4.A) *It was my MUM who won the lottery.*

(4.B) *I asked for RED roses.*

In addition to the dialogues, seven pairs of photos of toy animals and everyday objects as well as two photos depicting a clock which showed times 10 minutes to and 10 minutes past the hour respectively were used to elicit spontaneous speech in the pre-test. For the post-test, an additional three pairs of photos of stuffed animals and everyday objects were used (see Figure 4.5).

Figure 4.5.

Example of pictures used in picture description task in Study 2.

Unfortunately, the pair of photos depicting clocks had to be removed as none of the participants seemed to understand how they were supposed to contrast the two (i.e., the participants seemed unable to tell the time in English).

For the perception pre-test (as shown in Table 4.9 above), 24 individual four-word sentences were recorded: six of each of the aforementioned three types of sentences as well as six four-word sentences with stress on the final word. The last group of sentences was only used to reduce the likelihood that the participants might select the correct answer by chance. However, due to how the participants appeared to perform on these items during the test, they were also included in the data analysis. For the post-test, an additional three sentences of each type were added to test generalisation to new items. All the stimuli were recorded in the same sessions using the same talkers and the same equipment that were used for the recording of the stimuli for Study 1. Before the recording began, the talkers received a brief explanation of the difference between the four sentence types. Although the talkers were somewhat familiar with all four, the instructions were nonetheless provided to eradicate any doubt and to make absolutely sure the talkers knew what their task was.

The materials for the pronunciation course consisted of a total of 30 dialogues; 10 for each of the first three weeks of the course. Using Audacity, test as well training items were normalised to -6dB and noise reduction was applied. Each dialogue was saved as an individual .wav file and later imported into NCH VideoPad along with its audio script to allow for later manipulation of the font for textual enhancement.

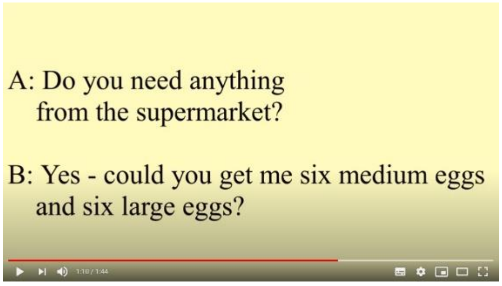
4.7.2. Course Design

This course also lasted four weeks with each of the first three weeks introducing a new rule and with the fourth week giving a recap of the content that had been presented during the course. On the Monday of each week of the course, the researcher uploaded to the YouTube channel training material consisting of ten short dialogues. To avoid the answers to the tasks being revealed on the channel, user comments were disabled. Although this did remove some of the social element from the experience, it seemed important to take this measure as the intention was for the participants to work out the answers for themselves.

As a first step, students listened to the dialogues (two to three exchanges) with the audio script provided on the screen. The reason for choosing short dialogues rather than individual sentences was that several authors suggest that suprasegmental aspects of a language are best taught in context (Goodwin, 2014; Hardison, 2005; Levis and Pickering, 2004). It is thus very likely that training that focuses explicitly on stress would also be more effective if taught in context. Although one additional sentence only creates limited context, it should help the learner better understand the meaning of the reply. As a second step, the students completed a task set in the video description asking them to identify the word with the greatest prominence (stress) in the last sentence of each of the dialogues. The final step in the procedure was the delivery of feedback, which took place on the Thursday of each week. The feedback came in the shape of the same dialogues, but with the stressed words textually enhanced in all dialogues (see Figure 4.6) and with the stress rule added in the video description. The audio was used to convey examples of sentence stress and to provide a model the participants could use to guide their own practice. The intention of the enhanced visual input was that it should help guide the participants' attention to the stressed element in the sentence (Sharwood Smith, 1993) so that if they had failed to spot the stressed word in the first video of the week, they would surely spot it in the second. Judging by the answers to the weekly questionnaires, this was a very necessary additional feature as the majority of the participants appeared to find the exercises challenging.

Figure 4.6.

Dialogues with and without textual enhancement.



Even though the dialogues were fairly simple, and thus should not require large amounts of cognitive capacity for the students to process, studies have shown that language learners tend to focus on meaning rather than form (Trofimovich and Gatbonton, 2006; van Patten, 1994). This is another reason why it makes sense to draw attention to the form of the sentence. Finally, the instructions provided an explicit explanation of the topic in each lesson to help make students aware of the rule in question, as explicit learning has been found to be more effective than implicit learning (DeKeyser, 2003). It should be added that explicit instruction about a topic can also be seen as input enhancement (Sharwood Smith and Truscott, 2014; Sharwood Smith, 1981). Taken together, this combination of elements was expected to create a good learning context for the students. Once the feedback with the explicit rule had been uploaded, the students were encouraged to listen to the dialogues multiple times and to practise producing them. It was briefly considered to ask the students to use shadowing (Foote and McDonough, 2017), but this idea was abandoned because it would have required giving the participants instructions about how to do that type of training. Instead, it was decided that it would be better to let the individual participant decide how they preferred to practise with the materials.

Finally, in addition to doing the exercises mentioned above, the participants were also asked to provide the course week, their participant number, and the amount of time they had spent working with the input (including the time spent repeating the dialogues). This served two purposes: 1) It functioned as an administrative tool allowing the researcher to see who had completed the exercises and 2) It was hoped that making students report the time they had spent on the task would make them spend a bit longer on the exercises. This last point could have been moot, as the participants could return a high value just to look good without actually putting in the work.

4.7.3. Test Design

A total of three different types of tests were used for this study. Apart from the perception and production tests that were used in the previous study, this study also included a picture description task.

Perception

As in Study 1, the perception test was carried out using TP (Rauber et al., 2011). Although TP is most frequently used for experiments with one-syllable words or non-words, it has been used successfully with sentence-level items as well (e.g., Brawerman-Albani et al., 2017). In this perception test, the participants were presented with individual four-word sentences and were asked to select the word they believed carried the nucleus. Using whole dialogues was initially considered in order to make the production task and the perception task more similar, but a potential issue with using whole dialogues is that if the participants mishear the first part of the dialogue, this might influence how they choose their answer. Instead of selecting actual words, the participants were asked to choose a number corresponding to the word in the sentence. Hence, if they believed the first word had the nucleus, they would select 'ONE', if they believed it was the second, they would select 'TWO', and so on. The interface is depicted in Figure 4.7 below.

Figure 4.7.
Options presented to the participants in the perception test in Study 2.

Before the test started, the participants received some brief instructions about how to take the test (see Appendix IV), and they were allowed four practice sentences to help them understand how the test worked. They were also encouraged to ask questions before the test began. As the test contained relatively few items, there were no breaks between the start and the end of the test.

Production:

The production test consisted of two types of tasks. Following Ortiz-Lira (1995) and Ramírez Verdugo (2006), a set of dialogues covering the three sentence types under investigation were written for the participants to read out loud. Additionally, a picture description task was used to elicit spontaneous speech. Although reading the dialogues in pairs would have made it more realistic, it was decided to let each participant do the dialogues individually, as pair-reading would make the design susceptible to participant attrition. Furthermore, the presence of an interlocutor could influence the speech production of the talker (Colantoni, Steele, and Escudero, 2015).

Before the dialogues were recorded, the participants were allowed to read through the materials at their own pace but were not allowed to take any notes to assist them in the reading task. Once they had finished reading, they were asked if any of the vocabulary was unclear or if they had any other questions. Although this type of test does not test the production of the items in spontaneous speech, it does provide valuable insight into whether the students have acquired sufficient knowledge about the English language to apply specific stress patterns correctly. Furthermore, the following picture description task elicited spontaneous speech so nothing was lost by using a read-out-loud task. On the contrary, it can be argued that using the two types of tasks together paint a better picture of the participants' development between pre-test and post-test.

As for the picture description task, this task required the participants to contrast a series of pairs of photos depicting toy animals and everyday objects. This test was adapted from Levis and Muller Levis (2018). Before this test started, the participants were told that they were about to see a series of photos which they had to compare. Specifically, they had to state what the difference between to two photos was. It was also pointed out to them that they should only use one sentence per photo. To help the participant get started, each slide came with the heading *In the picture on the left* which they were encouraged, but not forced, to use. Although this meant that the participants' speech production was not completely spontaneous, it was nonetheless seen as needed in order to give the participants some structure to rely on. After receiving the instruction, the participants were allowed to practise with two pairs of photos and were again encouraged to ask questions.

In the pre-test, the visuals were presented to the participants on a laptop using PowerPoint as shown in Figure 4.8. As in Study 1, several combinations of slideshows were created to avoid order effects.

Figure 4.8

Sample slide from the pre-test in Study 2.

Picture description tasks may be timed (e.g., Munro and Mann, 2005) or untimed (e.g., Levis and Muller Levis, 2018). Adding time constraints is often done in studies of morpho-syntax to make the task more cognitively demanding (Saito, 2012). However, as the aim of the task in this study was to discern whether the participants had managed to learn the given pronunciation rules, the participants were allowed as much time as they needed to describe the pictures. The reason for this was that because the intervention was only four weeks, it was seen as unlikely that the participants' ability to produce contrasts would have been automatised already meaning that they would have to rely on declarative knowledge to perform the task. Because the post-test had to be conducted online, the students watched the PowerPoint presentation on their own screens. To avoid materials being shared between the participants ahead of time, the researcher ran the slideshow on his laptop and used Zoom's screen sharing feature to allow the participants to view the visuals. The participants had the option to leave their microphone on or mute it depending on how comfortable they felt about the researcher hearing them speak in English. Participants who chose to mute their microphone were asked to use a hand signal, agreed upon before the test started, to be presented with the next slide.

4.7.4. Rating Procedure

The rating procedure was similar to the one used in Study 1 in the sense that all the collected speech samples were edited and prepared for later use in TP. For the read-out-loud task, this meant that each dialogue was edited so that only the reply was present. Between the experimental and the control group, seven tests were created with between 250 and 288 items in each test for a total of 1682 items. This was considerably fewer items than for Study 1, but as each item took longer to rate – for the simple reason that the rated tokens were sentences instead of words – the time spent rating was about the same for both studies. As in Study 1, the raters listened to a mixture of pre-test and post-test speech samples so that there was no way of telling which test each item belonged to.

The rating options for this task were again 'Acceptable' and 'Not acceptable'. This was not the first choice, however, as the initial rating design used more categories such as 'Correct', 'Incorrect', 'Incorrect contrast', and 'Don't know'. While discussing this design with the raters, it was decided that this was rather complicated, and thus, the current 'Acceptable/Not acceptable' dichotomy was used as it made both rating and scoring easier without losing too much information.

Before the rating procedure started, the raters received a copy of the dialogues with the stressed words printed in bold. It was important that the raters had a chance to read the dialogues to ensure they had the full context. For example, the sentence *I asked for red ROses* is a perfectly acceptable sentence, but would have had to be rated 'Not acceptable' for this purpose of this study, as it appeared in a contrastive context. The raters also received a reminder of the three sentence types they were going to listen to and could ask questions if anything was unclear. One issue was flagged up by both NS raters, namely, the sentence *20 prisoners escaped.* Although both NS judges found that stress on the last word would be more natural, they were asked to rate final stress on this item as 'Not acceptable'.

The speech samples for the picture description task were edited in a similar fashion to the read-out-loud task. The editing process for this part of the study was slightly more time consuming than expected. Each speech sample was edited so that only the description of the last of the two pictures was left. This was done for two reasons. Firstly, it was hypothesised that having the description of the first picture there might influence the extent to which the rater heard a contrast. Secondly, this was a time-saving exercise as only one sentence was needed from each participant. The speech samples were further edited so that any long pauses, fillers, or repeated words were removed. Whether the pauses were considered long was a subjective decision made by the researcher at the time of editing. Pauses were only removed if the participant spoke in the same key before and after the pause. This decision was made because a sudden change in key could make the speech sound unnatural and influence the rater.

A few options were considered for the rating of the extemporaneous speech samples. Initially, the idea was to let the raters look at the images while listening to the sentences and then state whether the speech sample matched their expectations based on the photo. This was thought to resemble the approach used by García Lecumberri (1995), who presented her raters with speech samples from her participants and asked them to pick a sentence that matched the speech sample. This idea was abandoned, however, as it could lead the raters to give poor ratings to samples where the stress was correct, but the grammar or semantics was wrong. For example, some of the participants confused prepositions and produced sentences such as 'The tiger is beLOW the glass' rather than 'The tiger is beHIND the glass'. Both of these sentences should be rated as 'Acceptable' for the purpose of this study.

Another issue that had to be resolved was the fact that the introduction of new images in the post-test meant that stating whether a sentence was correctly stressed or not was no longer a simple task. This is because whereas in the pre-test, all images related to differences in prepositions, the three additional items in the post-test looked at the participants' ability to stress nouns. This is illustrated in Figure 4.9 where the two pictures on the left would elicit a contrast between the nouns *dog* and *tiger*, whereas the pictures on the right would elicit a contrast between the prepositions *on* and *in*.

Figure 4.9.
Example pictures used to elicit a contrast in Study 2.

As a solution to this, the two sentence types were put into different tests in TP and the raters were given specific instructions as to how to rate each test. In other words, the raters were told to focus on either nouns or prepositions depending on which test they were rating.

Chapter 5 – Results

This chapter provides an overview of the results obtained from the various tests. The results, including a selection of responses from the post-test, from Study 1 (word stress) will be presented first. The second half of the chapter is dedicated to the results from Study 2 (sentence stress). This will also include post-test questionnaire responses.

5.1 Facebook in English Pronunciation Teaching (Word Stress)

This section presents the results from Study 1. More specifically, the participants' ability to learn the pronunciation rules will be described first followed by an analysis of the perception and production test results of the single-stressed items. Next, the test results for the double-stressed items are analysed. As a final point before the analysis of the data from Study 2, the questionnaire data from Study 1 is described.

5.1.1 Acquisition of Pronunciation Rules and Overall Improvement

One of the main foci of this study was the degree to which the participants were able to learn the pronunciation rules that were presented to them as part of the pronunciation course. This was done simply by asking the participants to state the rules as part of the post-test questionnaire. The participants' answers were rated so that each correctly stated rule received one point with an extra point added if the exceptions were also included. This led to a maximum score of 'five' as no exceptions were included for Rule 3 (OBJECT/AGENT). In some cases, despite most of the participants studying English at university level, the answers were so unclear that no meaning could be extracted from them. These answers received a score of zero. A breakdown of the scores is provided in Table 5.1 below.

Table 5.1.
Score frequencies for post-test questionnaire item 21 (Please state the rules you learnt on the pronunciation course).

Score	0	1	2	3	4	5
Frequency	6	3	4	5	4	7

Note: The participants were given example words to help them remember the rules.

The results obtained here were rather unexpected as it had been predicted that the participants would perform better on this task due to the relatively simple rules they had to learn.

To get an overview of how the participants performed overall, the perception and production results were analysed using two-way mixed measures ANOVAs. In order to ensure that the experimental group and control group were similar when the experimental group started their training, separate independent samples *t*-tests were performed in the pre-test data for both the perception and the production tests. Prior to this, the *t*-test's assumptions of data normality and homogeneity of variance were tested. The Shapiro-Wilk tests showed that the assumption of normality was not violated (Perception: p = 0.296; Production: p = .080). Similarly, the following Levene's Tests showed that the assumption of equal variance was not violated either (Perception: p = .068; Production: 0.959). Despite the groups being of slightly different sizes, the p-values of the independent samples *t*-test suggest that there were no significant differences between the two groups (Perception: p = 0.529; Production: p = 0.276) as shown in Table 5.2 below. This was an important finding as the students for the experimental group in this particular study were recruited across cohorts of the English Studies degree at the University of Murcia. Thus, one could have feared that the experimental group would be much better at pre-test than the control group, whose members were all recruited from the second year.

Table 5.2.
Pre-treatment comparison of the experimental and control group using all perception and production data in Study 1.

	N	Shapiro-Wilk	Levene's	*t*-test
Perception	35/29	$p = 0.296$	$p = .068$	$p = 0.529$
Production	25/22	$p = .080$	$p = 0.959$	$p = 0.276$

Note: The N-column represents the number of participants in the experimental/control group

As no significant differences between the two groups were found, further analysis of the data could be carried out. The following subsections will look at the results in more detail.

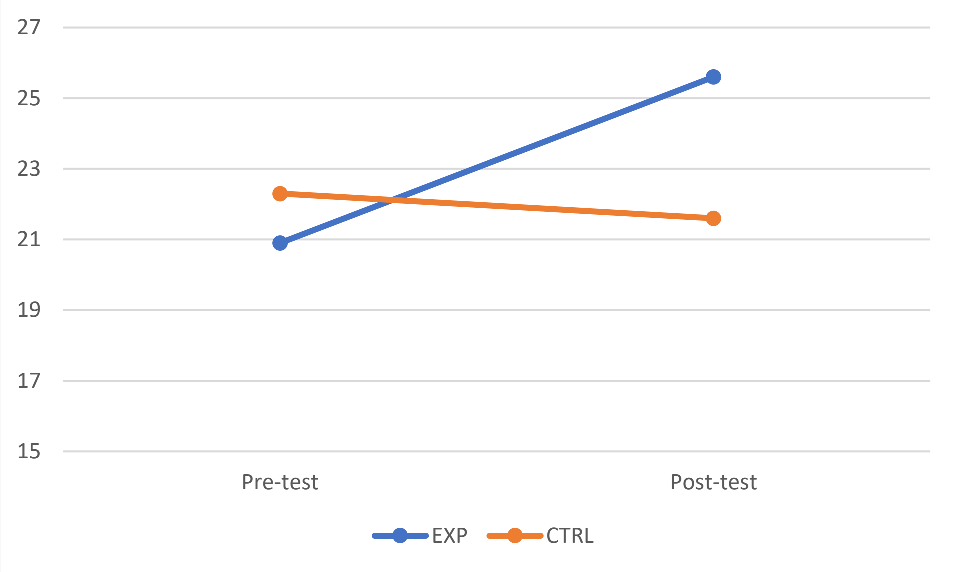
5.1.2. All Test Items in Study 1

Perception

To make sure the groups were comparable, an independent samples *t*-test was used to look for differences between the experimental group and the control group using the pre-test data. The t-test (Shapiro-Wilk: $p = .296$; Levene's: $p = .068$) found no significant difference between the groups ($p = .529$). As a result, their improvement between pre-test and post-test was analysed using a two-way mixed ANOVA. The ANOVA found a significant interaction between the Time and Treatment ($F(1,124) = 13.32$; $p < .001$) as well as a significant main effect of Time. The post-hoc Tukey tests revealed that the experimental group made statistically significant improvements ($p < .001$; $d = 1.12$; $M|D| = 1.43$). The control, on the other hand, made no such improvements ($p = .913$; $d = 0.17$). The development of the two groups is shown in Figure 5.1.

Figure 5.1.

Perception: Development across all items from pre-test to post-test in Study 1.

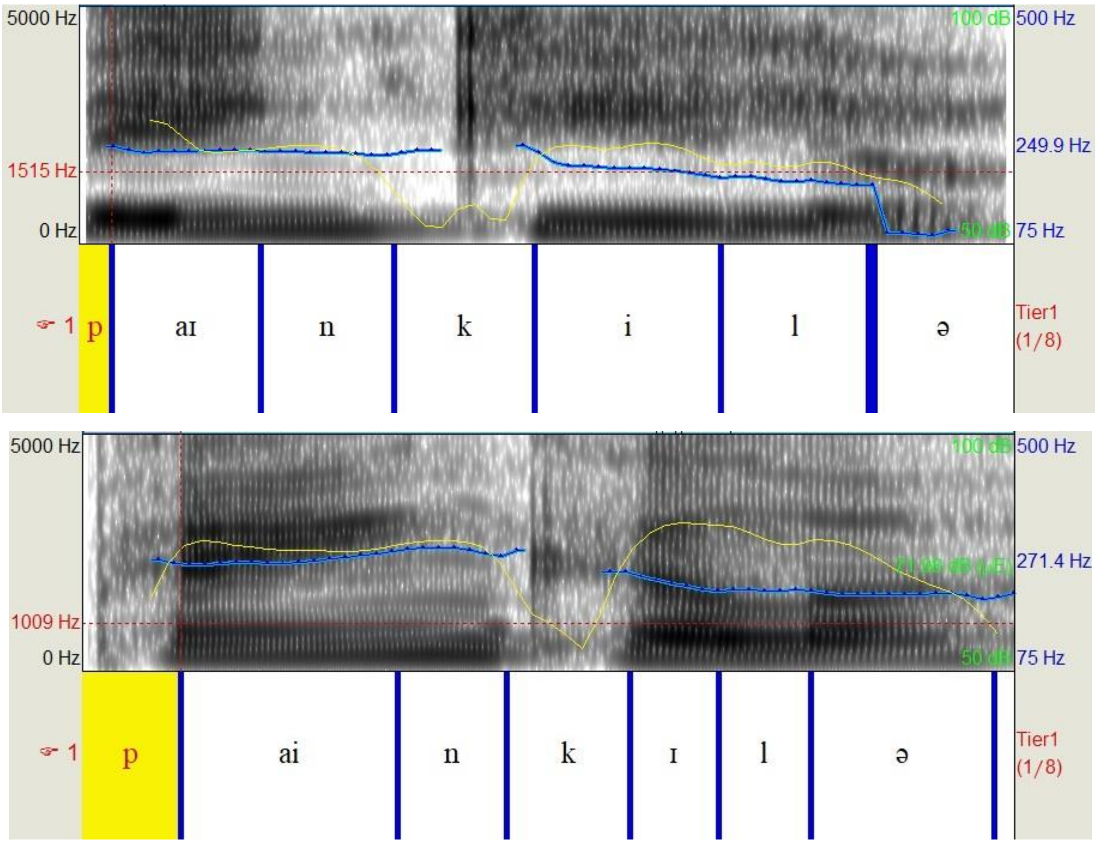


Production

Deciding whether word stress is produced correctly or not is not an easy task as the line is sometimes very fine and, essentially, a matter of interpretation. As described in Chapter 4, the items were rated by the researcher himself as well as by two native speakers of English. When all the raters had finished their ratings, the scores were compared. In a few cases, the disagreement simply stemmed from the wrong option being chosen on one rater's part – possibly due to a lapse in concentration. These could easily be corrected simply by listening to the items a second time. After an agreement had been reached, Praat was used to analyse the item to further confirm the decision that had been reached. In no cases did the visuals contradict the raters' decision. In other cases, Praat was used to solve disagreements that persisted after the second listening. This was done using the pitch contour and waveform (as a measure of intensity) as these tools can give a good indication of whether a compound noun has single or double stress. As mentioned in Chapter 2, vowel length could also have been included, but the two other cues were believed to be sufficient. In a few cases, though, even the visuals in Praat proved inconclusive. As a case in point, one of the instantiations of *painkiller* was a case of disagreement. As shown in the spectrograms obtained with Praat in Figure 5.2 below, the NNS produces only a very slight drop in F0. This could give the impression of double stress. However, the slightly higher intensity (yellow line) in the beginning of the first part of the compound could give the impression of single stress. Furthermore, the NNS pauses slightly between the two elements whereas the NS makes a seamless transition between the two elements. The slight pause is likely to help give the impression of double stress. Thus, it is almost impossible to draw any firm conclusions as to the nature of the given stress pattern. As a result, the few cases of this type were decided by the 2:1 majority among the raters. Fortunately, there were only a few cases which had to be decided by a third rater.

Figure 5.2.

Painkiller: Comparison of productions by an NNS (top) and an NS (bottom).



Note: The blue line indicates pitch movement.

Cohen's Kappa was used to calculate the interrater reliability. The values showed that there was only moderate agreement between the raters for both the experimental group (Kappa = 0.47 – 0.55) and the control group (0.60 – 0.77) as shown in Table 5.3.

Table 5.3.
Interrater values for experimental and control group (Study 1).

	Experimental Group		Control Group	
Raters	*R1 v R2*	*R1 v R3*	*R1 v R2*	*R1 v R3*
Agreement	73%	77%	80%	88%
Cohen's κ	0.47	0.55	0.60	0.77
p-value	< .001	<.001	< .001	< .001

As shown in Table 5.4, intra-rater reliability was also calculated based on a subset of the tokens (20 items). For this, Cohen's Kappa was used as this tool compares two raters assessing the same items. The intra-rater reliabilities were all substantial except for one value for Rater 3. These figures probably say more about the data than they say about the judges. As mentioned, some items were borderline cases where it was difficult to tell if the compound was stressed correctly or not. It is likely that items that produced the low intra-rater reliability were simply very difficult.

Table 5.4.
Intra-rater values for experimental and control group (Study 1).

	Experimental Group			Control Group		
Raters	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*
Agreement	95%	95%	65%	95%	90%	90%
Cohen's κ	.90	.90	.30	.0.90	.79	.90
p-value	< .001	< .001	.178	< .001	< .001	< .001

Similar to how the perception data was treated, the production data was first subjected to an independent samples *t*-test along with the corresponding assumption tests. Neither of the assumptions of normality or homogeneity of variance was violated (Shapiro-Wilk: $p = .080$; Levene's: $p = 0.959$). The *t*-test further revealed no reason to assume the groups were different at the onset of the Facebook course ($p = 0.276$).

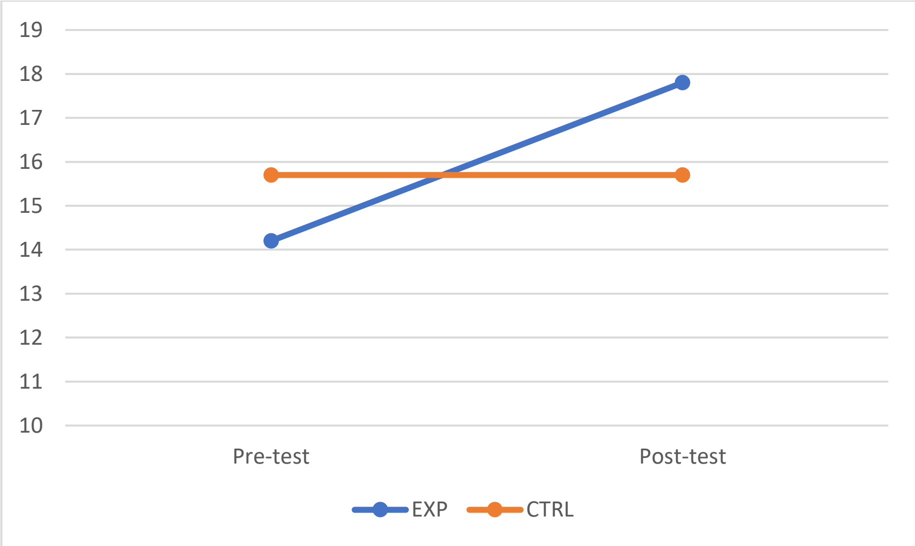
The first test that was conducted was a two-way mixed measures ANOVA to compare the two groups' performance on all the items in the pre-test and post-test regardless of stress pattern. Although not all items were the same on both tests, the comparison still makes sense. This is because an improvement in the experimental group would indicate an overall effect of the training, provided that no improvement was found in the control group.

The ANOVA found a significant interaction between Time and Treatment ($F(1,88) = 5.586$; $p = .020$) as well as a main effect of Time, but no effect of Treatment ($F(1,88) = .09$; $p = 0.757$). Due to the presence of an interaction effect, it is not possible to interpret the main effect of Time. However, this is not a major problem as the interaction effect in itself is a very useful finding. A post-hoc Tukey test showed that the experimental group made significant improvements from pre-test to post-test ($p < .01$; $d = 0.99$; $M|D| = 1.28$), whereas the control group made no improvements ($p = 1$, $d = 0.00$; $M|D| = 0.00$).

As Figure 5.3 below indicates, the mean score in the control group remained the same (15.7), but the experimental group's mean score increased from 14.2 to 17.8. Taken together with the fact that the standard deviation between pre- and post-test remained more or less the same ($SD_{pre} = 3.88$; $SD_{post} = 3.00$), there is clear evidence that the training was effective. The following subsections will look closer at this effect by dividing the test items into categories based on stress patterns and familiarity.

Figure 5.3.

Production: Line graph depicting the development from pre-test to post-test on all test items.



Having looked at all the items in both the perception and the production test, the next step in the analysis was to look more closely at each type of item – starting with the single-stressed items.

5.1.3. Single-stressed Items (Perception)

An initial analysis of the data showed that eight of the participants scored close to the maximum (i.e., within two points of the maximum score). In cases like this, it is not uncommon to exclude these participants from the data to avoid a ceiling effect. However, on inspection of the post-test data, this did not seem to be an issue, so all participants were included from the two groups.

To ensure that the ANOVA was an appropriate tool, tests of normality and equality of variance were once again conducted. The Shapiro-Wilk test returned a p-value of .001 thus indicating that the data was not normally distributed. However, as Bartlett's test returned a value of 0.156, the ANOVA was used even so as this test has proven quite robust to violations of the normality assumption (see 4.5.3.3).

As a result of the outcomes of the tests above, data from the two groups was compared using a two-way mixed measures ANOVA with Treatment as the between-subjects factor and Time as the within-subjects factor. The ANOVA found a significant interaction between Treatment and Time ($F(1,124) = 5.42$, $p = .022$), but the result of the following Tukey tests found no significant change from pre-test to post-test in either of the groups although the p-value for the experimental group approached significance ($Tukey_{exp.}$: $p = .052$; $d = 0.62$; $M|D| = 0.78$; $Tukey_{ctrl}$: $p = .859$; $d = -0.21$; $M|D| = -0.27$). It should be added that because the data was not normally distributed, the Bonferroni test was also applied. As this test is stricter than the Tukey test, it is not surprising that the test result was further from significance ($p = .064$) than that of the Tukey test (but see Chapter 6 for a discussion of this result).

At this point, however, it is worth looking at some of the descriptive statistics since the mean values indicate that the experimental group did improve as their pre-test had a mean score of 11.7 and a post-test score of 13.5 out of 17. In comparison, as can be seen from Figure 5.4 below, the mean score of the control group decreased slightly from 12.0 in the pre-test to 11.4 in the post-test.

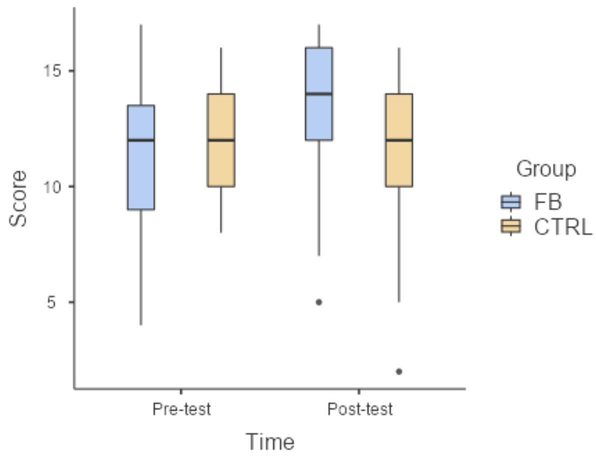
Figure 5.4.

Perception: Line graph depicting the development from pre-test to post-test on all single-stressed items in Study 1.

What is further noteworthy is that the standard deviation of the experimental group remained almost the same (3.12 vs. 2.96) whereas the standard deviation for the control group increased from 2.26 to 3.49. This latter finding could indicate that some of the participants' scores changed dramatically from pre-test to post-test – an idea that seems to be supported by the outlier in the boxplot below. However, whether this is coincidental is not easy to ascertain.

Figure 5.5.

Perception: Boxplot showing the development from pre-test to post-test on all single-stressed items.



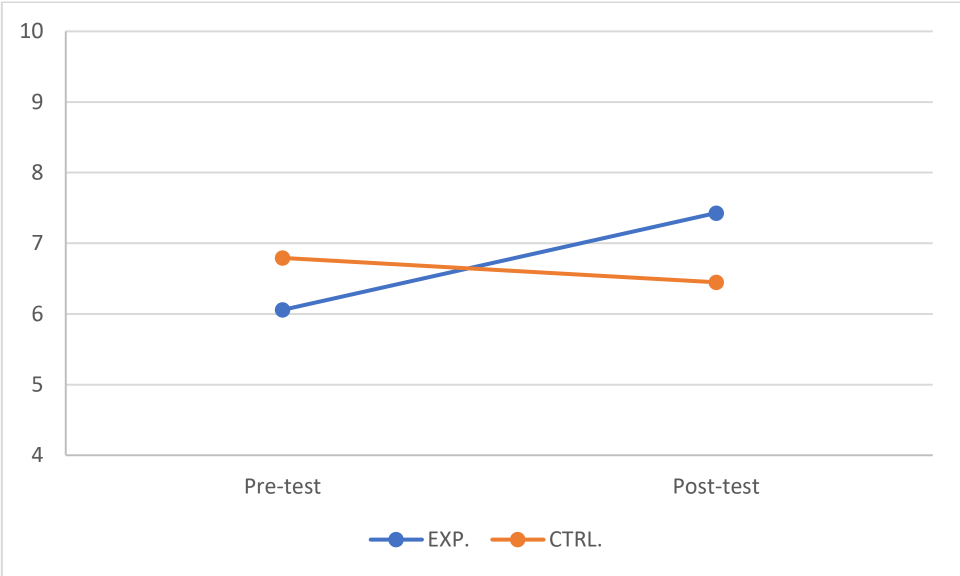
Note: Outliers are only observed in the post-test.

As for the outlier in the experimental group (Figure 5.5), it should be mentioned that this is not due to a drop in performance. In fact, the participant who achieved this score did slightly worse in the pre-tests. As such, the outlier, in this case, could be explained in one of two ways. It could be that the outlier is just a result of how the software creating the chart is programmed to define outliers; as this test was run in Jamovi, the outlier is defined as 1.5 x interquartile range. Alternatively, it could be explained by the other participants in the group either improving more from pre-test to post-test or scoring higher despite improving less.

The next step in the analysis was to look at the single-stressed items that were found on both the pre-test and the post-test (i.e., familiar single-stressed items). Similar to the tests in the previous step, the data was found to be non-normal (Shapiro-Wilk < .001), but equality of variance could be assumed (Bartlett's = .487). As can be seen from Figure 5.6 below, the ANOVA once again found a significant interaction between the two independent variables (F1,124) = 6.922; p = .009). However, unlike what was the case in the previous test, the post-hoc Tukey-test this time found a significant improvement for the experimental group (p = .011; d = 0.75; M|D| = 0.94), which was confirmed by the Bonferroni test (p = .013) and still no significant improvement for the control group (p = 0.891; d = -0.19; M|D| = -0.26).

Figure 5.6.

Perception: Line graph depicting the development from pre-test to post-test on familiar single-stressed items (Study 1).



As a final element of this analysis, the participants' performance on the new single-stressed items was compared to their performance in the pre-test and to their performance on the familiar items in the post-test also. As Table 5.5 shows, the experimental group scored 9% higher on the new items on the post-test than they did on the pre-test (67% → 76%). Although this was an improvement, it was much less pronounced than the 16% difference between the pre-test score and the score on familiar items in the post-test (67% → 83%). In contrast to this, the control group performed slightly worse (2%) on the new items than they did on the pre-test items (64% → 62%). When looking at the familiar post-test items, on the other hand, the control group showed an improvement of 8% (64% → 72%). These results could thus indicate that simply taking the pre-test helped the students to some extent in the post-test (see Chapter 6 for a discussion).

Table 5.5.
Perception: Comparison of familiar and unfamiliar single-stressed items at pre-test and post-test in Study 1.

	Pre-test	Familiar, post-test	New, post-test
Experimental Group	67%	83%	76%
Control Group	64%	72%	62%

5.1.4. Double-stressed Items (Perception)

Although the filler words were not initially planned to be a part of the analysis, the participants' performance during the tests seemed to deviate noticeably from what the researcher expected. That is, the first listening of the production tests and a quick glance at the perception results indicated that the participants had not performed as well on double stressed items as one would have expected given the stress patterns of Spanish compounds. As a result, the filler words were included in the data analysis in order to shed light on whether the training affected the participants' ability to perceive and produce English compound nouns.

The filler words can be divided into two categories, namely words that appeared on both the pre- and the post-test (familiar words), and words that only appeared once on either of the two tests. As a first step, all the items were taken into consideration. Again, a two-way mixed measures ANOVA was used (Shapiro-Wilk = .138; Levene's = .165). The test showed that there was a significant interaction between Time and Treatment ($F(1,124) = 8.71$; $p = .004$) as well as a significant main effect of both Time and Treatment. A post-hoc Tukey test showed that the Treatment group improved significantly ($p < .001$, d = 0.1.02; M|D| = 1.18) whereas the control group did not show any signs of improvement ($p = 0.99$, d = -.02; M|D| = -0.04). This translates to an improvement of 16% (Pre-test: 55%; Post-test: 71%) with the mean score rising from 9.26 at pre-test to 12.10 out of 17 at post-test (with SDs of 3.38 and 2.68 respectively). The two groups' development is depicted in Figure 5.7 below.

Figure 5.7.

Perception: Line graph depicting the development from pre-test to post-test for the experimental group and the control group on all double-stressed items.

These results are in stark contrast to the ones obtained for the single-stressed items in that the pre-test clearly did nothing in terms of improvement in the post-test because the control group showed no improvement. This becomes obvious when looking at the means and standard deviation for the control group: mean 9.48 (SD = 2.21) at pre-test, and mean 9.41 (SD = 2.51) at post-test.

The next step in the analysis was to focus on the familiar filler words. A two-way mixed measures ANOVA (Shapiro-Wilk = .129; Levene's = .379) was performed to compare these items. The ANOVA showed a significant interaction between Time and Treatment ($F(1,124) = 9.97$; $p = .002$) and also as a significant main effect of Time. The following Tukey test showed that the experimental participants' improvement from pre- to post-test was statistically significant ($p < .001$; $d = 1.17$;

M|D| = 1.33). In comparison, the control group showed very few signs of improvement (p = .99; d = 0.04; M|D| = 0.06).

As an extra measurement, the same pre-test/post-test statistics were calculated for the unfamiliar double-stressed items. Although it is hard to draw the same firm conclusions for items that are not the same, the calculations do make some sense on account of all the items presented on the tests shared the same stress pattern. Furthermore, if the items differed in terms of difficulty (e.g., due to word frequency) this should be reflected in the results of both groups, whereas an improvement only in the treatment group would have to be assigned to the effect of the training. Thus, similar to the previous two cases, a two-way mixed measures ANOVA (Shapiro-Wilk: p = .02; Bartlett's: p = .577) was run to assess potential main and interaction effects. The ANOVA found a significant main effect Treatment ($F(1,124)$ = 6.52; p = .012) but no interaction between Time and Treatment nor a main effect of Time. However, the post-hoc Tukey tests showed that the experimental group's improvement was not statistically significant (p = .089; d = 0.57; M|D| = 0.69), and that the control group showed a non-significant decrease in performance (p = .991; d = -0.08; M|D|= -0.09).

Finally, the participants' performance on the double-stressed unfamiliar items was compared to their performance in the pre-test as well as their performance on the familiar items in the post-test. As before, the performance was calculated as the percentage of correct items. In the pre-test, the experimental participants displayed an improvement of 22% (47% → 69%) on the familiar items. On the unfamiliar items, however, the improvement was 35% (47% →82). The data for the control group showed very limited improvement on the familiar items with an increase of just 1% (52% → 53%). Similar to the experimental group, the control group improved more on the new items, with an increase in post-test score of 13% (52%→65%). The data is shown in Table 5.6.

Table 5.6.

Perception: Comparison of the familiar and unfamiliar double-stressed items at pre-test and post-test in Study 1.

	Pre-test	**Familiar, post-test**	**New, post-test**
Experimental Group	47%	69%	82%
Control Group	52%	53%	65%

Interestingly, both groups performed better on the new double-stressed items than they did on the familiar double-stressed items, which indicates that simply hearing the word once in the pre-test did not help the participants in the post-test.

In summary, the perception tests showed that the experimental group generally made statistically significant improvements between pre-test and post-test. Table 5.7 summarises these results. It is worth noting that the experimental group also produced medium to large effect sizes ranging between 0.57 and 1.12. Additionally, the standard deviation in the post-test was lower in the post-test, which means that there was less variation between the participants at the time of the post-test.

Table 5.7

Summary of the results of the perception tests in Study 1.

	Perception							
	Experimental Group				*Control Group*			
	Pre	**Post**	**Impr.**	**d**	**Pre**	**Post**	**Impr.**	**d**
***M*(All items)**	20.9 (4.39)	25.6 (3.88)	**4.7**	1.12	22.3 (3.14)	21.6 (5.14)	-0.7	-0.17
***M*(All SS-items)**	11.7 (3.12)	13.5 (2.96)	1.8*	0.62	12.0 (2.26)	11.4 (3.49	-0.6	-0.21
***M*(Familiar SS-items)**	6.06 (1.94)	7.43 (1.82)	**1.37**	0.75	6.79 (1.52)	6.45 (2.01)	-.34	-0.19
***M*(All DS-items)**	9.26 (3.38)	12.1 (2.68)	**2.48**	1.02	9.48 (2.21)	9.41 (2.51)	-0.07	-0.02
***M*(Familiar DS-items)**	3.71 (1.79)	5.54 (1.56)	**1.83**	1.17	4.17 (1.31)	4.24 (1.53)	0.07	0.04
M(Novel DS-items)	5.54 (1.98)	6.54 (1.60)	1	0.57	5.31 (1.83)	5.17 (1.63)	-0.14	-0.08

Note: Standard deviations are listed in parentheses below each mean score.

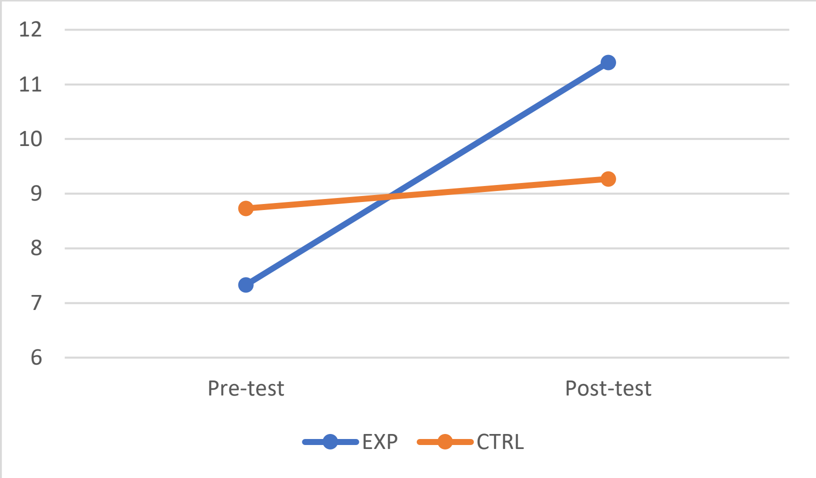
*See Chapter 6 for a discussion of this result.

5.1.5. Single-stressed Items (Production)

Having seen that the experimental group had improved overall (see 5.1.2), there was reason to believe that at least some improvement would be observed. The first subset of the data that was looked at was all the single-stressed items in the test. Once again, the first step in the procedure was to check the assumptions of normality and equality of variance. The Shapiro-Wilk test showed that the former was violated (p = .011), but Bartlett's test found equality of variance could be assumed (p = .329). Hence, a two-way ANOVA was used to analyse the data for all single-stressed items. The ANOVA found no interaction between Time and Treatment but did find a main effect of the Time variable ($F(1,88) = 4.306$; $p = .041$). As can be seen from Figure 5.8, the experimental group improved a great deal more than the control group.

Figure 5.8.

Production: Line graph depicting the development from pre-test to post-test across all single-stressed items (Study 1).



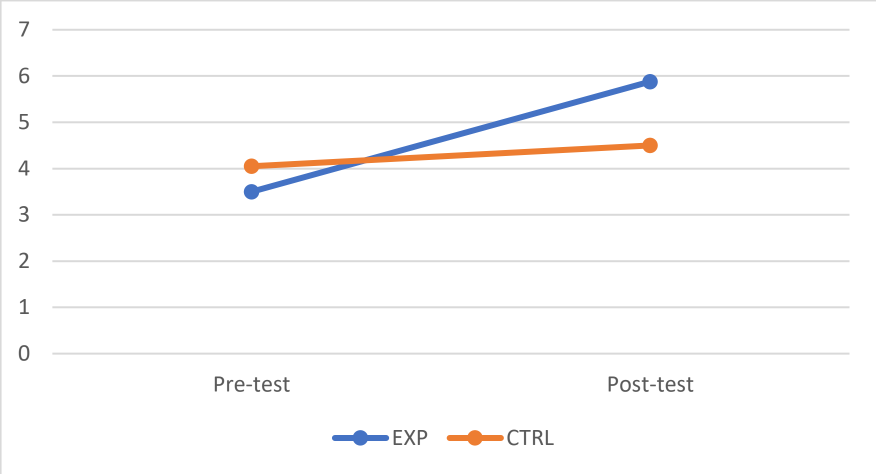
The post-hoc Tukey test found a statistically significant improvement in experimental group (p = .046; d = 0.76; M|D| = 1.04). However, as the data was not normally distributed, the Bonferroni test was also used. This test failed to find a statistically significant difference (p = 0.582). Thus, as will be discussed in Chapter 6, the extent to which the improvement was, in fact, statistically significant is very much a matter of interpretation. As for the control group, the post-hoc Tukey test found no statistically significant improvement (p = .986; d = 0.10; M|D| = 0.11).

Familiar Items (Single-stressed)

The participants' performance on familiar single-stressed items was tested using another two-way mixed ANOVA (Shapiro-Wilk: p = .016; Bartlett's: p = .391), which again showed that there was no interaction effect of Time x Treatment (F(1,88) = 2.915; p = .091) nor a main effect of Treatment (F(1,88) = 0.544; p = .463). However, there was a significant main effect of Time (F(1,88) = 6.327); p = .014). The development in mean scores from pre-test to post-test is illustrated in Figure 5.9.

Figure 5.9.

Production: Line graph depicting the development from pre-test to post-test on the familiar single-stressed items (Study 1).



Further investigation using a post-hoc Tukey test revealed a statistically significant improvement for the experimental group, (p = .016; d = 0.88; M|D| = 1.23). There was no significant change for the control group between pre-test and post-test (p = .943; d = 0.17; M|D| = 0.18). A Bonferroni test confirmed that the result obtained for the experimental group was significant despite the violation of the normality assumption (p = .018).

In order to investigate whether one rule had been easier to learn than the others, a one-way ANOVA was used to look for differences between the improvements on each of the rules. The comparison of the different rules did not yield a significant result ($F(2,8)$ = 3.57; p = .205). However, this is not surprising considering the very small sample size with only three comparable items per category. Bootstrapping was briefly considered as a solution to this, but as this technique has proven rather inaccurate with very small sample sizes (Plonsky, 2013), the idea was abandoned. As Table 5.8 below shows, there seems to be a clear difference between the scores, but due to the p-value obtained, it cannot be ruled out that this is a coincidence.

Table 5.8.
Number of correct scores in each category on pre- and post-test (Study 1).

Category	Pre-test	Post-test	Improvement
Place names	15	43	28
Materials/ingredients	35	51	16
OBJECT/AGENT	40	51	11

Note: Only items that appeared on both tests are included.

At first glance, it may appear as though the place names rule was easier for the participants to learn, as the improvement is bigger, but this might simply be due to the comparatively lower pre-test scores in this category.

Unfamiliar Items (Single-stressed)

Finally, in addition to the calculations above, the participants' performance on unfamiliar items was also measured. Unfamiliar items, here, are items that did not appear in the pre-test. As well as comparing the participants' performance on the new items to their performance in the pre-test, a comparison between new and familiar items in the post-test was also made. This could not be done as a 1:1 comparison on account of the fact that there were only nine unfamiliar items compared to the 18 items used in the pre-test. One possibility in such a scenario could be to calculate z-scores for all the unfamiliar items and then convert the scores to fit an 18-point scale. However, it was decided that rather than converting values, a more appropriate approach would be to compare percentage-wise performance in pre- and post-tests as done in 5.1.3 above. These comparisons revealed that the experimental group improved from pre-test to post-test when only considering familiar items, but further showed that an improvement on the new items, although this improvement was somewhat smaller as shown in Table 5.9.

Table 5.9.

Production: Comparison of familiar and unfamiliar single-stressed items at pre-test and post-test in Study 1.

	Pre-test	Familiar, post-test	New, post-test
Experimental Group	42%	65%	56%
Control Group	48%	50%	49%

In comparison, the control group showed very little variation between test times although it is worth pointing out that the control group did perform slightly better on both the familiar and the unfamiliar items in the post-test. However, as the increase was only 2% and 1% respectively, this difference is negligible. Table 5.9 further shows that the experimental group also performed better on familiar items than on the new items in the post-test (65% vs. 56% respectively). Hence, the experimental group showed an improvement from pre-test to post-test regardless of whether the items were familiar or not.

5.1.6. Double-stressed Items (Production)

As the perception tests yielded some quite interesting results with regard to the participants' ability to perceive double-stressed compound nouns, potential changes in their ability to produce these compounds were also investigated. Thus, a two-way mixed measures ANOVA (Shapiro-Wilk: $p < .001$; Bartlett's: $p = .691$) was run to compare the performance of the two groups at pre-test and post-test. The ANOVA found no interaction ($F(1,90) = .084$; $p = .772$) and no main effect of either of the two independent variables (Treatment: $F(1,90) = .251$ $p = .617$; Time $F(1,90) = 0.937$; $p = .336$). Both groups seemingly performed slightly worse in the post-test than in the pre-test with the mean score of the experimental group decreasing from 6.60 out of

10 in the pre-test to 5.84 in the post-test, and the mean score of the control group decreasing from 6.73 in the pre-test to 6.32 in the post-test. Thus, no real change was observed as the differences between pre- and post-test were very small.

In summary, as table 5.10 shows, the results of the tests only revealed significant findings in some of the cases.

Table 5.10.
Summary of the results of the production tests in Study 1.

	Production							
	Experimental Group				*Control Group*			
	Pre	**Post**	**Impr.**	**d**	**Pre**	**Post**	**Impr.**	**d**
***M*(All items)**	14.2 (3.88)	17.8 (3.00)	**3.6**	0.98	15.7 (4.04)	15.7 (3.37)	0	0
***M*(All SS-items)**	7.33 (5.22)	11.4 (4.14)	**4.07**	0.76	8.73 (5.62)	9.27 (6.11)	0.54	0.10
***M*(Familiar SS-items)**	3.50 (2.45)	5.88 (2.31)	**2.38**	0.88	4.05 (2.73)	4.50 (3.25)	0.45	0.17
***M*(All DS-items)**	6.60 (2.84)	5.85 (2.69)	-0.75	-0.26	6.73 (2.76)	6.32 (3.39)	-0.41	-0.14

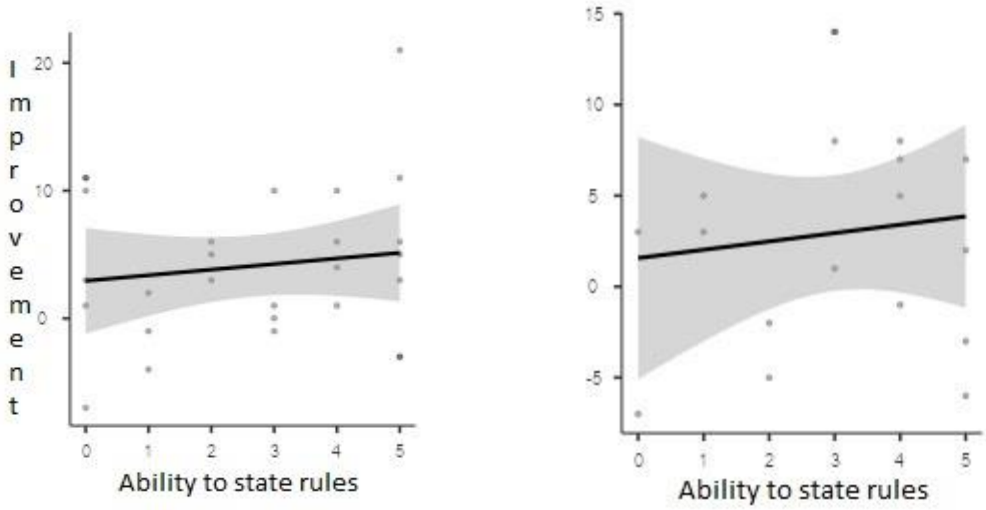
Note: *M* is the mean score. Standard deviations are listed in parentheses below each mean score.

What is important to point out is that the improvements found in the control group were never statistically significant and only showed very small effect sizes. In contrast, the experimental group showed statistically significant improvements in perception across the board. In terms of production, however, the only statistically significant result found was for the familiar single-stressed items.

5.1.7. Correlation between Rule Acquisition and Test Score (Study 1)

As shown in 5.1.1, not all participants managed to learn the rules that were presented to them in the pronunciation course. To investigate the extent to which a participant's ability to state the pronunciation rules affected their improvement, Pearson's r (also known as Pearson's Correlation Coefficient) was calculated for the two variables. The tests showed no correlation between the two variables for neither perception (r = 0.142) nor production (r = 0.124). The plots are shown in Figure 5.10.

Figure 5.10.
Plot of scores on the rule identification task against overall improvement in the perception test (left) and production tests (right).



Plonsky and Oswald (2017) mention that a different method sometimes used in a situation like this is to divide the group according to the median score and run a *t*-test to compare the top half to the bottom half. However, as these authors state, 'when analyses are based on artificial groups like this, any *p*-value, eta-squared, or other statistical result based on an ANOVA can be justifiably questioned or even dismissed out of hand' (p. 385), a correlation calculation seemed like a wiser approach.

Although this finding is not ideal, it is not entirely unique either. As a case in point, Alderson, Clapham, and Steel (1997) investigated the effect of experience and meta-linguistic knowledge on pronunciation accuracy. They found that participants who were aware of and able to describe the rules of French did not perform better in an oral performance test than participants with a poorer grasp of the L2's metalanguage.

5.1.8. Post-test Questionnaires

One of the main concerns upon commencing this research was that students would fail to engage with the materials. When the study was offered to them, it was made clear that it was not intended to take up much of their time, but that half an hour a week would suffice. As the videos are just over a minute long, this was not a bad estimate. Two videos of a minute's duration per week and a fairly simple task for the first video should allow the participants to watch the videos several times and practise the dialogues as they were prompted to do. Although teachers cannot tell how much time their students have spent with the materials, Facebook does allow teachers to see who has watched the uploaded videos. For this reason, much effort was put into keeping track of who watched the videos and when. As a general rule, the students who followed the course as requested, i.e., watched the videos as they were posted, also tended to complete the course.

Before going into any greater detail regarding the questionnaire data, it has to be said that, unfortunately, not all participants who completed the post-tests submitted the questionnaire. In the original design, each subject would have been asked to complete the post-test questionnaire in an adjacent room to where they had taken the post-test immediately after completing the perception and production tasks. However, as the design had to be altered due to the Covid-19 pandemic, students were sent a link to the questionnaire and asked to complete it as soon as possible. Reminder emails were sent to students who failed to complete the questionnaire within two weeks of the test, but despite this effort, not all the post-test questionnaire responses expected were submitted. Thus, of the 36 expected post-test questionnaires, only 29 were received.

One other detail should be mentioned as part of the analysis of this data, namely that the students were asked to supply their participant number in the post-test questionnaire. It is debatable to what extent this may have influenced the participants' answers. In some cases, complete anonymity may lead to more honest answers. However, as the researcher only ever saw the participants for the pre-test and the post-test, there was not a very close relationship between the participants and the researcher. This might have been different in a classroom setting, where the learner might have wanted to please their teacher by providing favourable answers. Finally, it was stressed before both tests that only the researcher would know the identity behind the participant number, so they would gain nothing in terms of good faith from their course tutor if they answered one way or another.

As mentioned in Chapter 4, the questionnaire used both quantitative and qualitative questions. The qualitative element of the questionnaire will be left for the discussion chapter as the data obtained from this section seems better suited for that chapter. The quantitative element of the questionnaire consisted of a series of Likert-style statements which the participants had to rate on a 7-point scale. The quantitative data has been summarised in Table 5.11, which contains the nine questions presented to the participants in addition to the scores and standard deviations for each question. The items selected here were intended to provide a solid overview of both the participants' overall impression of the course (items 1-5) as well as some information regarding specific aspects of the course content (items 6-9).

Table 5.11.

Mean scores and standard deviations from the post-test questionnaire in Study 1 (N = 29).

Question/statement	Mean	SD
1. I found the exercises useful.	6.27	0.83
2. I would have improved more if I had spent more time on the exercises.	5.37	1.30
3. How likely would you be to follow this type of course for an entire semester (3-4 months)?	5.77	1.41
4. I liked using Facebook as a language learning tool.	5.63	1.35
5. I'd rather do this type of exercises on Facebook than in class.	4.87	1.78
6. How would you rate the duration of the videos?	4.00	0.79
7. The talkers in the videos were easy to understand.	6.77	0.43
8. The altered text in the videos helped me focus on the stressed word.	6.43	0.86
9. I found the instructions posted in the group useful.	6.53	0.73

As can be seen from the standard deviations of the data in the table above, there was widespread agreement among the participants with regard to all the items in question. The item that brought about the most disagreement was Item 4, which asked the participants where they preferred to do the tasks, with the deviation for this statement amounting to 1.78. To gauge how the learners felt about this type of exercises, they were asked to explain their rating. This part will be treated in the following chapter.

The statement with the lowest mean was Item 5, which asked the participants to rate the duration of the videos. This is a fortunate outcome as this particular question used a scale ranging from 'too short' to 'too long'. Thus, a mean score of 4 could not be much better – especially considering the low standard deviation (0.79). Furthermore, the participants generally seemed to have found the exercises useful as shown in the responses to Item 1 (M = 6.27; SD = 0.83). This is a good result for the face-validity of the course, as one would obviously prefer learners to have the feeling what they are doing helps them. Naturally, this result should be seen in light of the test results, which also seem to suggest that the exercises were useful – at least to some extent. Also relating directly to the materials was the statement of whether the participants found the altered text useful (item 7). This appears to be the case with this particular item receiving a mean rating of 6.43 (SD = 0.86). As above, this only illuminates the face-validity of the materials but says little about whether the altered text was actually effective in helping the learners.

As described in 4.3.2, the talkers in the videos were both speakers of Northern British English. The decision to use these talkers was entirely vindicated by the unanimous agreement from the participants concerning item 6 in the questionnaire that the speakers were indeed easy to understand (M = 6.77; SD = 0.43). A look through the data set revealed that not a single participant rated this item below 6.

As the final part of this section, two somewhat linked items will be discussed, namely, items relating to whether the participants like using Facebook as a language learning tool (item 9) and to what extent they would be likely to follow a course of this type for a whole semester (item 3). As for the first of these two items, the participants generally responded that they had liked using Facebook (M = 5.63; SD = 1.35). Among these scores, just a single participant gave the overall experience a below neutral rating. This participant gave a score of 2. Upon closer investigation, however, this rating has to be questioned due to the participant's further comments. When asked 'What could have made the course better?' the participant responded 'Honestly I think it was really good, the only thing I would change is that I am not used to facebook and I would have prefer other plataform such as youtube or gmail'. Moreover, when asked to provide any other comments, the participant responded 'Everything was good (: '.

Regardless of whether this rating was a typographical error, the participants generally responded favourably to this item. As for the other two items, the responses were very similar (*M* = 5.77; SD = 1.41). Again, only one participant rated this item significantly lower than the rest (1). Oddly enough, though, this was a different participant to the one mentioned above. This is not to suggest, however, that this rating is not genuine. It is entirely possible that a participant could like the idea of short courses on Facebook, but does not like the idea of longer courses being run on the platform.

5.2 YouTube in English Pronunciation Teaching (Sentence Stress)

As detailed above, this study investigated L1 Spanish speakers' ability to perceive and produce English sentence stress in three different contexts. In addition, the study also looked at how the participants rated various aspects of the course in a post-test questionnaire.

Before the analysis, the possibility of removing two of the sentences was considered. The first was the sentence *It is definitely not my favourite* as it turned out to be a somewhat poorly constructed item. The sentence that was used to elicit this reply was *The Lion King is my favourite Disney film*, but in hindsight, several replies are possible which would all have acceptable foci. For example, little could be said against stressing *definitely* or *not*. However, rather than removing the item, it was decided to widen the number of acceptable replies so that participants who stress either *definitely* or *not* also had their reply accepted.

It was further contemplated whether to remove the sentence *Twenty prisoners escaped* from the data. The sentence was taken from Cruttenden (1997) as an example of an event sentence with the stress on the noun rather than the transitive verb in final position. This decision was taken after baseline data was collected from four native speakers, who all stressed the verb rather than the noun. It is not easy to say why the native speakers' utterances did not match the theory, but one possible explanation is that with that particular sentence the scenario is of a type where the speaker finds it more natural to highlight the action of the prisoners. In comparison, sentences like *My phone is ringing* or *My watch has stopped* are rather mundane. Cruttenden himself talks about this and compares sentences like *The kettle is boiling* which is theorised to be stressed on the noun, to a sentence like *The kettle exploded*, which is theorised to be stressed on the verb. The reason, Cruttenden explains, is that people expect a kettle to boil but probably do not expect it to explode. With this idea in mind, one might ask whether Cruttenden expects prisoners to escape regularly. Clearly, this is not the case for the native speakers used in this study as evidenced by the stress pattern they applied to this particular sentence. However, as this study was primarily concerned with whether L2 English learners would be able to learn a rule and apply it, the sentence was kept. In retrospect, this was the right decision as the results show that this was not the most problematic sentence for the participants to understand as will be shown below.

5.2.1. Acquisition of Sentence Stress Rules

The extent to which the participants had been able to learn the pronunciation rules presented on the pronunciation course was measured by asking them to state the rules in the post-test questionnaire and scoring their answers. As there were no exceptions involved, the participants received one point per correct answer. The scores can be found in Table 5.12 below.

Table 5.12.

Frequency counts for post-test questionnaire scores on item 18 (Please state the rules you learnt on the pronunciation course) in Study 2.

Score	0	1	2	3
Frequency	4	4	4	6

Note: The participants were given example sentences to help them remember the rules.

As in the study above, the participants' ability to correctly state the rules upon completion of the study was rather poor. These results will be dealt with in more detail in Chapter 6. For now, it suffices to say that with such a large part of the participants not being able to formulate the rules, the predictions were that only a marginal improvement (if any) would be found in the tests below.

5.2.2. Results of Perception Tests

Because an initial independent samples t-test (Shapiro-Wilk: p = .482; Levene's: p = .204) found no significant difference between the two groups (p = .155), the remaining data was compared.

As the first focal point, all the familiar items excluding the fillers (items with stress on the last lexical item) were compared using a two-way mixed measures ANOVA (Shapiro-Wilk: p = .128; Levene's: p = .418). The ANOVA found no evidence of interaction between Time and Treatment ($F(1,94) = 0.558$; p = .445). There was, however, a main effect of the Time variable ($F(1,94) = 19.286$; $p < .001$). Post-hoc Tukey tests showed that both the experimental group and the control group made significant improvements (p = .006 and p = .028 respectively). It should be pointed out, though, that the experimental group showed a bigger improvement than the control group as measured by effect size (Cohen's d = 1.06; M|D| = 1.39 and 0.75; M|D| = 0.96, respectively). Figure 5.11 below shows the development in means in both groups from pre-test to post-test.

Figure 5.11.

Perception: The development from pre- to post-test for the experimental group and control group on all familiar items – not including fillers.

The post-test also included nine new items, that is, items that appeared neither in the pre-test nor during the training. These were included as a sort of generalisation test. The improvement on these items was compared to both pre-test scores and post-test scores on the previously heard items. The two groups performed in this test to a fairly equal level. As such, the experimental group showed an improvement of 15% on the familiar items (56% → 71%), whereas the control group improved by 11% (61% → 72%. With regard to the new items, the experimental group improved by 16% (56% → 72%), whereas the control group improved by 17% (61% → 78%). This data can be found in Table 5.13 below.

Table 5.13.

Perception: Generalisation to new items across the three sentences types taught on the YouTube pronunciation course.

	Pre-test items	Familiar, post-test items	New, post-test items
YouTube	56%	71%	72%
Control	61%	72%	78%

Finally, given the unexpected findings with regard to the double-stressed items in the other study, the filler sentences were also analysed. It was initially hypothesised that these would not cause the participants any problems due to the stress patterns of Spanish (see 2.3.2). However, as it turned out, this was not the case. A two-way ANOVA (Shapiro-Wilk: $p = 0.04$; Bartlett's: $p = .461$) found a significant interaction between the two groups ($F(1,94) = 5.39$; $p = .022$), but the post-hoc Tukey tests showed no statistically significant differences between the groups. Somewhat surprisingly, although the results were not statistically significant, the results indicated that whereas the experimental group performed worse in the post-test ($M = 1.90$; SD = 2.00) than in the pre-test ($M = 2.95$; SD = 1.57), the control group seemed to improve from pre-test ($M = 1.90$; SD = 1.14) to post-test ($M = 2.17$; SD = 1.44), but as previously stated, the changes were not statistically significant.

5.2.3. Results of Production Tests

The judges who rated the speech samples for the word stress study also rated the speech samples for this study. Before the data was analysed, an initial test of interrater reliability was conducted using Cohen's Kappa. The interrater reliability tests (see Table 5.14) showed varying degrees of agreement – from moderate to substantial agreement – among the raters in terms of the dialogues (Kappa: 0.56 – 0.73) and substantial in terms of the picture description task (Kappa: 0.61 – 0.71).

Table 5.14.

Interrater values for experimental and control group (Study 2).

	Dialogues			
	Experimental Group		Control Group	
Raters	*R1 - R2*	*R1 - R3*	*R1 - R2*	*R1 - R3*
Agreement	85%	82%	90%	89%
Cohen's Kappa	0.56	0.56	0.73	0.72
p-value	< .001	< .001	< .001	< .001
	Picture Description Task			
Agreement	87%	86%	86%	85%
Cohen's Kappa	0.61	0.61	0.64	0.71
p-value	< .001	< .001	< .001	< .001

As with Study 1, there were cases of disagreement among the raters. These were also analysed acoustically. As mentioned in Study1, acoustic analysis on this type of data may not be 100% accurate due to the manner in which the data was obtained. These issues were particularly noticeable in this study due to the much longer speech samples. However, using a combination of human judges and visual aids was still considered the best approach for rating the files (see Chapter 6 for a discussion).

Intra-rater reliability was also measured using a subset of the data (Table 5.15). The raters assessed another 20 tokens from each of the tests they had already rated. The tokens were analysed together so that each rater produced two values – one for the items produced by the experimental group, and one for the items produced by the control group.

Table 5.15.

Intra-rater scores for experimental and control group (Study 2).

	Experimental Group			Control Group		
Raters	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*
Agreement	95%	95%	90%	93%	93%	95%
Cohen's κ	0.88	0.81	0.55	0.82	0.78	0.77
p-value	< .001	< .001	< .001	< .001	< .001	< .001

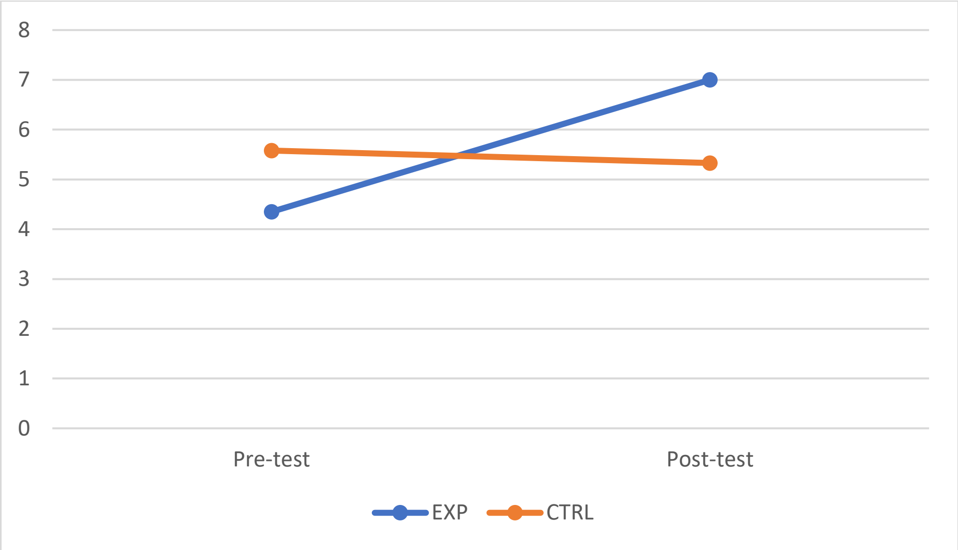
5.2.3.1 Dialogues

The data was first analysed looking at all the pre-test and post-test scores without considering whether the items were familiar or not. As with the perception data, before any data analysis was conducted an independent samples *t*-test was run in order to ensure that the two groups were similar at the time of the pre-test. The prior Shapiro-Wilk test showed that the data was normally distributed (p = .134). As a consequence, the independent samples *t*-test was used to compare the two groups. This test found no statistically significant differences at the time of the pre-test (p = .270). Considering that all participants were recruited from the same phonetics course, this was to be expected.

Similar to the analysis of the perception data, a two-way mixed measures ANOVA (Shapiro-Wilk: p = .022; Bartlett's: p = .199) was used to look for interaction/main effects. The test found a significant interaction between Treatment and Time (F(1,90) = 6.680; p = .011) as well as a significant main effect of Time (F(1,90) = 4.577; p = .035). The post-hoc Tukey test further revealed a significant improvement from pre- to post-test for the experimental group (p = .007, Cohen's d = 0.97; M|D| = 1.22). The result of the Bonferroni test confirmed that the improvement was significant (p = .008) The control group, on the other hand, performed slightly worse on the post-test, but the change was not statistically significant (p = .989; d = -0.09; M|D| = -0.12). These findings are visually represented in Figure 5.12 below.

Figure 5.12.

Production: Development from pre-test to post-test on all the dialogues.



The next step in the analysis was to compare the familiar items on the test only. As above, the Shapiro-Wilk test showed that the data was not normally distributed (p = .020), but Bartlett's test showed that equality of variance could be assumed (p = .052). Hence, a two-way ANOVA was used to analyse the results. The ANOVA found no interaction effect between Time and Treatment (F(1, 90) = 1.594; p = .210), but a main effect was found for the Time variable (F(1, 90) = 8.954; p = .004). The post-hoc Tukey test revealed that the experimental group had made statistically significant improvements (p_{Tukey} = .019; d = 0.88; M|D| = 1.20). Once again, Bonferroni test confirmed the statistical significance of the improvement (p = .022). The control group also improved slightly, but improvement was not statistically significant (p_{Tukey} = .606; d = .36; M|D| = 0.41). This finding is consistent with the line graph of the data in Figure 5.13 below, which clearly illustrates an improvement in both the experimental group and the control group as indicated by the slope of the lines.

Figure 5.13.

Production: Development from pre-test to post-test on the familiar dialogues only.

A result pointing to a benefit after training with familiar items could suggest that simply taking the test improves the participant's performance. While it is possible that familiarity with the test format helped the test-takers, it is important to point out that the experimental group improved more from pre-test to post-test than the control group did. It can thus be suggested that while there could be an effect of taking the test a second time, following the course may have added a further effect (see Chapter 6).

As an additional test for this data set, the participants' performance on the new items was compared to their performance on the familiar items. Since the experimental group had shown improvement on the tests of all items and the familiar items only, it comes as no surprise that this group improved on the new items as well. As such, while the participants had scored an average of 20% in the pre-test, they improved to 42% on the test of the new items. This should be seen in comparison with the remainder of the post-test items on which the participants achieved an average of 35%. In comparison, the control participants averaged 31% in the pre-test, 39% on the familiar items, and 26% on the new items. Hence, whereas the experimental group displayed gains on both familiar and novel items, the control group only seemed to make slight improvements on the familiar items. The data is summarised in Table 5.16 below.

Table 5.16.

Production: Generalisation to new items across the three sentence types taught in Study 2.

	Pre-test items	**Familiar, post-test items**	**New, post-test items**
YouTube	20%	35%	42%
Control	31%	39%	26%

Although not originally planned, it was decided to compare the sentence types to see whether any sentence types had been easier for the participants to learn than others. Unfortunately, due to the low number of different sentences, the one-way ANOVA failed to find any significant difference between the rates of improvement (p = .169). However, it is worth noticing that only the 'contrast' category produced consistent improvements as shown in Table 5.17.

Table 5.17.

Category-wise level of improvement (Impr.) displayed by the experimental and control group on the familiar items.

Category/Item	Experimental Group			Control Group		
Objects of general reference	**Pre-test**	**Post-test**	**Impr.**	**Pre-test**	**Post-test**	**Impr.**
What's wrong with people?	5	7	**2**	10	8	**-2**
Let's go back to my place.	0	2	**2**	0	0	**0**
I quite like the guy.	12	12	**0**	6	11	**5**
Event sentences						
20 prisoners escaped.	1	6	**5**	2	4	**2**
I think there is rain coming.	9	13	**4**	4	5	**1**
My watch has stopped.	4	1	**-3**	1	4	**3**
Contrasts						
Indian elephants have smaller ears.	4	9	**5**	8	9	**1**
It was my mum who won the lottery.	5	11	**6**	10	11	**1**
… I asked for red roses.	2	12	**10**	10	12	**2**

With any intervention study, the aim is to see whether the chosen intervention has an effect. Furthermore, it is of interest to see whether the degree of intervention is important, and how different amounts change the outcome variable. This study was not strictly designed to test the latter, but even so, the correlation between the test scores and the amount of time spent on task (as measured through the students' self-reports) was calculated. Pearson's r was calculated and returned a value of -.071. However, with $p = 0.766$, this result was not significant. As such, it cannot be stated studying more necessarily leads to bigger improvements for this type of intervention.

However, it must be added that Pearson's r is sensitive to sample size, so it is possible that the values presented here are just a reflection of an insufficient number of data points.

5.2.3.2 Free Speech

The other component of the production test was a picture description task in which the participants had to describe various scenarios involving stuffed toys and everyday items (e.g., a bag, a chair, or a bowl). This task tested the students' ability to create the correct contrast between items in the sentence. As expected, this was not an easy task for the majority of participants because, as explained in Chapter 2, Spanish speakers generally use word order rather than stress displacement to signal focus. Furthermore, this task had the added cognitive demand of requiring the participants to create their own sentences, though the amount of possible variation in sentence structure was limited.

Comparing the experimental and the control group after the pre-test, the initial Shapiro-Wilk test showed that the data was not normally distributed ($p = .008$), and hence a Mann-Whitney U-test was used to compare the two groups. The test found no significant differences between the two groups ($U = 151$; $p = .115$).

The following two-way mixed measures ANOVA (Shapiro-Wilk: $p = .003$; Bartlett's: $p = .766$) found no interaction effects ($F(1,78) = 1.789$; $p = .184$) and no main effects of neither Time ($F(1,78) = 0.832$; $p = .365$) nor Treatment ($F(1,78) = 0.238$; $p = .627$). Figure 5.14 below shows the development in the two groups from pre-test to post-test on the familiar items. As can be seen, although the ANOVA failed to find either a statistically significant interaction effect or a main effect, there does appear to be a difference between the two groups.

Figure 5.14.

Production: Development from pre-test to post-test in the Picture Description Task (familiar items only).

Whereas the control group displayed a slight decrease from pre-test to post-test, the experimental group improved. A paired samples *t*-test showed that the improvement displayed by the experimental group was significant (p = .023; d = 0.57).

A test of generalisation was also carried out to see if the course had been effective in teaching participants to use stress in new scenarios/contexts. However, these results were far from encouraging. The experimental group averaged 29% in the pre-test items, but only 11% on the unfamiliar items in the post-test. Similarly, the control participants averaged 39% in the pre-test items, but only 3% on the unfamiliar items.

Despite these results, there is reason to believe that the training did have an effect, owing to the fact that the experimental group did perform better in the post-test than in the pre-test despite the improvement not being statistically significant. The experimental group improved by an average of 15% (29% → 44%) when not considering the new items. This should be seen in contrast to the control group whose average dropped by 2% (39% → 37%). Thus, it is entirely possible that the lack of a significant finding is down to this particular part of the study being underpowered due to the relatively small number of items on the test.

In summary, whereas the study on word stress showed no significant improvement for the control participants, in this study, there were several instances where the control participants did improve significantly. The main results of the study are summarised in Table 5.18 below.

Table 5.18

Summary of the production data from Study 2.

	Production							
	Experimental Group				*Control Group*			
	Pre	**Post**	**Impr.**	**d**	**Pre**	**Post**	**Impr.**	**d**
***M*(All dialogues)**	4.35 (2.04)	7.00 (3.12)	**2.65**	0.97	5.58 (2.55)	5.33 (3.03)	-0.25	-0.09
***M*(familiar dialogues)**	1.83 (1.15)	3.17 (1.83)	**1.34**	0.88	2.13 (1.45)	2.71 (2.10)	0.54	0.36
Generalisation	20%	35%	15%	N/A	31%	39%	8%	N/A
***M*(picture description)**	1.95 (1.73)	2.95 (1.99)	1	0.49	2.76 (2.07)	2.57 (2.20)	--0.19	-0.09
Generalisation	29%	11%	-18%	N/A	39%	3%	-36%	N/A

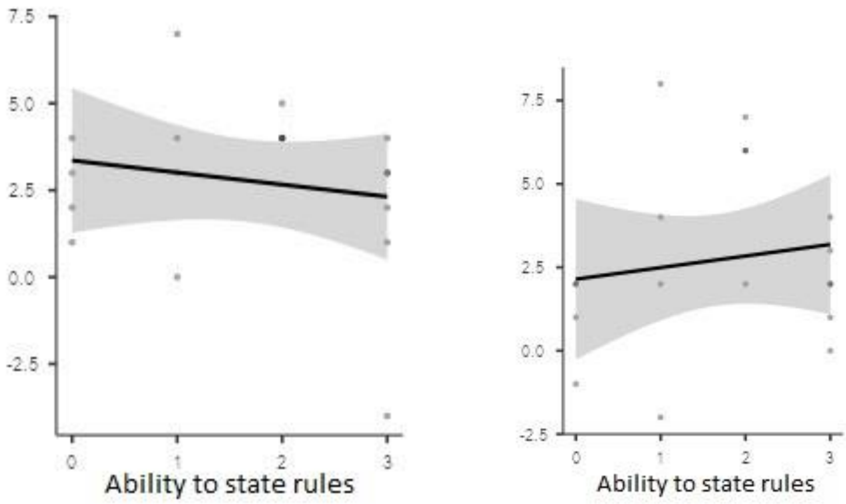
At first glance, there does not appear to be much difference between the two groups apart from two exceptions. First, the experimental group showed some improvement in the picture description task, whereas the control group did not. Second, the control group's improvement only approached significance in the dialogue production task while the experimental group showed a significant improvement. Nevertheless, it is worth keeping in mind that in all cases where improvements took place, the experimental group showed bigger improvements than the control group.

5.2.4. Correlation between Rule Acquisition and Test Score (Study 2)

As in Study 1, there was a clear difference in Study 2 between the experimental participants in terms of how many of them managed to learn the targeted pronunciation rules. To find out if there was a correlation between the students' ability to learn the pronunciation rules and their improvement from pre-test to post-test, Pearson's r was calculated.

Figure 5.15.

Plot of scores on the rule identification task against overall improvement on the perception (left) and production tests (right) in Study 2.



As can be gathered from Figure 5.15, no correlation was found between the two variables. Pearson's r showed a weak, non-significant negative correlation ($r = -.176$) for perception and a weak, non-significant positive correlation (0.153) for production.

5.2.5. Post-test Questionnaires

Similar to what was the case in the Facebook study, not all the participants who completed the post-test returned the post-test questionnaire. More specifically, four participants failed to return the questionnaire leaving just 19 submissions. The participants who had signed up for the post-test but failed to attend were also invited to fill out the questionnaire, but none of them did so despite reminder emails being sent two weeks after their scheduled test date. The responses given by the participants are summarised in Table 5.19 below.

Table 5.19.

Likert-type questions in the post-test questionnaire for Study 2 (N = 18).

Question/statement	Mean	SD
1. I found the exercises useful.	6.20	0.69
2. I would have improved more if I had spent more time on the exercises.	5.35	1.69
3. How likely would you be to follow this type of course for an entire semester (3-4 months)?	5.45	1.57
4. I'd rather do this type of exercises on YouTube than in class.	5.10	1.71
5. How would you rate the duration of the videos?	3.95	1.03
6. The talkers in the videos were easy to understand.	6.45	0.89
7. The altered text in the videos helped me focus on the stress in the sentence.	6.35	1.18
8. I found the instructions in the video description useful.	6.50	0.69
9. I liked using YouTube as a language learning tool.	6.20	0.95

The first thing that becomes apparent when looking at the data is that they are generally very similar to the responses from Study 1 (Facebook). The responses to item 5 showed that the participants generally agreed that the duration of the videos was appropriate (*M* = 3.95; SD = 1.03) although it must be said that the variation in this group was slightly higher than in the Facebook group. A total of four of the 19 participants rated the videos too short and three rated them too long. It was rather unexpected that some of the participants would think the videos should have been shorter, as the videos were just under two minutes long. Another similar set of ratings is the one relating to preferences regarding online or in-class practice (item 4). Here, their mean score was 5.10 (slightly higher than the Facebook group's 4.87), and the standard deviation was 1.71, thus indicating that there was some disagreement among the participants. That being said, a closer look at the data reveals that only four of the participants gave a less than neutral rating and three neither agreed nor disagreed.

A further confirmation that using speakers of Northern British English did not pose a problem came from the vast majority of respondents stating that the talkers were easy to understand (item 6) with just one neither agreeing nor disagreeing (M = 6.45; SD = 0.89).

As for the course itself (item 1), all the respondents rated the exercises useful (M = 6.20; SD = 0.69), and further agreed that the instructions in the video descriptions were useful too (M = 6.50; SD = 0.69). The latter statement was particularly encouraging because it had been a cause for concern whether the participants would actually read the information in the description. Of course, as mentioned in the previous study, the high rating that these items received do not directly translate to the course being effective, but they do encourage future projects of the same style. Finally, it was very encouraging to see that the participants enjoyed the course (M = 6.20, SD = 0.95) with the lowest score coming from just one participant who neither agreed nor disagreed. However, despite this very good overall rating, the participants showed less enthusiasm when asked whether they would be likely to follow a course of this type for a whole semester (M = 5.45; SD = 1.57). An in-depth look at what exactly the participants said will be provided in Chapter 6.

Chapter 6 – General Discussion

The two studies presented in this thesis looked at the applicability of the social media platforms Facebook (Study 1) and YouTube (Study 2) in teaching either word stress rules (Study 1) or sentence stress rules (Study 2) to L1 speakers of Spanish. In both studies, an experimental group receiving four weeks of training was compared to a control group. The training consisted of short videos containing dialogues of two to three turns. With each video, the participants were asked to identify either the stressed element of a compound noun or the word carrying the nucleus in a sentence. The feedback to the students consisted of the same dialogues, but with an orthographically enhanced audio script intended to direct the participants' attention to the correct stress placement.

Upon completion of the two pronunciation courses, their effectiveness was assessed. First, the degree to which the experimental participants had been able to learn the required rules was assessed. Next, the participants' ability to identify the correct stress pattern in target words or sentences as well as their ability to produce target stress patterns correctly was investigated.

Through a step-wise treatment of the research questions presented in Chapter 1, this chapter provides an interpretation and discussion of the results provided in Chapter 5. The results of the two studies will be discussed separately, starting with Study 1. While each study will be discussed independently, the data from the post-test questionnaires (RQs 5 and 11) will be treated as one at the end of the chapter.

6.1 Word Stress Taught Using Facebook (Study 1)

6.1.1. RQ1: Rule Learning

The first research question looked at whether the L1 Spanish experimental participants would be able to learn the three stress rules (and their exceptions) that were taught on the four-week pronunciation course. Considering the relative simplicity of the rules, it was expected that the participants would be able to state the three rules they worked with once the course had finished.

As mentioned previously, this was tested by asking the participants to state the rules as part of the post-test questionnaire. The intention had been to ask the participants to fill out the questionnaire in an adjacent room to where the tests were taking place. This would have allowed the researcher to ensure that the participants did not use their notes to answer the question as the two rooms were separated by a glass wall. However, as the location of the post-test had to be changed, there was no way for the researcher to guarantee that the participants did not use their notes. However, judging by the language used by many of the participants, this was not a major issue. As shown in Chapter 5, the number of participants who provided a satisfactory answer was rather low with only seven out of 29 able to fully state all three rules and their exceptions.

The analysis of the relationship between the ability to declare the pronunciation rules and improvement in the post-test were not encouraging. The results showed that the participants who scored in the top half of this test (i.e., between three and five on the scale from 0-5) were no more likely to improve from pre-test to post-test than the participants who scored in the bottom half. This assertion is based on the lack of correlation between the individual participant's ability to state the rule and the overall improvement score with Pearson's r of 0.142 for perception and 0.124 for production (see 5.1.7). Thus, in Study 1, it seems as though being able to state the word stress rules did not generally help the participants, although a great deal of variation was seen in the results.

It is important to keep in mind that this way of testing whether the students were able to learn the pronunciation rules is somewhat implicit in the sense that the participants were never told that they would be asked this question. If they had been informed about this, it is likely that they would have tried harder to memorise the few lines of text given to them on the course. However, as the students' affinity for rote learning was not a concern for the studies in this thesis, this method seemed more appropriate. To this should be added that the fact that a participant was unable to state the rule in their second language does not necessarily show that no uptake has taken place. It is possible that the participants knew how to stress certain types of compound nouns, but were unable to state exactly why. As shown in Table 6.1 below, some of the rules were rather difficult to interpret.

Table 6.1.

Examples of participants descriptions of the pronunciation rules taught in the online pronunciation course in Study 1 (Facebook).

- Singles stress when it goes with stress and other names places Long words in the middle And compounds that have one element that refers to the action or object second stress also with some compounds that talks about food like mud pie (502028)
- In place names: stress is in the second word compound words: stress is on the second word (202027)
- Place names, ingredients and materials... (502022)

Note: The participants were given an example of each type of compound noun to help them remember the rule.

It is not unheard of that participants find it difficult to state meta-linguistic knowledge in a second language as evidenced by Couper (2011). Although an ideal scenario would be one in which all the participants had learnt the pronunciation rules, what is equally important in a language learning context is that the participants improved. This will be the focus of the sections below.

6.1.2. RQ2: Rule Application to Aural Input

The second research question looked at whether L1 Spanish learners of English would be able to apply what they had learnt on the course to aural input. Initially, the students had been expected to be able to apply the rules to aural input by identifying the type of word they heard during the test and select the correct option on the screen ('Single stress' or 'Double stress'). However, given the overall poor performance in terms of stating the pronunciation rules, the participants were no longer expected to fare particularly well in any of the following tests.

As described in Chapter 5, this was tested by presenting the participants with a series of compound nouns through the TP software. There were two different hypotheses for this question, each pertaining to a different stress pattern. Due to the stress pattern of Spanish compound nouns (see 2.2.4), it was expected that the participants would find single-stressed compound nouns difficult, but struggle less with double-stressed ones. These predictions turned out not to be correct as the mean score for the experimental group at pre-test was 11.7 for the single-stressed compound nouns, but only 9.29 for the double-stressed compounds. In terms of improvement, the participants improved more on the double-stressed items (30%) than they did on the single-stressed items (15%). However, the relatively large difference in percentages can in part be explained by the lower pre-test score for the double-stressed items.

What is important to keep in mind is that the treatment proved effective. This can be seen by comparing the improvements between the experimental group and the control group. Whereas the experimental group showed an overall mean score improvement of 4.71, the control group actually performed slightly worse in the post-test than in the pre-test with an overall improvement of -0.72.

With regard to whether the results were statistically significant, the initial test of single-stressed items showed a statistically significant interaction between group and time. Unfortunately, upon further inspection, the post-hoc Tukey test did not yield a statistically significant result (see 5.1.2). However, it should be pointed out that the post-hoc Tukey test could be slightly too stringent. For example, a Bonferroni test for four pair-wise comparisons would yield a p-value of .041. The four planned tests would be:

- Pre-test – post-test (Experimental)
- Pre-test – post-test (Control)
- Pre-test – pre-test (Experimental vs. Control)
- Post-test – Post-test (Experimental vs. Control)

More tests could be constructed such as Pre-test(Exp.) – Post-test(CTRL.), but as this test would add very little information to the overall picture, it should not be included. This is, in fact, not an entirely novel idea, as discussion about how to approach family-wise error control is currently being debated among researchers (Lakens, 2020). As a consequence, although the expression 'approaching significance' is disapproved of by some researchers, this seems to be a case where it is rather apt.

6.1.3. RQ3: Rule Application in Speech Production

The third research question asked whether the participants would be able to apply single stress correctly in production following the course. This question can be divided into two due to the fact that the tests included two different stress patterns, namely single stress and double stress. As with the previous research question, the original hypothesis was that the participants would be able to apply the rule and successfully complete the test. However, as the participants had struggled to state the rules in the post-test questionnaire, a case could be made that the participants would find it hard to produce the words correctly. On the other hand, as many studies have shown that an improvement in perception can lead to an improvement in production – at least for segmental features (Akahane-Yamada et al., 1996; Flege, 1995), it was difficult to predict how the participants would fare on this part of the test.

Overall, the results showed that whereas the experimental participants improved, the control participants did not. However, the improvement that the experimental participants made was not statistically significant in all cases. Looking first at the participants' ability to produce single stress, it was observed that while the control group did not improve, the experimental group did make statistically significant improvements.

It has to be said that when looking at all the single-stressed items, the improvements shown by the experimental group only approached significance ($p_{Tukey} = .069$). Although this value seems larger than the commonly used alpha of 0.5, it must once again be taken into consideration that the uncorrected p-value was .0149. Thus, whether or not this would be significant with Bonferroni correction would be a matter of how many planned pair-wise comparisons were made. Admittedly, this value would not hold in the scenario of four pair-wise comparisons like the one mentioned in 6.1.2. but would do so with three pair-wise comparisons. When looking at the familiar single-stressed items, however, the improvements were statistically significant ($p_{Tukey} = .016$). These are important findings because, taken together with the findings from the perception tests, they suggest that the training the participants received was effective in helping them perceive and produce single-stressed compound nouns more accurately. They further indicate that the training procedure would be useful for English learners of a similar demographic.

Despite these encouraging results, it has to be said that the improvements seem to have come at a cost as the participants' performance on double-stressed items decreased slightly. Similarly, the control group performed marginally worse although neither change was statistically significant.

It is somewhat surprising that the experimental participants improved on just one category of items considering that the rules they have been working with explained how to treat both single- and double-stressed compound nouns. However, as pointed out in 6.1.1, only a minority of the participants were able to state the rule so perhaps it is not so surprising after all. The reason that the production of double-stressed compounds was considered at all was that the experimental group improved their perception of these compounds. In this respect, it should not be entirely unexpected that no improvement in production was observed, as previous studies on the perception-production link suggest that improvements in perception are often seen before improvements in production (Bradlow et al., 1997). It should be added that this literature deals with segmental phonology, whereas the findings in this study would support a link at the suprasegmental level as well. Additionally, as García Lecumberri (2006) found, some Spanish learners of English do well with perception of English compounds, so perhaps the actual conundrum is why the pre-test scores were so relatively low in the first place. The caveat here is that the learners in García Lecumberri's study were all 'at least upper-intermediate level' (p. 189) which means they were probably more proficient users of English than the participants used in this study.

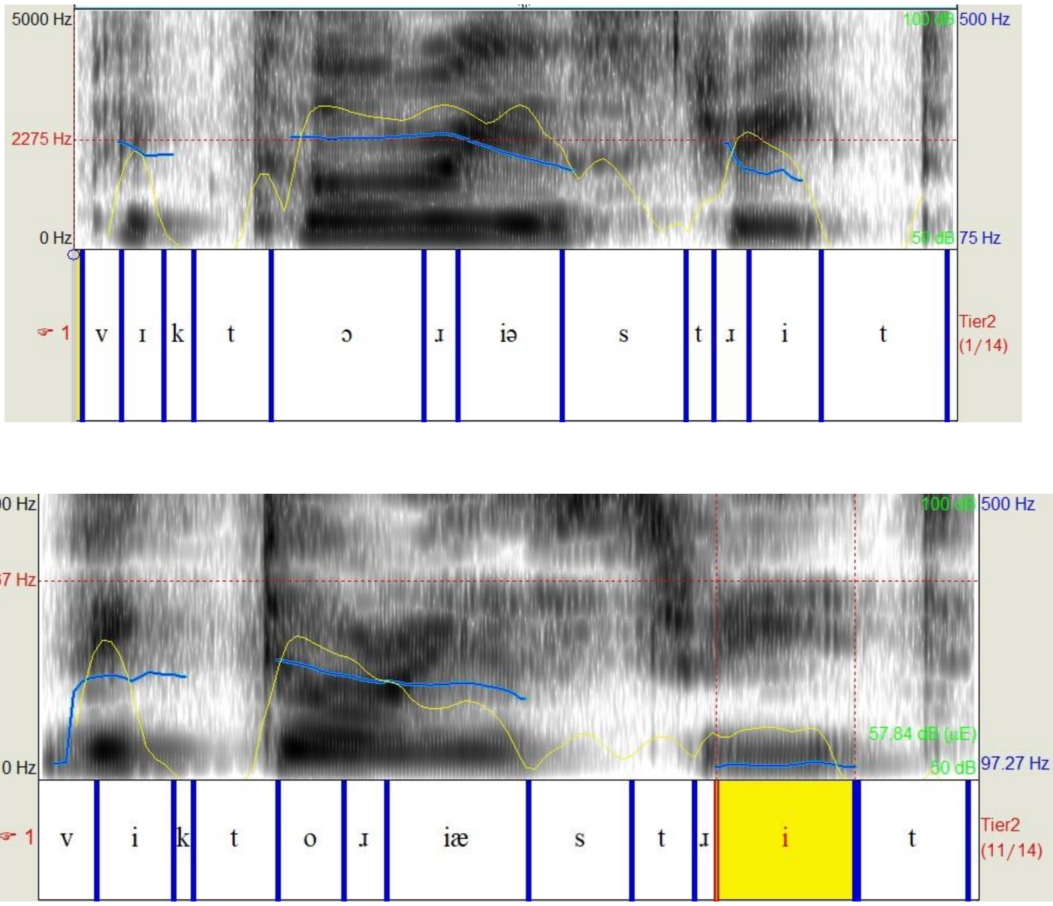
Production ratings

At this stage, a few comments should be made about the rating procedure. As shown in 5.1.3. the inter-rater reliability coefficients were very low. In some cases, this could present a major problem due to the fact that it could indicate that the raters are not sure how to rate the items or that the items are simply not suitable for the chosen rating procedure. However, because each item was rated at least three times (twice by the researcher and once by a native speaker judge), and because all disagreements were analysed acoustically, the method is nonetheless rigorous enough for the data to be analysed further.

Given the large amount of data that had to be rated, it is not surprising that some items are mis-labelled simply due to lapses in concentration. This is regardless of rater fatigue which, as described earlier (4.6.4), was kept to a minimum by asking the judges to take frequent breaks. Furthermore, there were some cases where it was almost impossible to tell whether the stress was placed on one element or the other, so cases like these are likely to lead to disagreement between the raters. Additionally, although great care was taken to ensure that the raters knew exactly what to listen out for, the comparison of the ratings did seem to exhibit a tendency on the native listeners' part to be influenced by features of intensity and pitch. This can be illustrated with Figure 6.1 below.

Figure 6.1.

Victoria Street: A native speaker's production (top) and one participant's production (bottom).



Note: The longer vowel in 'street' is highlighted in yellow.

As can be seen from the spectrogram, both intensity and pitch are higher on the stressed syllable of *Victoria Street*, which is exactly what you would expect in this type of word. However, the vowel in *Street* was occasionally markedly longer than is usually heard in the production of native speakers, and this may have prompted the raters to rate this sort of item as 'Not acceptable'. As mentioned in Chapter 1, duration can function as a cue for stress, but whether the raters heard the second element as stressed due to the longer vowel or simply heard it as distinctly different from what they would have expected is not obvious.

Finally, the ratings used for this analysis were the judges' original ratings. In hindsight, it would have made more sense to use the ratings after the raters had had a chance to discuss the cases of disagreement. This way, instances where the disagreement was due to a lapse in concentration on one rater's part would not have influenced the interrater agreement. Undoubtedly, this would have given a more accurate picture of the real differences in how the judges heard the speech samples.

6.1.4. RQ4: Generalisation to New Items

The participants' ability to perceive and produce novel compound nouns was measured through a percentage-wise comparison between the participants' scores on the novel and familiar items in the post-test and the corresponding items in the pre-test. As shown in 5.1.3 and 5.1.5, results varied depending on what specific category was looked at. However, one pattern that seems to emerge is that the experimental group improved on all items in the post-test. This improvement was only 9% (67% → 76%) on the perception test of single-stressed items, but 14% (42% → 56%) on the production test of single-stressed items. In comparison, the control group showed no signs of improvement on these items. Thus, what the participants learnt on the course transferred to some degree novel words (the picture was slightly different for familiar words as will be discussed below).

As mentioned above, it is unlikely that the improvements came about as a consequence of the participants learning and applying the rules that were taught on the course. Instead, what seems to have happened is that the pronunciation course raised the participants' awareness of the stress differences in English and allowed them to better perceive these differences. With this in mind, the fact that the training also affected novel words matches the findings of previous research on training of stress perception and production (e.g., Brawerman-Albani et al., 2017). Additionally, these results support theories and findings from cognitive linguistics as presented in Couper (2012), and add to the volume of research suggesting that awareness raising is an important part of SLA (Coniam, 2002; Schmidt, 1990; Sharwood Smith, 1981).

Possible Familiarity Effect

At this stage, a comment should be added regarding the issue of whether taking the pre-test helped the participants in the post-test regardless of what group they were in. As shown in 5.1.3, both groups seemed to perform better on the familiar single-stressed items than they did on the unfamiliar single-stressed items. This could indicate that simply taking the pre-test help the participants improve on the post-test. However, when looking at the figures for the double-stressed items (5.1.4), this claim is difficult to uphold, since the control group's performance on the familiar items only improved by 1% between the two test times. By contrast, control participants improved by 13% on the unfamiliar items. What is more, the pattern is the complete opposite for the experimental group. This group performed better on the unfamiliar double-stressed items than they did on the familiar ones, but performed better on the familiar single-stress items than unfamiliar ones. These figures are summarised in Table 6.2 below.

Table 6.2.

Perception: Comparison between the two groups' performance on familiar and unfamiliar single-stressed and double-stressed items.

	Pre-test	**Familiar, post-test**	**New, post-test**
Single stress			
Exp. Group	67%	83%	76%
Ctrl. Group	64%	72%	62%
Double stress			
Exp. Group	47%	69%	82%
Ctrl. Group	52%	53%	65%

Thus, there seems to be no particular pattern in how the groups performed on the two tests.

6.2. Sentence Stress Taught Using YouTube (Study 2)

6.2.1. RQ6: Rule Learning

As in the previous study, the first research question in Study 2 looked at whether the L1 Spanish participants would be able to learn three pronunciation rules following a pronunciation course on YouTube. In this study, the rules were all related to sentence stress – more specifically to objects of general reference, event sentences, and contrasts. These rules were considered slightly more complicated than the word stress rules because applying the rules required an understanding of the grammar of the sentence. What may have made this group's task easier is the fact that the participants did not have to worry about exceptions to the rules, as there were none. All in all, given that the participants had an entire week to learn each rule, it was nonetheless expected that the majority of the participants would be successful.

Because some participants' productions had to be left out of the analysis due to audio quality issues, only 18 participants' answers were taken into consideration for this part of the post-test questionnaire. After rating the responses, this expectation was only partly met. As described in 5.2.1. the responses were rated on a scale from 0-3 based on how accurately the participants were able to state the pronunciation rules. Eight of the respondents scored in the bottom half, and 10 of the respondents scored in the top half with four achieving the minimum and six achieving the maximum score. To see if there was any relationship between the ability to state the rules and test performance, the scores from the rule application test were compared to the test scores using Pearson's r. This test found no correlation between the ability to state the rule and improvement between pre-and post-test (r =

0.153, p = 0.544). These numbers are emphasised by the fact that the participant who improved the most was only able to state one of the three pronunciation rules. Similar to Study 1, some of the rule descriptions were difficult to interpret due to the language used as shown in Table 6.3 below.

Table 6.3.

Examples of participants descriptions of the pronunciation rules taught in the online pronunciation course in Study 2 (YouTube).

- 1) Prominence in a sentence, the word before the last one is usually the stressed one 2) Event sentence, SUBJECT+INTRANSITIVE VERB 3) Contrastive stress between two opposite words (212155).
- 1-New information 2-default position 3-contrastive purpose (212149).

Note: The participants were given an example of each sentence type to help them remember the rule.

Although the results were not as good as one could have expected, they nonetheless more than matched the achievements of the participants in Study 1. Overall, the participants in Study 2 answered much more accurately than their Study 1 counterparts. This was seen in the vocabulary used in the responses with words like ‘event sentence’ and ‘contrast’ featuring heavily. This was quite surprising considering that the rules taught on the sentences stress course were by no means easier than the ones taught on the word stress course. If anything, they might be considered more difficult because some of the vocabulary is more abstract. For example, terms like ‘event sentence’ and ‘object of general reference’ are used much less frequently than terms like ‘place name’ or ‘materials/ingredients’. The one difference that could speak in favour of the sentences stress rules being easier to learn is that, unlike the word stress rules, they did not contain exceptions.

6.2.2. RQ7: Rule Application to Aural Input

The seventh research question looked at whether the L1 Spanish learners of English would be able to apply the sentence stress rules for non-final sentence stress to aural input. Although the participants were not particularly good at stating the pronunciation rules, the results from Study 1 nonetheless suggested that improvement could still be expected.

The results from the perception tests in this study were the most intriguing of all the results in the two studies. As predicted, the experimental group made statistically significant improvements. However, the control group also made statistically significant improvements despite not participating in the YouTube pronunciation course.

In order to explain this curious finding, it might be worth looking at the participants' academic experience. The participants for Study 2 were all recruited during their first year. As explained in Chapter 4, this group of students took a mandatory course in English pronunciation which included a focus on sentences stress. The control participants came from the same group but did their first test eight months later than the experimental group. Thus, they may have had time to consolidate what they learnt on their course to some extent. Furthermore, they had the advantage of the additional teaching and practice they received in other courses such as *Lengua Inglesa III* and *Fonética y Fonología Inglesas* – both part of their second year. It is possible that this additional practice helped them improve between tests without receiving the instructions provided to the experimental participants. This claim is supported by the fact that the control group performed better in the pre-test than the experimental group did; although it is important to point out that the difference between the two groups was not found to be statistically significant.

What must be pointed out, though, is that the experimental group did improve more than the control group yielding a large and medium effect size respectively ($d_{EXP.}$ = 1.06; $d_{CTRL.}$ = 0.75). As the mean for the control group at post-test was 12.9 out of 18, the reason for the smaller improvement cannot be attributed to a ceiling effect. This means that even if all the participants benefited from what they learnt at university, the extra practice that the experimental group received helped them improve even further.

6.2.3. RQ8: Rule Application in Speech Production

This research question was divided into two as there were two different production tests, namely a control read-out-loud task and a less controlled picture description task. Looking first at the results of all the items from the controlled production task (i.e., the dialogues), improvements were once again found in both the experimental group and the control group. However, the improvements made by the control group were minimal, and only the experimental group achieved a statistically significant improvement. As with the perception results, the improvements made by the experimental participants showed a medium to large effect size (d = 0.97).

Similar results were obtained when analysing the familiar items in isolation. Once again, the experimental group produced a statistically significant improvement. The control group, on the other hand, did not, despite showing a very small improvement (0.583 of a possible 5.875). Thus, there is reason to believe that the training helped the participants produce sentence stress with greater accuracy in a controlled environment after the course.

The second part of the research question asked whether the participants would be able to apply contrastive stress correctly in a spontaneous production task as described in Chapter 4. This task was expected to be difficult for the participants because they had to put contrastive focus on an element which was not sentence-final, a linguistic resource not frequently used in Spanish (García Lecumberri, 1995). The analysis of the data found that the experimental made a statistically significant improvement from pre-test to post-test (p = .023; d = 0.57). However, this is an unmodified score. If this value were to undergo the same family-wise error correction as previous post-hoc tests, the result would not reach significance. Thus, again, it very much seems a matter of interpretation whether this result is, in fact, statistically significant or not. This issue notwithstanding, given the results from the read-out-loud tasks which suggested that the participants found contrasts relatively easy compared to other sentence types, it seems plausible that the result from the picture description task is not one that was achieved by chance, thus indicating that the training was effective.

Production ratings

As was the case in Study 1, the judges' task of rating the items were made more difficult by features not directly related to pitch and intensity. One such feature was a tendency from some of the participants to produce the dialogues in a rather staccato manner. This may have given the judges the impression that every word was stressed, which consequently, makes it difficult to tell if there is a difference between the degree of stress. Additionally, some of the participants divided tone units into two although they were supposed to have been produced as one (e.g. |*The TRAIN*| |*is COming*|. This would naturally lead to an incorrect production of the phrase. The reason for both these issues should probably be found in the participants' level of English. The read-out-loud task was made easier for the participants by giving them time to read the dialogues to themselves and ask questions regarding vocabulary, etc. before reading them out loud. However, it would seem that some of the participants may have needed additional time to work with dialogues in order to improve their oral productions.

6.2.4. RQ9: Generalisation to New Items

The ninth research question looked at whether the participants' ability to perceive and produce the three sentence types generalised to new items. This was done in a similar manner to Study 1, namely by doing a percentage-wise comparison of the pre-test items, familiar post-test items, and unfamiliar post-test items. The results from the read-out-loud exercises and the picture description tasks will be treated separately.

Dialogues

As shown in 5.2.2, in terms of perception, the experimental group's improvement on the familiar items was similar to the improvement on the new items (71% and 72% correct respectively compared to 56% in the pre-test). This would indicate that the training had been effective. However, it should be noted that the control group made similar improvements to those made by the experimental group, namely 61% correct in the pre-test, 72% correct on the familiar post-test items and 78% correct on the new post-test items.

Because the two groups made similar gains in perception, it could be expected that they would also make similar gains in production. However, as shown in 5.2.3., the experimental group performed much better in the production post-test on both familiar and new items. As such, the experimental group made gains of 15% on the familiar items (20% → 35%) and 22% on the new items (20% → 42%). In comparison, the control group only produced gains of 6% on the familiar items (24% → 30%) and

4% on the new items (24% → 28%).

As mentioned earlier, the control group may have been able to take advantage of what they were learning in their university courses while participating in the study. However, even if that is the case, there was clearly a significant difference between the two groups in terms of speech production. This again suggests that the pronunciation course had a positive effect on the participants.

With the regard to the items on the production test, it has to be acknowledged that a couple of the sentences may have been more difficult for the participants because they were questions. For example, the sentence *Could you bring me two black bowls and two white bowls?* seemed to cause many of the participants some trouble as they tended to finish the sentence with rising intonation, thus increasing the likelihood that the final word be perceived as stressed. This can possibly be explained by the formal English teaching the participants have received. It is not uncommon to tell students that you have to use rising intonation at the end of Yes/No-questions in English. Thus, even if the participants had understood that they had to emphasise the colour of the bowls in the aforementioned sentence, they might still fail to produce the sentence correctly. This is because they would be required to produce two adjacent stressed words with rising intonation but different degrees of stress. This is not at all impossible, but it might be too much to ask of participants at this level.

A similar issue was observed for the sentence *What's wrong with people?*. Orthographically, this sentence looks like a question, but in this particular context (i.e., as a response to hearing about a shooting), it is a statement. Therefore, it cannot have a rising pitch at the end, but should instead be treated as a statement which would be shown by a fall in pitch. However, many of the participants failed to grasp this.

As for the phrase *Let's go back to my place*, disregarding the many participants who transferred the Spanish nucleus placement, there was a strong tendency among the respondents to stress *back* rather than *my*. This is interesting for two reasons. First of all, it suggests that the participants realised that the expected L2 tonicity was different than the one in their L1. Second, it is possible that *back* is chosen as the focal point because the participants want to create a link to a past event in the scenario (i.e., the speakers were at speaker B's place, and B wants to emphasise that this is where they should return to). This is of course speculative, but in future studies of similar designs, it would be interesting to do follow-up interviews with the participants to further investigate their motivation for stressing that particular word.

Finally, an interesting observation regarding the items in the 'contrast' category was that many of these were stressed incorrectly, but in a rather unpredictable pattern. As described in Chapter 2, the tonic syllable is usually sentence-final in Spanish, and one might expect this pattern to be carried over to the L2. This was indeed the case in the majority of instances. However, this pattern was far from all-encompassing as many of the items in which the context required a contrast, the wrong contrast was made. As a result, old information was put into focus in a non-final position. As a case in point, item 12 from the pre-test was in some instances produced, as in:

**I said to meet at TEN to eight not TEN past eight*

or in some cases, a verb was put into focus without it being in any obvious contrast to previous elements, as in:

**These are blue roses. I ASKed for red roses*

Picture description task

Regarding the picture description tasks, as can be seen from the results, a clear difference was observed in subject performance between test items focusing on nouns and test items focusing on prepositions. As shown in 5.2.3.2, the experimental group only managed a total of six correct answers out of 57 items (11%). Although this was a better performance than the control group, which scored two of 62 (3%), it was still a markedly worse result than expected considering the results of the preposition focus domain items. It is not immediately clear why this is the case. One possibility is that because all the items on the pre-test followed the same pattern (i.e., contrastive stress on a preposition), the participants may have thought that this pattern could be applied across the board on the post-test as well. However, as the participants were not pressed for time, they should have been able to take their time and consider what the contrasts were between the images. Also, if the participants had indeed spotted the pattern, one would have expected a better performance on the items where stress on the preposition was required. On a positive note, the fact that many of the participants incorrectly stressed the prepositions does indicate that they understood that a contrast had to be created.

6.3. RQ5 and RQ10: Students' Views of Using Social Media for Language Learning

As explained in the introduction to this chapter, this section deals with the questionnaire responses from both studies. The questionnaire aimed to gauge the applicability of the materials used in the pronunciation course, and more importantly, the two social media platforms in general. This was primarily done through Likert-type items, which the participants were asked to rate on a 7-point scale. To add a qualitative element to the questionnaire and to elicit more detailed information from the participants, a few open questions were added as well.

Despite the findings laid out above, that the participants were only somewhat able to state the rules that were taught on the course, the questionnaire revealed that the participants found the pronunciation course useful. The average scores were 6.27 (SD = 0.83) for the Facebook group and 6.20 (SD = 0.69) for the YouTube group. Considering the findings of previous research that has looked at students' perceptions of the use of technology in language learning (Eren, 2012; Kabilan, Ahmad, and Abidin, 2010; Kho and Chuah, 2012; Promnitz-Hayashi, 2011), this was very much expected.

With regard to course duration, the participants who submitted the post-test questionnaire generally agreed that the course duration was appropriate. Just three participants across the two studies stated that it was too short. Evidently, this is important as you do not want to risk students getting bored.

Following their experience with a four-week pronunciation course, it was hoped that the participants would be able to contemplate engaging with a longer course. Thus, the participants were asked about the likelihood of them participating in a similar course for an entire semester. The two groups responded with very similar figures, with the Facebook group giving an average rating of 5.77 (SD = 1.41) and the YouTube group giving an average rating of 5.45 (SD = 1.57). Although these scores tell us nothing about whether the participants actually would follow a course for a whole semester, they at least indicate that they were not opposed to the idea. As the standard deviation shows, there was quite some variation in this part of the data, which probably reflects students' varying attitudes towards this type of learning. Previous research investigating learning through social media has found that student attrition and decreased participation is a problem on longer courses

(Buga et al., 2014; Mompeán and Fouz-González, 2016). It would be interesting to know exactly was causes this so that future courses could be tailored in such a way as to keep learners engaged. Even though it may be prudent to keep online courses relatively short, at the same time the course has to be of a duration sufficient to help the students improve. As seen earlier in this chapter, the participants did improve for the most part, but it is very likely that a longer course would have benefitted some of the participants.

Apart from the duration being right, the content has to appeal to the students as well. As mentioned, the course consisted of short videos accompanied by some exercises in the shape of a few brief questions. With regard to the videos, the students seemed to prefer the format with short videos as only three out of 48 respondents gave scores in either the high or low end of the scale (i.e., 2 and 7). It has to be questioned whether the students giving the 7 had actually understood the question. This is because it is hard to see how videos of approximately a minute's duration can be judged to be 'much too long'. As for the participants saying the videos were too short, this is perfectly possible since some students like to be presented with example sentences in abundance. However, all in all, considering the mean score and the low variance, this particular aspect must be deemed highly successful.

To see whether the so-called digital natives (Prensky, 2001) preferred to do the exercises in class or online, they were asked to state to which extent they agreed with the statement 'I'd rather do this type of exercises on Facebook than in class' and to explain why. As it was shown in the previous chapter, the results varied greatly, with some participants much preferring a classroom-based approach, while others seemed delighted with the online format. The reason for this becomes evident when looking at the student comments, which show how students vary in their preferences. The students who would have preferred the classroom-based format primarily focused on the fact that in the classroom they can get immediate feedback from their tutor, whereas online they feel somewhat left to themselves. This can be gathered from the comments in Table 6.4 below.

Table 6.4.

Participants' post-test comments on learning modality (1).

- It would be easier to be corrected at the moment by a teacher (502031).
- In class I tend to be more concentrated and there are physical people who can solve my doubts, whereas in Facebook, despite being also assisted, it is more complicated to get the concepts (202020).
- I prefer in-class pronunciation exercises because you can get a direct feedback from your teacher (202018).
- The teacher can provide me some advices in order to identify properly the stress of some words and why, unlike Facebook, where the student has norms of the stress with barely explanation (502020).
- I prefer in-class pronunciation exercises rather than through YouTube because it would be easier to ask for doubts (212159).
- I prefer class because I like to have a direct conversation with my teacher in person (212143).

Note: Participant numbers (in brackets) starting with 21 are from Study 2. Other participant numbers belong to Study 1.

What is interesting is that in contrast to these statements, which generally focus on the presence of a teacher as the sole argument, the students at the other end of the scale mentioned a variety of reasons for preferring to work online, as shown in Table 6.5.

Table 6.5.

Participants' post-test comments on learning modality (2).

- Facebook does more dynamic, you can do it when you want and it helps you to organize yourself (202030).
- Being at home alone makes me spend more time and reflect on what I am doing and in class I feel shy (502030).
- It's more accessible (502044).
- Because I think that a student can focus more in their house rather than in class, which it can be crowded and noisy (212137).
- I have more time for practice by myself. Plus, I don't get nervous as I tend to do in front of the whole class and I could ask any doubts I have next lesson (212138).
- On youtube I have my own time and I don't depend on anyone. I know my skills and I can use the time I want in order to correct myself or to pay attention to the details (212146).

Participant numbers (in brackets) starting in 21 are from Study 2. Other participant numbers belong to Study 1.

These comments quite neatly cover the advantages described in Chapter 1 and also echo the advantages previously suggested in research on the use of technology in language learning and teaching (Neri, Cucchiarini, and Strik, 2002; Pennington, 1999). However, this should not be used as a reason to disregard the comments made by the participants who preferred learning in the classroom. It is important that new ways are continually developed to make IT ever more inclusive, so that, eventually, every learner feels comfortable learning with technology.

It must be added that the students' evaluation of working with the software may be artificially high due to the novelty effect (Stockwell and Hubbard, 2013; Stockwell, 2007). An attempt to account for this effect was made by adding the question 'Would you like to use this type of instruction throughout the semester?' to the post-test questionnaire. As mentioned in sections 5.1.7 and 5.2.5, students generally stated that they would be likely to want to follow a similar course for a whole semester. This was based on the scores on a 7-point scale of 5.77 (SD = 1.41) in Study 1 and 5.45 (SD = 1.47) in Study 2. This question of course assumes that the participants can detach themselves from their present experience – which could be influenced by a novelty effect – and give a relatively objective answer to the question. To investigate this in more detail, longer studies looking at both effectiveness and student attrition are needed.

One of the advantages of online learning is that it is possible to access content from anywhere as long as you have a smartphone and access to data. However, with this easy access also comes the risk of learners engaging with content in less-than-ideal conditions. For example, working with listening exercises will not be ideal on a noisy bus. Such a scenario was a potential issue for this study, but much to the relief of the researcher, 100% of the respondents stated that they worked with the materials at home. Interestingly, the vast majority of the student chose to work on their phones rather than on a PC or laptop.

The final, and somewhat less encouraging result that should be mentioned is that to the question of whether the students would rather have used a different platform to take the course, only seven of the 29 respondents answered 'No'. Thus, 22 of the participants would rather have used a variety of other platforms, with YouTube being mentioned most frequently (18 times). There can be several explanations for this – and in hindsight, it would have been a good idea to ask the students to elaborate on this point – but the most likely reason is that the students simply do not see Facebook as a platform for learning. As mentioned, in the pre-test questionnaires, none of the respondents indicated that they used Facebook for language learning, although eight stated that they did use Facebook to learn about things that interest them. One would be excused for assuming that the students who already use Facebook as a learning tool were the ones who preferred to use Facebook over other platforms, but as can be seen from Table 6.6 below, this is not the case.

Table. 6.6.

Participants using Facebook as a learning tool before the study and participants who stated that they would prefer Facebook to other platforms.

Prefer to use FB as a learning tool	202005	202014	202015	202018	202030	502024	502031	502044
Prefer to use YT as a learning tool	202005	202011	202021	202031	505022	505041	505043	

Note: Only one participant (202005) of the former group preferred to use Facebook after the study.

It must be added that some of the participants did not have a Facebook account when they started. For this group, it would be surprising if they converted to staunch Facebook believers after a four-week course. Moreover, of the participants who already had an account, only three reported using Facebook on a daily basis. With this in mind, the overall positive feedback the course received could be seen as even more encouraging. However, at the same time, there are warning signs that Facebook may be going out of fashion with younger generations of learners (see 7.6); or at least that their use of Facebook is changing, both in terms of how frequently they use it, and in terms of what they use it for.

6.4. Additional Remarks

Although the two studies showed varying degrees of effect depending on item types and test modality, the effects were generally significant, and some showed medium to large effect sizes with Cohen’s d being as 1 or above in some cases. These results are very encouraging considering the fact that the effectiveness of instruction has been suggested to be larger in laboratory settings than in classroom studies (Li, 2010). It is important to note that Li (2010) did not actually analyse the effects of pronunciation instruction. However, the arguments provided for the greater effects, such as the greater control afforded by the laboratory setting, should also apply to pronunciation. Both classroom studies and laboratory studies have an advantage over the studies presented here in that a teacher or researcher can ensure that the participants spend the intended time on the tasks.

However, it seems that even without this supervision, participants generally engaged enough with the course content to improve their pronunciation skills. Furthermore, research has also suggested that results of pronunciation instruction are larger in an ESL context than in an EFL context (Derwing, 2003). This is most likely because participants in an ESL context have the advantage of being surrounded by native-speaker input and from being able to engage with native speakers. The relatively large effects found in the studies presented here could be explained with reference to Derwing and Munro (2005), who argue that bigger effects are often found with low-level learners. That could easily be the case here as many of the participants showed significant room for improvement – at least on the production tests. On the perception tests, on the other hand, many of the participants performed remarkably well as described above.

As a final point, a few words should be said about the factor of student motivation. When the recruitment material was uploaded to the university's VLE, the researcher received emails from students saying that 'our teachers has told us to take part in your study', which is not exactly an indication of intrinsic motivation. However, it is interesting to contrast this with the very positive feedback that the participants provided. This could indicate that students like learning languages through social media, even if they do not see these platforms as language learning tools (but further research would have to confirm that). Furthermore, it is possible that the fact that the participants were not the researcher's own students had a negative effect on the results in the sense that some students may have put in less effort because there was no real consequence for not doing the exercises. Judging by the completion rate for the weekly questionnaires, this was not a big problem at first, but in the third and fourth week, participation did seem to drop somewhat. If used as part of a classroom-based English course, the teacher would be able to make sure the participants complete all the exercises, thereby ensuring that the learners maximise the benefit of using the course content.

When discussing the effectiveness of any kind of intervention, it is important to look at whether the participants actively take part in the treatment. Unfortunately, a large number of students failed to complete the post-test. In a study that runs over several weeks, it is not uncommon to see some participant attrition (MacKay and Gass, 2005), however, the attrition was most likely exacerbated by the Covid-19 pandemic. This claim is supported by the fact that a large portion of the student attrition was observed in relation to the university declaring that face-to-face teaching had been cancelled (16th March, 2020). As the two studies ran a week apart, Week 4 in Study 1 happened at the same time as Week 3 in Study 2. It can be gathered from Table 6.7 that this week (beginning on the 16th March, 2020) saw a significant increase in student attrition compared to previous weeks.

Table 6.7.

Weekly participation in the two studies.

	Week 1	Week 2	Week 3	Week 4*
Study 1 (Facebook)	50	42	42	31
Study 2 (YouTube)	36	31	16	21

Note: Participation was measured using the weekly questionnaires in the first three weeks. In the fourth week, participation was measured through submission of audio recordings.

Chapter 7 – Conclusion

7.1. Social Media in L2 Pronunciation: Summary of the Current Studies

The overarching problem that this thesis attempted to address is the well-known fact – at least among teachers and pronunciation researchers – that pronunciation is difficult to teach and that teachers find they have insufficient time, training, and materials to teach it (Henderson et al., 2012). There is ample evidence in the literature that language learners can improve their pronunciation with or without the use of technology. In the past, a suggestion for the former has been audio-visual training tools (de Bot and Mailfert, 1982). One issue with these tools is that it takes time for students to learn how to use them, which is not helpful to a teacher who is already pressed for time. More recently, commercial language learning software has been launched, which claims to help learners with their pronunciation and do so without the involvement of a teacher. However, many of these software solutions have been found to care more about sales than pedagogy, and are generally of little use in terms of pronunciation learning (Neri et al., 2002).

Despite the issues pertaining to commercial language learning software, a possible solution to the current issues that many teachers face is technology. This would have to be technology that students can use autonomously outside the classroom without prior time-consuming instruction in how to use the materials. It goes without saying that this technology must be shown to be effective, so it is not used just for the sake of it. Although Golonka et al. (2014) found limited effectiveness of the 350 CALL studies they surveyed, they did highlight CAPT as an area where technology seems to have an effect.

For these reasons, this study investigated the applicability of two popular social media platforms: Facebook and YouTube. Previous research on the use of Twitter for pronunciation teaching has shown promising results (Fouz-González, 2017; Mompeán and Fouz-González, 2016), so there was reason to believe that other social media platforms can be used too. Furthermore, studies on both Facebook and YouTube in language teaching have already been carried out, but these either did not focus on pronunciation (AbuSa'aleek, 2015; Arzu, 2014; Ghani, 2015; Saylag, 2013) or were limited to classroom use (Alwehaibi, 2015; Eren, 2012; Hismanoglu, 2012). Studies looking at social media and pronunciation teaching have been conducted, but here the focus has been on student perceptions rather than the actual effectiveness of the technology (Balbay and Kilis, 2017). Consequently, there was a gap in the literature which was filled partially by the studies presented in this thesis, that is, studies that look at the effectiveness of these social media platforms specifically for pronunciation learning outside the classroom.

The first study looked at the applicability of Facebook in teaching three stress rules relating to compound nouns in English, whereas the second study looked at the applicability of YouTube in teaching three specific stress rules relating to English sentence stress. The two studies employed the same overall approach in that the teaching materials were short videos containing dialogues between two native speakers with added audio scripts. The participants were first asked to deduce the pronunciation rule based on the input and were later given the rule as part of the weekly feedback. The feedback also included a revised audio script for the dialogues, which used orthographically enhanced text to direct the participants' attention to the relevant stress pattern.

In the pre-test, the participants were first tested in their ability to correctly identify the stressed word in aural input. Next, they were asked to produce a number of items in a read-out-loud test. In Study 2, the participants also completed a picture description task in order to elicit more spontaneous input. This process was repeated in the post-test, but in addition to the two tests just mentioned, the participants were also asked to state the rules they had learnt on the course. This was done as part of a post-test questionnaire. The main findings from these tests will be described in the section below (7.2). Overall, the two studies provided evidence that Facebook and YouTube can be effective in teaching word and sentence stress to L1 Spanish learners of English and have paved the way for future studies into the use of social media in English pronunciation teaching (see 7.4).

7.2. Summary of Results and Findings

Test Data

Looking first at word stress, this is an aspect of English pronunciation which has been taught successfully in the classroom both without technology (e.g., Couper, 2012) and with technology (e.g., Hismanoglu, 2012). What these approaches have in common with the current studies is that they all, in one way or another, draw the learner's attention to what stress is. Hence, it was hoped that the approach used in this study would have a positive outcome.

As mentioned above, a range of factors contributed to the expected pronunciation learning in Study 1; among these was awareness raising, which was achieved through textually enhanced input and explicit instruction. Although the study was originally designed to focus on single-stressed items only, the data nonetheless showed that the pronunciation course had improved the participants' ability to correctly identify double-stressed compounds as well. As for speech production, the study once again showed that the training had been effective. The analysis of all the test items as a whole found significant improvements. However, these improvements were only found when looking at single-stressed items. One unexpected finding was that the data for the production of double-stressed items showed a non-significant reduction in post-test scores. This was particularly surprising given the good results obtained in the perception test.

When interpreting these results, it is important to see them in relation to the results of the control group. In comparison, the control group only made improvements in two categories, namely perception of novel single-stressed items and production of familiar single-stressed items, but neither of these improvements was statistically significant – or even approached significance. Hence, Study 1 provided empirical evidence that teaching word stress in English compound nouns can be done effectively through Facebook.

The focus of Study 2 was sentence stress. Like word stress, this is an aspect of English pronunciation which has been taught with good results in the classroom and through technology. Specifically, AbuSeileek (2007) used the *Mouton Interactive Introduction to Phonetics and Phonology* software as part of a pronunciation course for L1 Arabic speakers. Although this software proved effective, it might be too specific in scope to appeal to the average language learner.

Most studies seem to have focused on contrasts, which might be because it is a rather salient feature in the language (Levis and Muller Levis, 2018). Indeed, the results from Study 2 suggested that contrasts were easier for the participants to learn, although this could not be proved statistically.

In a similar vein to Study 1, Study 2 found that the training provided on the YouTube course had been effective. Although only half of the participants were able to state the rules they had been working with, improvements were observed in both speech perception and speech production.

Looking first at speech perception, both the experimental group and the control group showed significant improvements between the two test times, although the experimental group showed the biggest improvements of the two groups. With regard to speech production, the course proved highly effective in the controlled environment (the read-out-loud task). When comparing the participants' performance on familiar and new items, the analysis showed that they performed slightly better on new items than on familiar ones. Hence, the training generalised to unfamiliar items in a controlled environment. The experimental group also improved in spontaneous speech (picture description task), but these were not statistically significant. In comparison, the control group made no statistically significant improvements in speech production. These results, therefore, suggest that the course was effective, but that longer intervention is probably needed for the effects of the course to be observed in spontaneous speech production.

Questionnaire Data

In addition to offering empirical evidence that the learning method is effective, the two studies added to the existing literature on student perceptions of working with technology in language learning. Echoing the findings of several previous studies on learning with technology (Eren, 2012; Fleck et al., 2014; Kabilan, Ahmad, and Abidin, 2010; Yunus and Salehi, 2010), the comments made regarding the course were, to a large degree, positive. One finding that stands out is that YouTube seems by far the more popular platform of the two with regard to language learning activities such as the ones used in the two studies conducted for this thesis. Whereas the majority of respondents in Study 1 (Facebook) declared that they would have preferred to use YouTube, not a single participant in Study 2 (YouTube) said they would rather have used Facebook.

These positive remarks should be seen in the light of the difficulties that the researcher encountered when initially trying to recruit participants for the study (see

4.1). The disconnect between the initial recruitment issues and the positive feedback could indicate that students are not, initially, particularly drawn to the idea of learning a language through social media. However, once they get started, they seem to enjoy the experience. What is more, the participants found the courses useful and could generally see themselves following a course on social media for at least a whole semester - although future research will have to judge whether students really are able to follow a course of such a long duration.

In summary, when looking at the two studies together, there is empirical evidence that both Facebook and YouTube can be used effectively in second language pronunciation teaching. Although the courses did not lead to the expected rule learning, the courses, nevertheless, helped the participants improve their pronunciation in only four weeks. The improvement could be a result of working with the videos containing orthographically enhanced audio scripts (Sharwood Smith, 1981). Alternatively, the improvement could be explained from a cognitive perspective. When the participants were made aware of the differences between English and Spanish stress patterns and were allowed to practise with the materials on the online course, new phonological concepts may have started to form in the minds of the learners (Fraser, 2006b). Whatever the explanation, it is very encouraging indeed to see a change in speech patterns which often cause problems for L1 learners of Spanish (Gutierrez-Díez, 2012; Ortiz-Lira, 1998) in such a relatively short space of time.

7.3. Limitations to the Studies

The studies presented here both have limitations which should be considered when assessing the applicability of Facebook and YouTube in pronunciation teaching.

One rather serious limitation to both studies is the lack of a delayed post-test.

While there can be little doubt that the participants performed better immediately after the course than they did in the pre-test, the studies give no indication of whether the benefits of the courses were still present a month or more after the course finished. When planning the recruitment and data collection it seemed unlikely that participants would be available for a delayed post-test as they would have progressed to a new academic term. Considering the participant attrition that was observed when the Covid-19 pandemic hit, it seems even more unlikely that a delayed post-test would have been feasible.

Another thing that should be mentioned is that the participants' ability to state the pronunciation rules was not tested directly in the pre-test. Hence, it is impossible to say whether the rules stated in the post-test questionnaire reflect learning, or they are simply a reflection of the participants' prior knowledge. However, considering the pre-test results and the participants' prior formal English education as stated in the pre-test questionnaire, it seems fair to assume that the participants did not already know the rules they worked with on the pronunciation course. Furthermore, the elicitation of the pronunciation rules did not proceed as intended because, since the test conditions had to be altered due to the Covid-19 pandemic. Therefore, it cannot be guaranteed that the answers provided by the participants regarding the pronunciation rules were genuine.

It should also be noted that the intervention only lasted four weeks. The length of the intervention was, in part, decided by the availability of the participants in that an effort was made to avoid the test time clashing with the participants' exam dates. It is possible that a longer course would have led to larger effects (Lee, Jang, and Plonsky, 2015), but given that each week introduced a new topic and that the topics were only somewhat related, this should not have been too much of a limitation to the current studies. However, as stated above, a longer course duration could have led to more practice, and consequently, statistically significant improvements even in spontaneous speech.

Word familiarity was not tested prior to the studies, which could be a potential issue. However, as described in 4.6.1, in Study 1, words were chosen that learners at CEFR B1 level should have some degree of familiarity with. At the same time, however, too much familiarity would mean that students would produce the words correctly, which could have led to a ceiling effect. Considering the range of scores in terms of correct productions in the pre-test (2-20; 25 was the maximum), a word familiarity test would not have been out of place.

Another issue pertaining to Study 1 is that production was not tested in spontaneous speech. This was partly due to the assumption that a four-week training course would be insufficient for improvements to be measured in this type of speech production. However, the results from the picture description task in Study 2 to some extent refutes this assumption, since the experimental group did show a significant improvement in this task. Future studies should include spontaneous speech production regardless of intervention length.

A final limitation worth mentioning is the use of remote data collection. As described in 4.5.1, the production data was primarily obtained through having the participants record themselves using their mobile phones. Some test times yielded remotely recorded speech samples of generally high quality. Unfortunately, the quality at other test times was significantly poorer. The big difference between the conditions of the test times was that during one test time (March/April, 2020), the Murcia region was in a complete lockdown, which meant that ambient noise was limited. During the other (November/December, 2020), however, there was only a partial lockdown in place. This meant that there was more traffic on the roads and more noise from people talking in the streets. Admittedly, it cannot be said with certainty, that the poorer quality of some speech samples was not due to the participants failing to find a suitable place to make their recordings. Regardless of whether the varying quality of the speech samples was related to the lockdown or to the participants themselves, it must be concluded that remote data collection is far from ideal. It can, however, be used in some cases, but large amounts of data will need to be collected to allow for the discarding of several speech samples.

7.4. Future Research

The first point that should be reiterated here is that based on the questionnaire responses obtained from the participants, YouTube proved a more popular platform than Facebook. However, this should not discourage future studies on Facebook as a language learning tool. Rather it should be seen as an encouragement to design tasks that are even better suited for the tools that Facebook has to offer. Two of the features that should make Facebook stand out as superior to YouTube in terms of providing teaching are the 'Group' and the 'Social Learning Units'. These make it possible to give students the feeling that they are part of a learning community and affords the option to ask questions as well as discuss problems and solutions to tasks with other members of the group. This was an element which the participants in these studies rarely – if ever – took advantage of, despite being encouraged to do so. Also, the point should be made that 'less popular' is not the same as 'unpopular'. The feedback on Facebook as a language learning tool was generally very good, so there is certainly reason to believe that this platform can be an interesting topic for future research.

There seem to be vast opportunities for future research to investigate other aspects of social media and pronunciation teaching. More specifically, the extent to which more online social activity affects both the effectiveness and the students' views of using social media for pronunciation learning should be explored. This should ideally be done by someone who sees their students regularly, so that the tutor can help create group cohesion and ensure that the online group is a safe space. Moreover, as this study only focused on certain aspects of word and sentence stress, future research should also explore other aspects of suprasegmental phonology such as speech rhythm and intonation. This would most likely require the inclusion of some audio-visual feedback, but given the ease with which hyperlinks can be used to direct learners to new websites, this should be a simple task. Alternatively, both concepts could be covered in instructional videos uploaded to either platform.

As described in 4.2, the initial idea for the studies for this thesis was to use participants from local academies. It is a well-known issue that much research in applied linguistics as a whole has relied too heavily on the participation of university students. As a consequence, the results obtained do not necessarily apply to the general population. Future studies should look into how other demographic groups respond to the type of learning used in the present studies.

For example, in terms of language proficiency, it might be interesting to see how this type of training affects both less and more proficient learners. The students in this study were all studying English at university level, and many had at least some knowledge of English phonetics and phonology. Similar to García Lecumberri (2000), this study found that the participants struggled with certain types of English sentence stress as evidenced by the overall poor performance in the category of event sentences. Given the educational background of both sets of participants, it can easily be imagined that lower-level students would find this area of L2 acquisition rather difficult. Thus, future projects could look at the effects of computer-assisted training for English sentence stress for both lower-level and higher-level learners.

As mentioned above, the age range of the participants in this thesis was very limited. As a consequence, a look at how different age groups respond to different platforms could also be very important. Some age groups may already see Facebook as antiquated. This is particularly plausible considering the participants' overwhelming preference for YouTube. These studies could yield undesired results, though. If it is found that certain age groups are actually discouraged from engaging with learning materials due to the platform through which these are delivered, teachers in classes with a large age range might not be able to use social media after all. This is a particular worry, as the increased focus on Lifelong Learning and Continued Professional Development means that it is no longer uncommon for learners beyond their formative years to start learning a language. This should not discourage undertaking such research, however, as such knowledge would be valuable in itself.

Because the learning method used in the studies for this thesis was rather rigid, it would be interesting to see whether students would benefit from using the same technology with a larger degree of freedom. For example, future studies could look at how YouTube could be used for creating dialogues or other spoken output recorded by students, thereby adding an element of collaboration, as this seems to further increase motivation and learner autonomy (Hafner and Miller, 2011). This would also be in line with the suggestion of O'Bryan and Hegelheimer (2007) to link online and classroom-based learning.

One thing that has set this study apart from most other studies on social media in ELT is the very limited involvement of the tutor. In this study, the tutor's role was limited to creating and uploading the materials, and giving weekly feedback to the group (except for the last week of each course when each participant received individual feedback on their recordings). It is possible that the participants would have felt a stronger obligation to practise, if they had regularly met face-to-face with their tutor. Thus, future studies should look at the importance of having a tutor as a part of the learning experience. That is not to say that the teacher/tutor should take a large amount of time out of their schedule to engage with the students online. Instead, the focus should simply be on whether students perform better after doing a pronunciation course on social media if the course is introduced to them by their teacher.

Although the responses from the post-test questionnaires were very positive indeed, a small group of respondents did mention that they would prefer to practise pronunciation in the classroom. This would allow them to receive immediate feedback from their tutor. This leads the thoughts directly to the potential incorporation of ASR technology into pronunciation courses like the ones used in this thesis. As shown by Imoto et al. (2002) and Lee et al. (2017), software is being developed that can detect problematic stress patterns, and using such software would give the students the immediate feedback they desire. Of course, this technology would not be directly part of the social media platform, but as both Facebook and YouTube allow links to external sites, it should be fairly simple to lead learners to this technology if/when it is readily accessible. One might hope that this would make the students spend more time with the materials, which could lead to better results. Considering the findings from de Bot (1983), which showed that the right learning modality can lead the user to engage significantly more with the activities, this is not a naïve hope in the least.

As mentioned in Chapter 6, a few of the participants mentioned that they would have liked to use a different platform. Other types of social media have already been investigated for their applicability in English pronunciation instruction such as Twitter (Mompeán and Fouz-González, 2016) and Telegram (Xodabande, 2017), but as the number of applications increases, so should the research into how we can make the most of these applications in education. Instagram and TikTok, with their huge popularity – especially among certain age groups – would be two obvious choices. However, it would be important to ensure the developed materials are suitable for these platforms. Materials such as the ones used in this study would be unlikely to work well because they require the use of both video and written instructions. As always when using technology, the target should always be to help students learn and not simply keeping them entertained.

From a linguistic point of view, although the studies found no statistically significant difference in the students' ability to learn the various pronunciation rules, future studies should look closer at whether some rules are easier to learn than others. On the surface, the results from this thesis (Study 1 in particular) indicate that a difference might exist. In addition, Ortiz-Lira (1995) and Pennington and Ellis (2000) seemed to suggest that stress rules are not created equal when it comes to SLA. It should be pointed out that these studies looked at sentence stress, but this gives even more reason to investigate the acquisition of word stress rules.

Finally, as the results obtained in Study 1 showed, although only a handful of the experimental participants managed to learn the stress rules, they improved both their perception and production of English compound nouns nonetheless. With this in mind, future research should explore the perception-production link in word stress application for L2 learners in more detail. A large, and growing, volume of literature has described this link for segmental phonology (see Aliaga-García, 2017 for a review), but only a small number of studies have looked at both perception and production of word stress in SLA, and it is currently impossible to say whether a similar link exists.

7.5. Recommendations

Previous studies on social media platforms in pronunciation teaching have highlighted the fact that this type of instruction could prove particularly popular or effective because users can incorporate language learning into their e-routine (Blattner and Fiori, 2009). While it had been hoped from the outset of this study to obtain similar findings, this proved impossible for Study 1, as only a minority of the participants used Facebook regularly (every week). However, given the positive feedback and the positive test results, it seems reasonable to conclude that both Facebook and YouTube have a place as future language learning tools. In the case of the current study, pronunciation was the focus, but other parts of L2 English instruction should be taught using social media as well.

In the studies presented here, short videos were uploaded to the platforms. However, in the case of Facebook, any type of material could be used. YouTube is more limited in this respect, but there are still vast opportunities yet unexplored. Therefore, one recommendation is that teachers start collaborating to create materials for various aspects of pronunciation teaching – and indeed teaching as a whole. There is already a host of different websites that offer teaching tips and lesson plans for a large variety of topics. However, what is needed is something more centralised. In a research context, we have seen the development of open-source databases such as IRIS (an online portal for research materials in SLA). Similarly, there is a large community for researchers using Praat in which scripts are shared freely. A similar type of open-source database for instructional videos would empower teachers to use any social media platform in their teaching, assuming that the platform is compatible with the given file format.

Employing social media to provide pronunciation instruction would be something very new to most language teachers, and, as Darcy et al. (2012) point out, teachers may, at times, be reluctant to make radical changes to the way they teach. This is supported by Pierce (2014) who found teachers were very reluctant to implement HVPT training as part of their lessons. She speculates that the ideology of the communicative classroom still has a strong hold in many language teachers, thus leaving them unwilling to consider activities that are not communicative in nature. While this could, at first, seem like an obstacle to the application of social media to pronunciation teaching, it is important to keep in mind that – depending on the exercises chosen – social media could be used entirely outside the classroom. This means that teachers would only have to make minimal changes to their lessons to provide their students with potentially effective learning materials.

With that said, this study found an interesting divide between the participants in terms of preferences with regard to instruction format. Despite all the students being very similar in terms of age, they differed significantly in terms of whether they preferred online instruction or not. Although many indicated that they enjoyed the freedom online pronunciation instructions offers, some preferred to have the teacher present to help them through the exercises. Similar findings were made by Stenson et al. (1992), but it is still somewhat surprising that this divide persists considering how comfortable learners of the participants' age are with technology. This point should not be overlooked, as it would be a mistake to put these learners at a disadvantage by moving all pronunciation instruction online. Thus, despite the issues teachers face when it comes to teaching pronunciation in class, the data from this study suggests that there is no other alternative, if we want to give all students a fair chance. This is, at least, until sufficiently accurate ASR is developed to provide these learners with the immediate feedback they need.

Finally, when using any kind of social media platform for pronunciation teaching, it is equally important that the reasons for using these are founded in research, and that the targeted issues have real implications for pronunciation teaching. Using TikTok to deliver any old pronunciation activity just because the platform appeals to a specific group of learners is not a sufficient reason to use it. However, this is not to say that TikTok or similar platforms cannot be used in pronunciation teaching; at the moment there is little research evidence to suggest one way or another. It is simply to underline two facts about technology for autonomous use in language teaching: 1) technology-enhanced teaching should appeal to the learner; otherwise, the student may not use engage with the materials, and 2) there must be a well-founded, research-based reason for using the materials; otherwise, the students could be wasting their time on ineffective activities.

7.6 Final Remarks

It is important to keep in mind that a major reason for choosing the platforms used here (i.e., Facebook and YouTube) was their popularity. It is a well-known fact that most products undergo what is known as a 'product life cycle', which eventually ends with the product going out of fashion (Klepper, 1996). Similarly, Facebook and YouTube may eventually be replaced by other social media platforms. Both platforms have gone from strength to strength in recent years, but that does not mean they will remain popular forever, as the decline of Myspace clearly shows. However, this does not subtract from the contributions of the studies presented here, as the essence of the studies was the language learning materials used in the courses, namely the videos. Thus, even if users should abandon Facebook and/or YouTube, the materials can be used for any other platform that allows the uploading of video material. In the case of Facebook, even if the platform should cease to be perceived as trendy, it has the features to function as an easily accessible learning management system as pointed out by Wang et al. (2012). Thus, rather than having learners engage with lessons and activities while socialising on Facebook, they could instead enjoy the social aspect of Facebook while logged in, with learning in mind.

The materials used in the two studies presented here were in many ways simple, in the sense that they were short dialogues which were orthographically enhanced. However, the strength of these materials is that they could be accessed any time the learner could fit them into their schedule. Thus, while it is true that technology should not be used for the sake of it, one should not dismiss the use of a new technology simply because what it offers could be replicated in the classroom.

When it comes to technology and language learning, there are already several tools available for students to use outside the classroom. This study found that teaching English pronunciation can be done effectively through the use of social media, thus providing a tool for language teachers to engage their students in pronunciation activities without sacrificing their teaching time in the classroom.

Appendices

Appendix I – Background Questionnaire

1. Participant number:

   ____

2. Gender:
   - Female
   - Male
   - Other

3. Age:

   ____

4. What is your nationality?
   - Spanish
   - Dual-Spanish
   - A nationality other than Spanish

5. What is your first language?
   - Spanish
   - Spanish and another language (early bilingual)
   - A language other than Spanish

6. Do you speak any languages other than Spanish? If yes, please state which

7. How would you describe your English level?
   - C2 (Proficient)
   - C1 (Advanced)
   - B2 (Upper-intermediate)
   - B1 (Lower-intermediate)
   - A2 (Elementary)
   - A1 (Beginner)

Education and Language Use

8. How many years have you been learning English?

9. Have your teachers mainly been native speakers of Spanish or native speakers of English?
    - Mainly native speakers of Spanish
    - Mainly native speakers of English
    - Mainly native speakers of languages other than Spanish and English

10. In your English lessons, what aspects of pronunciation did you focus on the most?
    - Mainly speech sounds (vowels and consonants)
    - Mainly intonation and stress
    - An equal mix of speech sounds, intonation, and stress
    - We didn't focus on pronunciation

11. Have you ever taken a specific pronunciation course?
    - Yes
    - No

12. If you answered 'yes' to the previous question, please describe the course in a few words:

____

13. How much time do you spend speaking English outside of class?
    - Less than an hour a week
    - Less than an hour a day
    - Between one and two hours a day
    - More than two hours a day

14. How much time do you spend listening to English outside of class?
    - Less than an hour a week
    - Less than an hour a day
    - Between one and two hours a day
    - More than two hours a day

15. Have you ever spent time in an English-speaking country?
   - Yes
   - No

16. If you answered 'yes' to the question above, please state how much time you have spent in an English-speaking country:

   ____

17. To what extent do you agree with the following statement 'Sounding like a native speaker is important to me

	1	2	3	4	5	6	7	
Completely disagree	○	○	○	○	○	○	○	Completely agree

Use of Technology

18. Are you on Facebook?
   - Yes
   - No

19. If you answered 'yes' to the question above, please state how often you use Facebook
   - Several times a day
   - Once or twice a day
   - A few times a week
   - Less than once a week

20. If you are on Facebook, what do you use Facebook for? (You can select more than one)
   - See what my friends are doing
   - Chatting with friends in English
   - Chatting with friends in Spanish
   - Following the news
   - Learning about things that interest me
   - Other: ____

21. Do you use YouTube?
    - Yes
    - No

22. If you answered 'yes' to the question above, please state how often you use YouTube
    - Several times a day
    - Once or twice a day
    - A few times a week
    - Less than once a week

23. If you use YouTube, what do you use if for? (You can select more than one)
    - Streaming music
    - Entertainment
    - Learning about things that interest me
    - Other:____

Appendix II – Test items – Study 1

Facebook

Production test (Pre-test)

Single stress (18 Items)

Carrot cake
Chocolate cake
Housekeeper
Orange juice
Painkiller
Regent Street
Skyscraper
Stanley Street
Temple Street
Banana bread
Bricklayer
Brook Street
Downing Street
Lemon cake
Lord Street
Melon juice
Proof-reader
Taxi driver

Double stress (10 Items)

Bacon sandwich
Euston Road
Fruit salad
Gold watch
Hyde Park
Meat pie
Oxford Road
Plastic bag
Prenton Park
Third Avenue

Facebook

Production test (Post-test)

Single stress (18 Items)

Carrot cake
Chocolate cake
Housekeeper
Orange juice
Painkiller
Regent Street
Skyscraper
Stanley Street
Temple Street
Cup holder
Drug dealer
Edmund Street
Leeds Street
Lie detector
Pineapple juice
Strawberry cake
Tomato juice
Victoria Street

Double stress (10 Items)

Apple pie
Brooklyn Bridge
East Village
Fourth Avenue
Ham Sandwich
Hamilton Square
Moscow Road
Noodle soup
Plastic spoon
Silver plate

Facebook

Perception test (Pre-test)

Single stress (17 Items)

Carrot cake
Dishwasher
Edmund Street
Grape juice
Orange juice
Skyscraper
Songwriter
Stanley Street
Temple Street
Bottle opener
Bull fighter
Chocolate cake
Cornflakes
Hood Street
Lord Street
Painkiller
Paradise Street

Double stress (17 Items)

Acid rain
Civil war
Clock radio
Ground floor
Group therapy
Kitchen sink
School uniform
Square root
Baby boy
Citric acid
Direct object
Native speaker
Natural science
Prime Minister
Red Cross
Single bed
Twin brother

Facebook

Perception (Post-test)

Single stress (17 items)

Carrot cake
Dishwasher
Edmund Street
Grape juice
Orange juice
Skyscraper
Songwriter
Stanley Street
Temple Street
Banana bread
Blood donor
Butter cake
Goalkeeper
Henry Street
Hope Street
Record player
York Street

Double stress (17 items)

Acid rain
Civil war
Clock radio
Ground floor
Group therapy
Kitchen sink
School uniform
Square root
City centre
First aid
Front door
Front row
Giant panda
National anthem
Nuclear energy
Red carpet
Town hall

Appendix III – Test items – Study 2

Production.
Pre- and post-test:
1
A: Any idea what Fred is looking for?
B: I don't know, but it looks like he has dropped something.
2
A: What time is it?
B: I don't know. My watch has stopped.
3
A: Can you help me with something for a second?
B: Not right now, sorry. My phone is ringing.
4
A: Did you find that book you've been looking for?
B: Yes, finally. I thought I'd left it on the bed, but I found it under the bed.
5
A: What's the difference between African and Indian elephants?
B: African elephants have big ears. Indian elephants have smaller ears.
6
A: Did you hear the news about that attack in Paris?
B: Yeah! Horrible! What's wrong with people?
7
A: Did you hear what happened in Manchester yesterday?
B: Yes! 20 prisoners escaped.
8
A: Wow… look at those black clouds…
B: Yeah, it looks like there's rain coming.
9
A: I got you the flowers you asked for.
B: But… these are blue roses, I asked for red roses.
10
A: How is your paper going?
B: Great, thanks! I really feel like I'm getting somewhere.
11
A: What do you think about the new teacher?
B: I quite like the guy.
12
A: There you are! Why are you late?
B: What do you mean? It's ten past eight.
A: I said to meet at ten to eight, not ten past eight.
13
A: I heard your brother won the lottery.
B: No – it was my mum who won the lottery.
14
A: How do you think we get in to the museum?
B: I don't know. We may have to ask someone.
15
A: Is it time to go down to the platform yet?
B: Yeah, let's go – the train is coming.

16
A: Did H&M have that jumper you wanted?
B: No. They only had grey jumpers, but I wanted a black jumper.
17
A: Anna has been looking really sad lately.
B: Yeah, haven't you heard? Her mum died.
18
A: This bar is much too noisy.
B: Yeah, you're right. Let's go back to my place.

Post-test only:

19
A: Why are there so many people over there?
B: I think an accident has happened.
20
A: The Lion King is my favourite Disney film.
B: Really??? It's definitely not my favourite.
21
A: What did I miss while I was out?
B: Your friend Jane called.
22
A: John has seemed really happy lately.
B: Yes. I've heard he's seeing someone.
23
A: Do you want me to get you anything from IKEA?
B: Yes, please – could you get me two black bowls and two white bowls?
24
A: Wow! It's really hot in here.
B: I know! And the fan broke yesterday.
25
A: Do you know what I've always wondered?
B: What?
A: Are zebras black with white stripes or white with black stripes?
26
A: Why is the dog barking.
B: He must have heard something.
27
A: I'm really bored.
B: If you want, we can go for a drive somewhere.

Perception (Pre-test)

1. Are you going somewhere?
2. Did you drop something?
3. I like the guy.
4. John is seeing things.
5. The dog heard something.
6. You should see someone.
7. About twenty prisoners escaped.
8. My watch has stopped.
9. Our train is coming.
10. The fan broke yesterday.
11. The taxi is waiting.
12. Yesterday my dog died.
13. Ten minutes past six.
14. Black coffee is nice.
15. I like that painting.
16. I needed red roses.
17. These bowls are white.
18. Zebras have black stripes.
19. I need more paint.
20. Here comes the sun.
21. I bought fresh meat.
22. Peter wrote a book.
23. Look through the window.
24. My teacher is funny.

Perception (Post-test)

1. Are you going somewhere?
2. Did you drop something?
3. He should take something.
4. I hate the guy.
5. I like English people.
6. I like the guy.
7. John is seeing things.
8. The dog heard something.
9. You should see someone.
10. About twenty prisoners escaped.
11. My car won't start.
12. My finger is hurting.
13. My watch has stopped.
14. Our train is coming.
15. The bus has stopped.
16. The fan broke yesterday.
17. The taxi is waiting.
18. Yesterday my dog died.
19. Ten minutes past six.
20. Black coffee is nice.
21. Elephants are big animals.
22. I like that painting.
23. I needed red roses.
24. I said five dollars.
25. I think Belgium won.
26. These bowls are white.
27. Zebras have black stripes.
28. I need more paint.
29. Here comes the sun.
30. I bought fresh meat.
31. Peter wrote a book.
32. Look through the window.
33. My teacher is funny.

Appendix IV – Test instructions

Study 1

In the following test you will be asked to listen to several compound nouns. These are two words that are put together to form a new noun e.g., *police car*. *Police* is a noun. *Car* is a noun. Together they form a third noun *police car*.

You have to decide which of the two elements has the most stress. If you think the first element has the most stress (poLIce car), you choose the option 'Single stress'. If you think the last element has the most stress (police CAR), you choose 'Double stress'. If you think both words are stressed the same (poLIce CAR), you also choose 'Double stress'.

Study 2

In the following test you will be asked to listen to a number of sentences. Each sentence has exactly four words. You have to decide which of the words has the most stress – which word 'stands out more'. For example, in the sentence *John STOLE the bag*, *stole* stands out.

On the screen, you will see four buttons labelled 'One', 'Two', 'Three', and 'Four'. Each number represents a word in the sentence. 'One' is the first word, 'Two' is the second word and so on. So, in the sentence *John STOLE the bag* the answer is 'Two' because the second word in the sentence has the most stress.

Before the real test starts, you get four practice sentences to help you understand how the test and the software work.

References

Abberton, E., & Fourcin, A. (1975). Visual feedback in the acquisition of intonation. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development 2* (pp. 157–165). Academic Press.

Abberton, E., Howard, D., & Fourcin, A. (1989). Laryngographic assessment of normal voice: a tutorial. *Clinical Linguistics and Phonetics, 3*, 281–296.

Abe, A. (2011). Effects of form–focused instruction on the acquisition of English weak forms by Japanese EFL learners. In W.–S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (184–187). Hong Kong.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language: Listener perception vs. linguistic scrutiny. *Language Learning 59*, 249–306.

AbuSa'aleek, A. O. (2015). Students' perceptions of English Language Learning in the Facebook Context. *Teaching English with Technology, 15*(4), 60-75.

AbuSeileek, A. F. (2007). Computer-assisted pronunciation instruction as an effective means for teaching stress. *The JALT CALL Journal, 3(1-2)*, 3-14.

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly 18*(1), 69–83.

Adler–Bock, M., Bernhardt, B., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of English /r/ in two adolescents. *American Journal of Speech–Language Pathology. 6*(2), 128–139.

Akahane–Yamada, R., Adachi, T., & Kawahara, H. (1997). Second language production training using spectrographic representations as feedback. *Journal of the Acoustical Society of Japan, 18,* 341–343.

Akahane–Yamada, R., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer–based second language production training by using spectrographic representation and HMM–based speech recognition scores. *Proceedings of ICSLP*, Sydney, Australia.

Akahane–Yamada, R., Tohkura, Y., Bradlow, A., & Pisoni, D. (1996). Does training in speech perception modify speech production? *Proceedings of the International Conference on Spoken Language Processing*. (pp. 606–609). Philadelphia, PA.

Albertson, K. (1982). Teaching pronunciation with visual feedback. *NALLD Journal, 17*, 18–33.

Alderson, J., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude, and language proficiency. *Language Teaching Research, 1*, 93–121.

Al-Hejin, B. (2004). Attention and Awareness: Evidence from Cognitive and Second Language Acquisition Research. *Working Papers in TESOL and Applied Linguistics 4*(1). Teachers College, Columbia University.

Aliaga–García, C. (2007). The Role of Phonetic Training in L2 Speech Learning. *Proceedings of the Phonetics Teaching and Learning Conference* (1–5), University College London.

Aliaga–García, C. (2009). Effects of audiovisual auditory vs. articulatory training on L2 vowel category learning. *Proceedings of the PTLC 2009 Phonetics Teaching and Learning Conference*. University College London.

Aliaga–García, C. (2017). *The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan–Spanish learners of English*. Unpublished doctoral dissertation. University of Barcelona.

Aliaga–García, C., & Mora, C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent research in second language phonetics/phonology: Perception and production*, (2–31). Cambridge Scholars Publishing.

Altmann, H. (2006). *The perception and production of second language stress: A crosslinguistic experimental study*. Unpublished doctoral dissertation. University of Delaware.

Alwehaibi, H. O. (2015). The impact of using YouTube in EFL classroom on enhancing EFL students' content learning. *Journal of College Teaching & Learning (Online), 12*(2), 121–126.

Available at: https://files.eric.ed.gov/fulltext/EJ1061416.pdf.

Amer, M., & Amer, W. (2011). The role of explicit instruction in English word stress patterns in an EFL Arab university context. *Indonesian Journal of English Language Teaching 7*(2), 71–87.

Anderson–Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field–specific materials. *English for Specific Purposes*, *9*, 195–214.

Anderson–Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5–22.

Anderson–Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529–555.

Anderson–Hsieh, J., & Kohler, K. (1988). The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension. *Language Learning 38*(4), 561–611.

Archibald, J. (1993). Metrical phonology and the acquisition of L2 stress. In F. R. Eckman (Ed.), *Confluence: linguistics, L2 acquisition and speech pathology*, (pp. 37–48). John Benjamins.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics, 35*, 167–181.

Armstrong, L. (1932). *The Phonetics of French.* Bell and Hyman.

Arndt, H. L., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning & Technology, 22*(1), 124–142.

Arnold, J. H., & Brown, D. (1999). A map of the terrain. In J. Arnold (Ed.), *Affect in Language Learning* (1–24). Cambridge University Press.

Arzu, E. (2014). Facebook groups as a supporting tool for language classrooms. *Turkish Online Journal of Distance Education*, *15*(3), 18–26.

Ashby M., & Maidment, J. (2005). *Introducing Phonetic Science*. Cambridge University Press.

Asher, J. J. (1969). The Total Physical Response approach to second language learning. *The Modern Language Journal, 53*, 3–17.

Baker, A., & Murphy, J. (2011). Knowledge base of pronunciation teaching: Staking out the territory. *TESL Canada Journal, 28*(2), 29–50.

Balbay, S., & Kilis, S. (2017). Students' Perceptions of the Use of a YouTube Channel Specifically Designed for an Academic Speaking Skills Course. Eurasian *Journal of Applied Linguistics, 3*(2), 235–251.

Bang, J., Kang, S., & Lee, G. (2013). An automatic feedback system for English speaking integrating pronunciation and prosody assessments. *Proceedings of SLaTE,* 83–89.

Barriuso, T., & Hayes–Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal, 30*(1), 177–194.

Barrot, J. (2018). Facebook as a learning environment for language teaching and learning: A critical analysis of the literature from 2010 to 2017. *Journal of Computer Assisted Language Learning, 34*, 863–875.

Beckman, M.E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In A. Keating (Ed.), *Phonological structure and phonetic form. Papers* *in Laboratory Phonology 3*, 7–33.

Bennett, S. J., Maton, K. A., & Kervin, L. K. (2008). The 'digital natives' debate: a critical review of the evidence. *British Journal of Educational Technology, 39*(5), 775–786.

Benrabah, M. (1997). Word–stress – a source of unintelligibility in English. *IRAL, 35*(3), 157–165. Benson, P. (2015). Commenting to learn: Evidence of language and intercultural learning in comments on YouTube videos. *Language Learning & Technology, 19*(3), 88–105.

Benson, P. (2015). Commenting to learn: Evidence of language and intercultural learning in comments on YouTube videos. *Language Learning & Technology, 19*(3), 88–105.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn & M. J. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 331-347). John Benjamins.

Benzeguiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., & Ris, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Commun*ication, *49,* 763–786.

Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler–Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics, 19*(6/7), 605–617.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second–language speech perception: Commonalities and complementarities. In M. J. Munro, & O. S. Bohn (Eds.), Second language speech learning: The Role of language experience in speech perception and production. (pp. 13–34). John Benjamins.

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, *13*, 116–137.

Blattner, G., & Fiori, M. (2009). Facebook in the Language Classroom – Promises and Possibilities. *Instructional Technology and Distance Learning, 6*(1), 17–28.

Blattner G., & Fiori, M. (2011). Virtual Social Network Communities – An investigation of language learners' development of sociopragmatic awareness and multiliteracy skills, *CALICO Journal*, *29*(1), 24–43.

Blattner, G., & Lomicka, D. (2012). Facebook–ing and the social generation: A new era of Language learning, *Alsic, 15*(1).
Available at: http://alsic.revues.org/2413.

Bloomfield, L. (1933). *Language*. Harper & Row.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.35. Retrieved 15th of February, 2020 from: http://www.praat.org/.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). York Press.

Bohn, O.–S. (2005). A fond farewell to the Critical Period Hypothesis for nonprimary language acquisition. In A. Saleemi, O.–S. Bohn, & A. Gjedde (Eds.), *In search of a language for the mind–brain: Can the multiple perspectives be unified?* (285–310). Aarhus University Press.

Bohn, O.–S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language Experience in L2 vowel perception. *Applied Psycholinguistics, 11*, 303–328.

Bongaerts, T., Mennen, S., & van der Silk, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica 54,* 298–308.

Bolinger, D. (1958). On Intensity as a Qualitative Improvement of Pitch Accent. *Lingua, 7*, 175–182.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language 37*(1), 83–96.

Bolinger, D. (1985). Two views of Accent. *Journal of Linguistics, 21*, (79–124).

Bolinger, D. (1986). *Intonation and its parts.* Stanford University Press.

Bosch, T. E. (2009). Using online social networking for teaching and learning: Facebook use at the University of Cape Town. *Communication, 35*(2), 185–200.

Bowen, D. (1972). Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly, 6*, 83–94.

Bowen, D. (1975). *Patterns of English pronunciation.* Newbury House.

Bowers–Campbell, J. (2008). Cyber 'pokes': Motivational antidote for developmental college readers. *Journal of College Reading and Learning, 39*(1)*,* 74–87.

Bradlow, A. R., Pisoni, D. B., Akahane–Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.

Brawerman–Albini, A., & Becker, M. R. (2014). Perception and production of English stress by Brazilian speakers. *Concordia Working Papers in Applied Linguistics, 5*, 73–84.

Brawerman–Albini, A., Kluge, D. C., Silva, A. H. & Albini, L. C. P. (2017). Perceptual training effects on the acquisition of English stress by Brazilian learners. *International Journal of English Linguistics, 7*(6), 1–15.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2002). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19,* 51–61.

Brick, B. (2011). How effective are web 2.0 language learning sites in facilitating language learning? *Compass: The Journal of Learning and Teaching at the University of Greenwich, 3*, 57–63.

Brown, G., & Yule, G. (1983). Teaching the spoken language: An approach based on the conversational analysis of English. Cambridge University Press.

Buga, R., Căpeneață, I., Chirasnel, C., & Popa, A. (2014). Facebook in foreign language teaching – A tool to improve communication competences. *Procedia Social and Behavioral Sciences, 128*, 93–98.

Burleson, D. F. (2007). Improving intelligibility of non–native speech with computer assisted phonological training*, IULC Working Papers Online, 7*. Available at: https://scholarworks.iu.edu/journals/index.php/iulcwp/article/view/25801.

Burston, J. (2014). The reality of MALL: still on the fringes. *CALICO Journal, 31*, 103–125.

Busà, M. G. (2008). New perspectives in teaching pronunciation. In A. Baldry, M. Pavesi, C. Taylor Torsello, & C. Taylor (Eds.), *From DIDACTAS to ECOLINGUA. An ongoing research project on translation and corpus linguistics* (165–182). Università degli Studi di Trieste.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1–47.

Cardoso, W. (2018). Learning L2 pronunciation with a text–to–speech synthesizer. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds.), *Future–proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018,* 16–21.

Carlet, A. (2017). L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross–training study. Unpublished PhD dissertation. Universitat Autònoma de Barcelona.

Carlet, A. (2019). Different high variability procedures for training L2 vowels and consonants. In *Proceedings of the International Congress of Phonetic Sciences* (944–948). Melbourne, Australia.

Carlet, A., & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short–term high variability training study. *Concordia University Working Papers in Applied Linguistics*, 85–98.

Carpenter, A. (2015). Phonetic training significantly mitigates the stress 'deafness' of French speakers. *International Journal of Linguistics, 7*(3), 94–108.

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal, 54*, 477–481.

Cauldwell, R. T. (1996). Stress-timing: Observations, beliefs, and evidence. *Eger Journal of English Studies, 1*, 33-48

Cebrian J., & Carlet, A. (2014). Second language learners' identification of target language phonemes: A short–term phonetic training study. *Canadian Modern Language Review, 70*(4), 474–499.

Celce–Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation – A course book and reference guide*. Cambridge University Press.

Celik, M. (2004). Teaching English intonation to EFL/ESL students. In M. Raman (Ed.), *English Language Teaching*. Atlantic Publishers and Distributers.

Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2015). Listen, attend and spell. arXiv:1508.01211.

Chapelle, C. (2003). *English language learning and technology*. John Benjamins.

Chinnery, G. M. (2006). Going to the MALL: Mobile assisted language learning. *Language Learning & Technology, 10*(1), 9–16.

Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese–English bilingual listeners. *Cognition, 189*, 188–192.

Chomsky, N. (1959). A review of B.F. Skinner's 'Verbal Behavior'. *Language, 35*(1), 26–58.

Chomsky, N. (1968). Language and the mind. Cambridge University Press.

Chomsky, N. (1980). In M. Piatelli–Palmerini (Ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Harvard University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* Harper and Row.

Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal,* 7(1), 21–47.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 61–77.

Chun, D. M. (2002). *Discourse Intonation in L2: From theory and research to practice.* John Benjamins.

Chun, D. M., Hardison, D. M., & Pennington, M. C. (2008). Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards &, M. L. Zampini (Eds.), *Phonology and second language acquisition* (323–346). John Benjamins.

Chun, D. M., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal, 100*(1), 64–80.

Coffield, F. (2013). Learning styles: Time to move on. *Lifelong Learning in Europe 4*. 2013

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second Language Speech: Theory and Practice*. Cambridge University Press.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*, 49–64.

Coniam, D. (2002). Technology as an awareness–raising tool for sensitising teachers to features of stress and rhythm in English. *Language Awareness, 11*(1), 30–42.

Counselman, D. (2010). *Improving pronunciation instruction in the second language classroom*. Unpublished Ph.D. dissertation. Pennsylvania State University.

Couper, G. (2006). The short and long–term effects of pronunciation instruction. *Prospect, 21,* 46–66.

Couper, G. (2011). What makes pronunciation teaching work? Testing the effect of two variables: Socially constructed metalanguage and critical listening. *Language Awareness, 20*(3), 159–182.

Couper, G. (2012). Teaching word stress: learning from learners' perceptions. *TESOL in Context* S3.

Available at: http://www.tesol.org.au/Publications/Special-Editions.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Edward Arnold.

Cruttenden, A. (1990). Nucleus placement and three classes of exceptions. In S. Ramsaran (Ed.), *Studies in the pronunciation of English: A commemorative volume in honour of A. C. Gimson* (pp. 9–18). Routledge.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Cruttenden, A. (2014). *Gimson's Pronunciation of English*. Routledge.

Cucchiarini, C., Neri, A. and Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication, 51*(10), 853–863.

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp.264–289). Blackwell.

Cutler, A., & Clifton, C. J. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Erlbaum.

Cutler, A., Pasveer, D. 2006. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. *Proceedings of the 3rd International Conference on Speech Prosody* (237–400). Dresden.

Dalby, J., & Kewly–Port, D. (1999). Explicit pronunciation training using automatic speech recognition. In M. Holland (Ed.), *Tutors that listen: Speech recognition for language training*, Special issue of the Journal of the Computer Assisted Language Learning Consortium, *CALICO, 16*(5), 425–445.

Dalby, J., Kewly–Port, D., & Sillings, R. (1998). Language–specific pronunciation training using automatic speech recognition technology. In *Proceedings of the European Speech Communication Association Conference on Speech Technology in Language Learning*, 25–28.

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford University Press.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In. J. Levis, & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, 93–108. Iowa State University.

Dauer, R. (1993). *Accurate English: A complete course in pronunciation*. Longman.

de Bot, K. (1980). The role of feedback and feedforward in the teaching of pronunciation. *System, 8*, 35–47.

de Bot, K. (1983). Visual feedback of English intonation, I: and induced practice behavior. *Language and Speech, 26*, 331–350.

de Bot, K., & Mailfert, K. (1982). The Teaching of Intonation: Fundamental Research and Classroom Application. *TESOL Quarterly, 16*(1), 71–77.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long

(Eds.), *Handbook of second language acquisition* (pp. 313–348). Blackwell.

DeKeyser, R. (2006). A critique of recent arguments against the critical period hypothesis. In C. Abello–Contesse, R. Chacon–Beltran, M. D. Lopez–Jiménez, & M. M. Torreblanca–Lopez (Eds.), *Age in L2 acquisition and teaching* (49–58). Peter Lang.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t–test instead of Student's *t*–test with unequal group sizes. *International Review of Social Psychology, 30*, 92–101.

DeLuca, V., Miller, D., Pliatsikas, C., Rothman, J., Schwieter, J., & Paradis, M. (2019). Brain adaptations and neurological indices of processing in adult second language acquisition: Challenges for the Critical Period Hypothesis. In J. W. Schwieter, & M. Paradis (Eds.), *The Handbook of the neuroscience of multilingualism* (170–196). Wiley–Blackwell. doi.org/10.1002/9781119387725.ch8.

Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 59*, 547-566.

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen–Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (347–369). John Benjamins.

Derwing, T. M. (2018). The efficacy of pronunciation instruction. In O. Kang, R. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (320–334). Routledge.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: a research–based approach. *TESOL Quarterly, 39*(3), 379–398.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476–490

Derwing, T. M., & Munro, M. J. (2015). Pronunciation Fundamentals: Evidence–based Perspectives for L2 Teaching and Research. John Benjamins.

Derwing, T. M., Munro, M. J., & Carbonaro, M. D. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*, 592–603.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29*, 359–380.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393–410.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*, 155–166.

Dewaele, J.–M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274.

Dey, P. (2012). *Visual speech in technology–enhanced learning*. Unpublished Ph.D. dissertation. University of Sheffield.

Dickerson, W. B. (2004). Should we teach construction stress to ESL/EFL students? *Jurnal Pendidik Dan Pendidikan*, *19*, 81–96.

Dimitrova, S., & Chernogorova, T. (2012). English pronunciation models and tertiary-level students: A Bulgarian perspective. In T. Paunović & B. Čubrović (Eds.), *Exploring English Phonetics* (pp. 208–224). Cambridge Scholars Publishing.

Dlaska, A., & Krekeler, C. (2013). The short–term effects of individual corrective feedback on L2 pronunciation. *System, 41*(1), 25–37. http://dx.doi.org/10.1016/j.system.2013.01.005

Dornyei, Z. (2010). Questionnaires in second language research: Construction, administration, and processing. Routledge.

Dowd A., Smith J., & Wolfe, J. (1998). Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech, 41*(1), 1–20.

Ducate, L., Lomicka, L., & Lord, G. (2012). Hybrid learning spaces: Re–envisioning language learning. In J. Thoms, & F. Rubio (Eds.), *Hybrid language teaching and learning* (67–91). Cengage Learning.

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology, 13*(3), 66–86.

Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education, 20*(2), 129–136.

Dupoux, E., Pallier, C., Sebastian–Gallés, N. & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language, 36*(3), 406–421.

Dupoux, E., Sebastián–Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106*(2), 682–706.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34,* 917–928.

Eckhart, R. A. (2009). Teaching Pronunciation with Facebook and Photobooth. *Ohio TESOL Journal, 2*(1).

Available at: http://works.bepress.com/bobeckhart/3/.

Egan, K. (1999). Speaking: A Critical Skill and Challenge. *CALICO Journal, 16*, 277–293.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer–aided language learning: Strengths and limitations of a new CALL paradigm. *Language*

Learning & Technology, 2(1), 45–60.

Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *JALT CALL Journal, 10*(1), 21–47.

Elliott, R. (1995). Foreign language phonology – Field independence, Attitude, and the success of Formal Instruction in Spanish pronunciation. *The Modern Language Journal, 79*(4), 530–542.

Engwall, O. (2008). Can audio–visual instructions help learners improve their articulation? An ultrasound study of short term changes. *Proceedings of Interspeech 2008*, 2631–2634.

Engwall, O., Bälter, O., Öster, A–M., & Kjellström, H. (2006). Designing the user interface of the computer–based speech training system ARTUR based on early user tests. *Journal of Behavioural and Information Technology, 25*(4), 353–365.

Eren, Ö. (2012). Students' attitudes towards using social networking in foreign language classes: A Facebook example. *International Journal of Business and Social Science, 3*(20), 288–294.

Eskenazi, M. (1996). Detection of foreign speakers' pronunciation errors for second language training – preliminary results. 4th *International Conference of Spoken Language Processing*. Philadelphia, PA, USA.

Eskenazi, M. (1999a). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal, 16*(3), 447–469.

Eskenazi, M. (1999b). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology, 2*(2), 62–76.

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication, 51*(10), 832–844.

Estebas–Vilaplana, E. (2015). The learning of English intonation by Spanish speakers in a distance education environment. In: J. Przedlacka, J. Maidment, & M. Ashby (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference* (39–43). UCL, London.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly 39*, 399–423.

Finocchiaro M., & Brumfit, C. (1983). *The functional–notional approach: from theory to practice*. New York: Oxford University Press.

Firth, M., & Mesureur, G. (2010). Innovative uses for Google Docs in a university language program. *The JALT CALL Journal, 6*(1), 3–16.

Fleck, B. K., Beckman, L. M., Sterns, J. L., & Hussey, H. D. (2014). YouTube in the Classroom: Helpful Tips and Students Perceptions. *The Journal of Effective Teaching, 14*(3), 21–37.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76*, 692–707.

Flege, J. E. (1987a). A critical period for learning to pronounce foreign languages? *Applied Linguistics, 8*(2), 162–177.

Flege, J. E. (1987b). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification, *Journal of Phonetics, 15*, 47–65.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders – a review* (pp. 224-401). Ablex.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross–language research* (233–276). Timonium, MD: York Press.

Flege, J. E. (2009). Give input a chance! In T. Piske, & M. Young–Scholten (Eds.), *Input matters in SLA* (175–190). Multilingual Matters.

Flege, J. E., & Bohn, O.–S. (1989). An instrumental study of vowel reduction and stress placement in Spanish accented English. *Studies in Second Language Acquisition, 11*, 35–62.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native–language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*, 169–186.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1–34.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23*, 567–598.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125–3134.

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation, 3*(1), 34–56.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal 29*(1), 1–22.

Fourcin, A. J. & Abberton, E. (1977). Laryngograph studies of vocal–fold vibration. *Phonetica 34*, 313–315.

Fouz–González, J. (2015). Trends and directions in computer–assisted pronunciation training. In J. A. Mompeán & J. Fouz–González, *Investigating English pronunciation: Trends and directions* (314–342). Palgrave Macmillan.

Fouz–González, J. (2017). Pronunciation instruction through Twitter: the case of commonly mispronounced words, *Computer Assisted Language Learning, 30*(7), 631–663. doi:10.1080/09588221.2017.1340309.

Fouz–González, J. (2019). Podcast–based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features. *ReCALL, 31*(2), 150–169.

Fouz–González, J., & Mompeán, J. A. (2012). Twitter and English pronunciation teaching. In Pixel (Ed.), *5th ICT for Language Learning* (341–344). Conference Proceedings. Florence, Italy.

Fouz–González, J., & Mompeán, J. A. (2021). Exploring the potential of phonetic symbols and keywords as labels for perceptual training. *Studies in Second Language Acquisition*, *43*(2), 297–328. doi:10.1017/S0272263120000455.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*, 1668–1680.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366. doi:10.1037//0096–1523.28.2.349.

Franke, F. (1890). Die praktische Spracherlernung auf Grund der Psychologie und der Physiologie der Sprache dargestellt. O.R. Reisland.

Fraser, H. (2001). *Teaching Pronunciation: A handbook for teachers and trainers*. TAFE NSW Access Division.

Fraser, H. (2006a). Helping teachers help students with pronunciation: A cognitive approach. *Prospect, 21*, 80–86.

Fraser, H. (2006b). Phonological concepts and concept formation. *IJES, International*

Journal of English Studies 6(2), 55–76.

Frieda E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception and Psychophysics, 61*(5), 561–577.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech 1*(2), 126–152.

Fryer, L., & Carpenter, R. (2006). Emerging technologies: BOTS as language learning tools. *Language Learning & Technology, 10*(3), 8–14.

Fullana, N., & Mora, J. (2009). Production and perception of voicing contrasts in English word–final obstruents: Assessing the effects of experience and starting age. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (97–117). Cambridge Scholars Publishing.

Gallacher, A., Thompson, A., & Howarth, M. (2018). "My robot is an idiot" – Students' perceptions of AI in the L2 classroom. In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouësny (Eds.), *Future–proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (70–76). doi.org/10.14705/rpnet.2018.26.9782490057221

García Lecumberri, M. L. (1995). *Intonational Signalling of Information Structure in English and Spanish*. Bilbao, Servicio de Publicaciones de la Universidad del País Vasco.

García Lecumberri, M. L. (1997). Perceptibility of nuclear focus in English. *Barcelona English language and literature studies*, 175–186.

García Lecumberri, M. L. (2000). Identificación del foco entonativo inglés por estudiantes frente a hablantes nativos. *Estudios de Lingüística Aplicada 32,* 19-28.

García Lecumberri, M. L. (2006). Perception of double stress by Spanish learners of English. In: B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin beat: Studies in Portuguese/Spanish–English interphonology* (185–198). John Benjamins.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. García Mayo, & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (115–135). Multilingual Matters.

Gattegno, C. (1972). *Teaching foreign languages in schools – The silent way.* Educational Solutions World Wide.

Ghani, M. B. A. (2015). Using Facebook in Teaching and Learning English. *3rd International Conference on Language, Literature, Culture & Education (ICLLCE)*, 97–102.

Gick, B., Bernhard, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen, & M. L Zampini (Eds.), *Phonology and Second Language Acquisition* (309–322). John Benjamins.

Giegerich, H. (2009). The English compound stress myth. *Word Structure, 2*, 1–17.

Gilbert, J. (2008). Teaching pronunciation using the prosody pyramid. Cambridge University Press.

Glass, J., & Zue, V. (2003). 6.345 Automatic Speech Recognition. Spring 2003. Massachusetts Institute of Technology: MIT OpenCourseWare https://ocw.mit.edu License: Creative Commons BY–NC–SA.

Gluszek A., & Dovidio, J. F. (2010). The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychological Review, 14*(2):214–237.

Godwin–Jones, R. (2011). Emerging Technologies – mobile apps for language learning. *Language Learning & Technology, 15*(2), 2–11.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L. & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105.

Gómez González, M., & Sánchez Roura, T. (2016). *English pronunciation for speakers of Spanish: From theory to practice*. De Gruyter Mouton.

Gómez Lacabex, E., & García Lecumberri, M. L. (2010). Investigating training effects in the production of English weak forms by Spanish learners. In K. Dziubalska Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech New Sounds* (137–143). Adam Mickiewicz University.

Gómez Lacabex, E., García Lecumberri, M. L., & Cooke, M. (2008). Identification of the contrast full vowel–schwa: training effects and generalization to a new perceptual context. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies, 55*, 173–196.

Gómez, P., Álvarez, A., Martínez, R., Bobadilla, J., Bernal, J., Rodellar, V., & Nieto, V. (2008). Applications of formant detection in language learning. In V. M. Holland, & F. P. Fisher (Eds.), *The path of speech technologies in computer assisted language learning* (44–65). Routledge.

Goodwin, J. (2014). Teaching pronunciation. In M. Celce–Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (136–152). National Geographic Learning/Cengage Learning.

Gorard, S. (2015). Introducing the mean absolute deviation 'effect' size. *International Journal Research and Methods in Education, 38*(2), 108–114.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation, 2*, 56–92.

Graham, C. (1978). *Jazz chants*. Oxford University Press.

Grandgent, C. H. (1892). The teaching of French and German in our Public High Schools. *School and College, 1*(3), 148–152.

Grosseck, G., & Holotescu, C. (2008). Can we use Twitter for educational activities? *The 4th International Scientific Conference eLSE "eLearning and Software for Education*" (17–18). Bucharest.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.–S. Bohn &, M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (55–77). John Benjamins.

Guiora, A. Z. (1972). Construct validity and transpositional research: toward an empirical study of psychoanalytic concepts. *Comprehensive Psychiatry 13*, 139–150.

Guiora, A. Z., Beit–Hallahmi B., Brannon, R. C. L., Dull, C. Y., & Scovel, T. (1972). The effects of experimentally induced changes into ego states on pronunciation ability in a second language: an exploratory study. *Comprehensive Psychiatry, 13,* 421–428.

Gussenhoven, C. (1983). Focus, mode, and the nucleus. *Journal of Linguistics 19,* 377–417.

Gussenhoven, C. (2007). Types of focus in English. In D. Büring, M. Gordon, & C. L. Lee (Eds.), *Topic and Focus: Papers from a Workshop on Intonation and Meaning*, (pp. 83–101). Springer.

Gutiérrez–Díez, F. (2005). Intonation focus in the interlanguage of a group of Spanish learners of English. *Revista Alicantina de Estudios Ingleses, 18*, 129–147.

Gutiérrez-Díez, F. (2008). Contrastive intonation and error analysis. Tonality and tonicity in the language of a group of Spanish learners of English. In M. A. Gómez, J. L. Mackenzie, & E. M. González (Eds.), *Languages and Cultures in Contrast and Comparison* (327–354). John Benjamins.

Gutiérrez–Díez, F. (2012). Trouble Spots in the Learning of English Intonation by Spanish Speakers. Tonality and Tonicity. In: J. Romero–Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching. Educational Linguistics,* 15 (219–230). Springer. doi.org/10.1007/978–94–007–3883–6_13.

Hafner, C. A., & Miller, L. (2011). Fostering learner autonomy in English for science: A collaborative digital video project in a technological learning environment. *Language Learning & Technology, 15*(3), 68–86.

Hahn, L. (1999). *Native speakers' reactions to non-native speech in English discourse.* Unpublished Ph.D. dissertation. University of Illinois at Urbana-Champaign

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201–223.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Mouton.

Han, Z., Park, E., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics, 29*, 597–618.

Hancock, M. (2003). *English pronunciation in use (intermediate)*. Cambridge University Press.

Hardison, D. M. (2002). Computer–assisted second–language learning: Generalization of prosody focused training. *Proceedings of the 7th International Conference on Spoken Language Processing* (1217–1220). Denver, Colorado.

Hardison, D. M. (2004). Generalization of computer–assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34–52.

Hardison, D. M. (2005). Contextualized computer–based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal, 22*, 175–190.

Hart, A. (2001). Mann–Whitney test is not just a test of medians: differences in spread can be important. *BMJ (Clinical research ed.),* 391–393. doi.org/10.1136/bmj.323.7309.391.

Hatchavanich, D. (2014). A comparison of Type I error and power of Bartlett's Test, Levene's Test and O'Brien's test for homogeneity of variance tests. *Southeast–Asian Journal of Sciences 3*(2), 181–194.

Hattem, D., & Lomicka, L. (2016). What the Tweets say: A critical analysis of Twitter research in language learning from 2009 to 2016. *E–Learning and Digital Media, 13*, 5–23.

Heil, C. R., Wu, J. S., Lee, J. J., & Schmidt, T. (2016). A review of mobile language learning applications: Trends, challenges, and opportunities. *EUROCALL Review*, *24*(2), 32–50.

Henderson, A. (2004). The settlement experiences of immigrants (excluding refugees) in New Zealand: An overview paper completed for the Auckland Regional Settlement Strategy. *International Pacific College, Palmerston North*

Henderson, A. (2010). A Corpus–based Study of Lexical Stress Variation in American English. *Research in Language, 8*(8), 1–15.

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek–Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). The English pronunciation teaching in Europe survey: Selected results. *Research in Language, 10*(1), 5–27.

Herron, D., Menzel, W., Atwell, E., Bisiani, R., Daneluzzi, F., Morton, R., & Schmidt,

J. A. (1999). Automatic localization and diagnosis of pronunciation errors for second–language learners of English. *Eurospeech 99* (855–858). Budapest.

Hew, S.–H., & Ohki, M. (2004). Effect of animated graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal, 21*(2), 397–420.

Hewings, M. (2007). *English pronunciation in use (advanced).* Cambridge University Press.

Hickey, J. (1992). The treatment of lateral fricatives and affricates using electropalatography: A case study of a 10–year–old girl. *Clinical Speech and Language Studies, 1*, 80–87.

Hincks, R. (2001). Using speech recognition to evaluate skills in spoken English. *Papers from Fonetik 2001* (58–61). Lund University Department of Linguistics.

Hincks, R. (2002a). Speech recognition for language teaching and evaluating: A study of existing commercial products. *Proceedings of the 7th International Conference on Spoken Language Processing* (733–736). Denver, Colorado.

Hincks, R. (2002b). Speech synthesis for teaching lexical stress. In *Proceedings of Fonetik, TMH–QPSR* (pp. 153–156). Department of Speech, Music and Hearing, KTH 100 44 Stockholm.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL, 15*(1), 3–20.

Hincks, R. (2012). Suprasegmentals: Stress. In C. A. Chapelle (Ed.), *The Encyclopaedia of Applied Linguistics*. John Wiley.

Hincks, R. & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning and Technology, 13*(3), 32–50.

Hirata, Y. (2004). Computer–assisted pronunciation training for native English speakers learning Japanese pitch and duration contrasts. *Computer Assisted Language Learning, 17*(3–4), 357–376.

Hismanoglu, M. (2012). Teaching word stress to Turkish EFL (English as a Foreign Language) learners through Internet–based video lessons. *US–China Education Review, 1*, 26–40.

Holt, E. (2018). Teaching tip: The use of MRI and ultrasound technology in teaching about Spanish (and general) phonetics and pronunciation. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (252–262). University of Iowa.

Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching, 43*, 154–167.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125–132.

Howatt, A. P. R., & Widdowson, H. G. (2004). *A History of English Language Teaching*. Oxford University Press.

Hualde, J. I. (2005). *The sounds of Spanish.* Cambridge University Press.

Hualde, J. I. (2007). Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics, 6*(1), 59–89. doi.org/10.5334/jpl.145.

Hualde, J. I., & Kim, J.–Y. (2015). The acquisition of Spanish lexical stress by Korean Learners. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Hymes, D. (1972). On communicative competence. In J. B. Pride, & J. Holmes (Eds.). *Sociolinguistics* (269–293). Penguin.

Imoto, K., Tsubota, Y., Raux, A., Kawahara, T., & Dantsuji, M. (2002). Modeling and automatic detection of English sentence stress for computer assisted English prosody learning system. *Proceedings of International Conference on Spoken Language Processing* (749–752). Denver, Colorado.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Ioup, G. (1995). Evaluating the need for input enhancement in post–critical period language acquisition. In D. Singleton, & Z. Lengyel, (Eds.), *The age factor in second language acquisition* (pp. 95–123). Multilingual Matters.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first–language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866–877.

Jenkins, J. (2000). The phonology of English as an international language. Oxford University Press.

Jenkins, J. (2002). A Sociolinguistically Based Empirically Researched Pronunciation Syllabus for English as an International Language. *Applied Linguistics, 23*(1), 83–103.

Jenkins, J. (2009). English as a Lingua Franca: Interpretations and attitudes. *World Englishes, 28*(2), 200–207.

Jespersen, O. (1904). *How to Teach a Foreign Language*. Allen and Unwin.

Johnson, C. P., & Marsh, D. (2014). Blended learning – an effective solution but not without its challenges. *Higher Learning Research Communications, 4*(3), 23–41.

Jongman, A., & Wade, T. (2007). Acoustic variability and perceptual learning. In M. J. Munro, & O.–S. Bohn (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (135–150). John Benjamins.

Jongman, A., Wade, T., & Sereno, J. (2003). On improving the perception of foreign–

accented speech. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (1561–1564). Barcelona, Spain: Universitat Autònoma de Barcelona.

Juang, B. H., & Rabiner, L. R. (2005). Automatic Speech Recognition – A Brief History of the Technology Development. In K. Brown (Ed.), *Elsevier Encyclopaedia of Language and Linguistics* (67–90). Elsevier.

Jügler, J., & Möbius, B. (2015). Auditory feedback methods to improve the pronunciation of stops by German learners of French. In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow.

Junco, R., Heiberger, G., & Loken, E. (2010). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning, 27*(2), 119–132.

Kabilan, M. K., Ahmad, N., & Abidin, M. J. Z. (2010). Facebook: An online environment for learning of English in institutions of higher education? *The Internet and Higher Education, 3*(4), 179–187.

Kaltenboeck, G. (2001). A multimedia approach to suprasegmentals: Using a CD–ROM for English intonation teaching. In *Phonetics Teaching and Learning Conference Proceedings*, 19–22.

Kang, Rubin (2009) native speaker judgements.

Karch, J. D. (2021). Psychologists should use Brunner–Munzel's instead of Mann–Whitney's U test as the default nonparametric procedure. *Advances in Methods and Practices in Psychological Science, 4*(2), 1–14.

Karal, H., Kokoc, M., & Cakir, O. (2017). Impact of the educational use of Facebook group on the high school students' proper usage of language. *Education and Information Technologies, 22*(2), 677–695.

Kawai, G., & Hirose, K. (2000). Teaching the pronunciation of Japanese double–mora phonemes using speech recognition technology. *Speech Communication, 30*, 131–143.

Kelly, G. (2000). *How to Teach Pronunciation*. Longman.

Kelly, L. G. (1969). *25 centuries of language teaching*. Newbury House.

Kelsen, B. (2009). Teaching EFL to the iGeneration: A survey of using YouTube as supplementary material with college EFL students in Taiwan. *CALL–EJ Online, 10*(2).

Available at: http://callej.org/journal/10–2/kelsen.html.

Kennedy, A. S. (2015). Non-native accent and listener perceptions of grammaticality. In J. A. Mompeán & J. Fouz-González (Eds.), *Investigating English pronunciation trends and directions* (pp. 131- 146). Palgrave Macmillan.

Kétyi, A. (2013). Using Smart Phones in Language Learning – A Pilot Study to Turn CALL into MALL. In I. Bradley, & S. Thouesny (Eds.), *20 years of EUROCALL: Language Learning from the Past Looking to the Future. Proceedings of the 2013 EUROCALL Conference* (129–134). Évora, Portugal.

Kétyi, A. (2015). Practical evaluation of a mobile language learning tool in higher education. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds.), *Critical CALL – Proceedings of the 2015 EUROCALL Conference* (306–311). Padova, Italy.

Kho, M. G. W., & Chuah, K. M. (2012). Encouraging ESL discourse exchanges via Facebook: A study on engineering students. *Proceedings of 2nd Innovation Conference and Exposition*. Padang, Indonesia.

Kijak, A. (2009). *How stressful is L2 stress?: A cross–linguistic study of L2 perception and production of metrical stress*. Utrecht: LOT.

Kim, A. (2012). Investigating the effectiveness of computer–assisted language learning (CALL) in improving pronunciation: A case study. *Multimedia–*

Assisted Language Learning, 15(3), 11–33.

Kim, I–S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society, 9*(1), 322–334.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *Modern Language Journal*, *97*(3), 720–744.

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research, 19*, 254–275.

Klepper, S. (1996). Entry, Exit, Growth, and Innovation over the Product Life Cycle. *The American Economic Review; 86*(3), 562–583.

Koet, T. & van den Bergh, H. (2018). Pride and prejudice? Judging non–native pronunciation of English. Making Connections: Studies of Language and Literature Education. *L1–Educational Studies in Language and Literature*, 18, 1–13.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Krashen, S. D. (2013). The effect of direct instruction on pronunciation: Only evident when conditions for monitor use are met? *GIST Education and Learning Research Journal, 7,* 271–275.

Krashen, S. D., & Terrell, T. (1983). *The Natural Approach: Language Acquisition in the Classroom*. Pergamon Press.

Krause, C. A. (1916). *The Direct Method in Modern Languages: Contributions to Methods and Didactics in Modern Languages*. Charles Scribner's Sons.

Kröger, B. J., Graf–Bortscheller, V., & Lowit, A. (2008). Two– and three–dimensional visual articulatory models for pronunciation training and for treatment of speech disorders. In: *Proceedings of Interspeech 2008* (2639–2642). Brisbane, Queensland, Australia.

Krynicki, G., & Michalski, G. (2019). Applications of Electrography in L2 Pronunciation Teaching and Phonetic Research. In M. Wrembel, A. Kielkiewicz–Janowiak, & P. Gasiorowski (Eds.), *Approaches to the Study of Speech* (270–279). Routledge.

Kuhl, P. K. (1992). Psychoacoustics and speech perception: internal standards, perceptual anchors, and prototypes. In: L. A. Werner, & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293–332). American Psychological Association.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812–822.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross–Language Research* (pp. 121–154). York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Kukulska–Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271–289.

Ladd, D. R. (1980). *The structure of intonational meaning: evidence from English*. Indiana University Press.

Ladd, D. R. (1983). Even, focus and normal stress. *Journal of Semantics 2*, 157–170.

Ladefoged, P. & Johnson, K. (2010). *A course in phonetics* (6th ed.). Thomson Wadsworth.

Lakens, D. (2020). What is a family in family–wise error control? *The 20% Statistician.* Available at: https://daniellakens.blogspot.com/2020/03/whats–family–in–family–wise–error.html?m=0 [last accessed on 21/4/21].

Lambacher, S. (1999). A CALL Tool for Improving Second Language Acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning, 12*(2), 137–156. doi.org/10.1076/call.12.2.137.5722.

Lambacher, S., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227–247.

Larsen–Freeman D., & Long, M. H. (2014). An introduction to second language acquisition research. Routledge.

Larson–Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research 24*(1), 35–63.

Larson–Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R.* Routledge.

Laroy, C. (1995). *Pronunciation.* Oxford University Press.

Le, H. T., & Brook, J. (2011). Using Praat to teach intonation to ESL students. *Hawaii Pacific University TESOL Working Paper Series, 9*, 2-15.

Lee, G. G., Lee, H., Song, J., Kim, B., Kang, S., Lee, J., & Hwang, H. (2017). Automatic sentence stress feedback for non–native English learners. *Computer Speech & Language, 41*, 29–42.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta–analysis. *Applied Linguistics, 36*, 345–366.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. Wiley.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. *Human Factors in Computing Systems*, 359–366.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3)*,* 369–377.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics 27*, 184–202.

Levis, J., & Grant, L. (2003). Integrating pronunciation in to ESL/EFL classrooms. *TESOL Journal, 12*, 13–19.

Levis, J., & Muller Levis, G. (2018). Teaching high–value pronunciation features: Contrastive stress for intermediate learners. *The CATESOL Journal, 30*(1), 139–160.

Levis J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*, 505–524.

Levis, J., & Sonsaat, S. (2017). Pronunciation in the CLT era. In O. Kang, R. Thomson, & J. Murphy (Eds.), *The Routledge handbook of English pronunciation* (267–283). Routledge.

Levy, M. (2009). Technologies in use for second language learning. *Modern Language Journal, 93*, 769–782.

Lewis, C., & Deterding, D. (2018). Word stress and pronunciation teaching in English as a lingua franca contexts. *The CATESOL Journal, 30*(1), 161–176.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta–analysis. *Language Learning, 60*, 309–365.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25.

Liberman, M., & Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry 8*, 249- 336.

Lightbown, P., & Spada, N. (2006). *How Languages are Learned.* Oxford University Press.

Littlewood, W. (1981). *Communicative Language Teaching – An introduction.* Cambridge University Press.

Liu, Y. (2008). *The effectiveness of integrating commercial pronunciation software into an ESL pronunciation class.* Unpublished M.A. dissertation. Iowa State University.

Liu, M., Evans, M., Horwitz, E., Lee, S., McRory, M., Park, J., & Parrish, C. (2013). A study of the use of social networking sites for language learning by university ESL students. In M.–N. Lamy, & K. Zorou (Eds.), *Social networking for language education* (15–41). Palgrave Macmillan.

Llisterri, J., Machuca, M., de la Mota, C., Riera, M., & Ríos, A. (2003). The perception of lexical stress in Spanish. In M.–J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (2023–2026). Barcelona, Spain.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874–886.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition 12*, 251–285.

Long, M. (2005). Problems with supposed counter–evidence to the Critical Period Hypothesis. *IRAL 43*(4), 287–317.

Lord, G. (2008). Podcasting Communities and Second Language Pronunciation. *Foreign Language Annals, 41*(2), 374–389.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*, 488–503.

Lu, J., Wang, R., & De Silva, L. C. (2012). Automatic stress exaggeration by prosody modification to assist language learners perceive sentence stress. *International Journal of Speech Technology 15*(2), 87–98.

Mackay, A., & Gass, S. (2005). *Second Language Research: Methodology and Design*. Lawrence Erlbaum Associates.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning 44*(1), 75–100.

MacDonald, S. (2002). Pronunciation – views and practices of reluctant teachers. *Prospect, 17*(3), 3–18.

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language teaching: A practical guide to creating a low–anxiety classroom atmosphere* (24–45). McGraw–Hill.

McBride, K. (2009). Social networking sites in foreign language classes: Opportunities for re–creation. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (35–58). San Marcos, TX: CALICO.

McCrocklin, S. (2012). The role of word stress in English as a lingua franca. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 249–256). Ames, IA: Iowa State University.

Madge, C., Meek, J., Wellens, J., & Hooley, T. (2009). Facebook, social integration and informal learning at university: 'It is more for socialising and talking to friends about work than for actually doing work', *Learning, Media and Technology, 34(2)*, 2 141–155. doi:10.1080/17439880902923606.

Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A critical review of the literature on Facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning, 29*(6), 487–504.

Marks, J. (2007). English pronunciation in use (elementary). Cambridge University Press.

Martinelli, O. (2016). Effectiveness of online language learning software (Duolingo) on Italian pronunciation features: a case study. Unpublished MA thesis. Oklahoma State University.

Massaro, D.W. (2003). A computer–animated tutor for spoken and written language learning. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (172–175). Vancouver, British Columbia, Canada.

Massaro, D. W., & Light, J. (2003). Read my tongue movements: bimodal learning to perceive and produce non–native speech /r/ and /l/. In: Proc. Eurospeech (2249–2252). Geneva, Switzerland.

Massaro, D. W., Bigler, S., Chen, T., Perlman, M., & Ouni, S. (2008). Pronunciation training: the role of eye and ear. In: *Proceedings of Interspeech (2623–2626).* Brisbane, Australia.

Mazer, J. P., Murphy, R. E., & Simonds, C. J. (2007). I'll see you on Facebook – The Effects of Computer–Mediated Teacher Self–Disclosure on Student Motivation, Affective Learning, and Classroom climate. *Communication Education. 56*(1), 1–17.

Menzel, W., Herron, D., Morton, R., Pezzotta, D., Bonaventura, P., & Howarth, P. (2001). Interactive pronunciation training. *ReCALL, 13*(1), 67–78.

Messum, P. (2007). *The role of imitation in learning to pronounce*. Unpublished Ph.D. dissertation. University College London.

Messum, P., & Howard, I. S. (2015). Howard Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics, 53*, 125–140.

Michael, L. (2010). The interaction of tone and stress in the prosodic system of Iquito (Zaparoan). In *UC Berkeley Phonology Lab annual report*, (57–79). Department of Linguistics, UC Berkeley.

Miksatko, J., Kipp, K., and Kipp, M. (2012). The persona zero–effect: Evaluating virtual character benefits on a learning task. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Proceedings of the 10th International Conference on Intelligent Virtual Agents* (475–481). Springer.

Mitchell, K. (2012). A social tool: Why and how ESOL students use Facebook. *CALICO Journal, 29*(3), 471–493.

Mompeán, J. A. (2010). A corpus–based study of phonological free variation in English. In A. Henderson (Ed.), *English Pronunciation: Issues and Practices (EPIP): Proceedings of the First International Conference* (13–36). Chambéry: Éditions de l'Université de Savoie.

Mompeán, J. A. (2014). Stress shift in English: The case of teen numbers. In R. Monroy & I. Arboleda–Guirao (Eds.), *Readings in English Phonetics and Phonology* (149–164). Valencia: Universitat de Valencia.

Mompeán, J. A., & Fouz–González, J. (2016). Twitter based EFL pronunciation instruction. *Language Learning and Technology 20*(1), 166–190.

Mompeán, J. A. & Fouz–González, J. (2020). Phonetic symbols in contemporary pronunciation instruction. *RELC Journal, 3*, 155–168.

Mompeán, J. A., & Lintunen, P. (2015). Phonetic notation in foreign language teaching and learning: Potential advantages and learners' views. *Research in Language,* 13(3), 292–314.

Mompeán, J. A., Ashby, M., & Fraser, H. (2011). Phonetics teaching and learning: an overview of recent trends and directions. *Proceedings from the 17th International Congress on Phonetic Sciences* (96–99). Hong Kong.

Mora, C., & Darcy, I. (2017). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs, & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (95–120). Multilingual Matters.

Mork, C. (2009). Using Twitter in EFL education. *JALT CALL Journal, 5*(3), 41–56.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481–520.

Morley, J. (1994). A multidimensional curriculum design for speech–pronunciation instruction. In J. Morley (Ed.), *Pronunciation, pedagogy and the theory: New views, new directions* (64–91). TESOL.

Morton and Jassem 1965

Motohashi–Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology, 13*(2), 29–47.

Mott, B. (1996). *A course in phonetics and phonology for Spanish learners of English.* EUB.

Munoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics 24*, 578–596.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: ratings of "altered speech. *Studies in Second Language Acquisition, 17*, 17–33.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards, & M. Zampini (Eds.), *Phonology and second language acquisition* (193–218). John Benjamins.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System 34*, 520–531.

Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching 44*(3), 316–27.

Munro, M. J., Derwing, T. M., & Burgess, C. S. (2003). The detection of foreign accent in backwards speech. In M–J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (535–538). Barcelona, Spain: Universitat Autonoma de Barcelona.

Munro, M., & Mann, V. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics, 26*, 311–341.

Murphy, J. & Baker, A. A. (2015). History of ESL pronunciation teaching. In M. Reed, & J. Levis (Eds.), The handbook of English pronunciation (36–65). Chichester, West Sussex, UK: John Wiley and Sons.

Murray, L., & Barnes, A. (1998). Beyond the 'wow' factor – Evaluating multimedia language learning software from a pedagogical viewpoint. *System, 26*, 249–259.

Neri, A., Cucchiarini, C., & Strik, H. (2002). Feedback in computer assisted pronunciation training: Technology push or demand pull? *Proceedings of International Conference on Spoken Language Processing* (1209–1212). Denver, Colorado.

Neri, A., Cucchiarini, C., & Strik, H. (2003). Automatic speech recognition for second language learning: how and why it actually works. Proceedings of 15th International Congress of Phonetic Sciences (1157–1160). Barcelona, Spain.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy–technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441–467.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning, 21*, 393–408.

Nielson, K. B. (2011). Self–study with language learning software in the workplace: What happens? *Language Learning & Technology 15*(3), 110–129.

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini–review. *Frontiers in Psychology 28*, 1–5. doi.org/10.3389/fpsyg.2012.00322.

Nishi, K., & Kewley–Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech Language and Hearing Research, 50*(6), 1496–1509.

Nolan, F., Jeon, H–S. (2014). Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society*. doi.org/10.1098/rstb.2013.0396.

O'Bryan, A., & Hegelheimer, V. (2007). Integrating CALL into the classroom: the role of podcasting in an ESL listening strategies course. ReCALL, 19(2), 162–180.

O'Reilly, T. (2005). *What is Web 2.0: Design Patterns and Business Models for the next generation of software*. Available at: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what–is–web–20.html [last accessed on 16/8/19].

Ortega–Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context–driven L2 stress perception. *Journal of Phonetics, 41*(3–4), 186–197.

Ortega–Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, 54(1), 73–97.

Ortiz–Lira, H. (1995). Nucleus placement in English and Spanish: a pilot study. In J. Windsor Lewis (Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor* (pp. 255–265). Routledge.

Ortiz–Lira, H. (1998). *Word stress and sentence accent*. Santiago: Universidad Metropolitana de Ciencias de la Educación.

Ota, F. (2011). A study of social networking sites for learners of Japanese. *New Voices, 4*, 144–167.

Ou, S. C. (2011). Training Taiwanese EFL learners to perceive English lexical stress contrast: A pilot study. In *Proceedings of the 17th International Congress of Phonetic Sciences* (150–153). Hong Kong, China.

Penfield W., & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton University Press.

Pennington, M. C. (1998). The teachability of phonology in adulthood: A re–examination. IRAL, *International Review of Applied Linguistics in Language Teaching, 36*(4), 323–341.

Pennington, M. C. (1999) Computer–Aided Pronunciation Pedagogy: Promise, Limitations, Directions, *Computer Assisted Language Learning, 12*(5), 427–440.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*, 372–389.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL*

Quarterly, 20, 207–225.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 203–240) Mouton de Gruyter.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38*(3), 422–430.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between differences and training paradigm design. *Journal of the Acoustical Society of America, 130*(1), 461–472.

Pierce, L. (2014). *Learning novel vowel contrasts: experimental methods in classroom applications*. Unpublished Ph.D. dissertation. University of Illinois at Urbana–Champaign.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics 29*, 191–215.

Plag, I. (2006). The variability of compound stress in English: Structural, semantic, and analogical factors. *English Language and Linguistics, 10*(1), 143–172.

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure and lexicalization in compound stress assignment in English. *Language 84*, 760–794.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition, 35*(4), 655–687.

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effects sizes in L2 research. *Applied Linguistics, 36*(3), 345–366.

Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to

ANOVA in L2 research. *Studies in Second Language Acquisition, 39,* 579–592.

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, *9*(5), 1–6.

Promnitz–Hayashi, L. (2011). A learning success story using Facebook. *Studies in Self–Access Learning Journal, 2*(4), 309–316.

Precoda, K., Halverson, C., & Franco, H. (2000). Effects of Speech Recognition–based Pronunciation Feedback of Second–Language Pronunciation Ability. In *Proceedings of InStil* 2000 (102–105). Dundee, Scotland.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: a reexamination. *Language Learning 30*(2), 271–287.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* Longman.

Ramirez–Verdugo, D. (2006). A study of intonation awareness and learning in non–native speakers of English. *Language Awareness. 15*(3), 141–159.

Rato, A. (2014). Effects of Perceptual Training on the Identification of English Vowels by Native Speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics, 5*, 529–546.

Rauber, A., Rato, A., Kluge, D., & Santos, G. (2011). TP–S (Version 3.1). [Software]. Brazil: Worken.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and Methods in Language Teaching*. Cambridge University Press.

Roach, P. (2009). *English phonetics and phonology*. Cambridge University Press

Romanelli, S., Menegotto, A. C., & Smyth, R. (2015). Stress perception: Effects of training and a study abroad program for L1 English late learners of Spanish. *Journal of Second Language Pronunciation, 1*(2), 181–210.

Rosell–Aguilar, F. (2007). Top of the Pods – In search of a podcasting 'podagogy' for language learning. *Computer Assisted Language Learning, 20*(5), 471–492.

Rosell–Aguilar, F. (2009). Podcasting for language learning: Re–examining the potential. In L. Lomicka, & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (13–34). CALICO Monograph Series.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation in EFL settings: The case of English vowels and Japanese learners of English. *The Linguistics Journal 3*(3), 16–40.

Saito, K. (2012). Re–examining effects of form–focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*(1), 1–29.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37*, 813-840.

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain–general auditory processing as an anchor of post–pubertal L2 pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language, 115*, 1–15.

Saito, K., & Lyster, R. (2012). Effects of form–focused instruction and corrective feedback on L2 pronunciation development of / ɹ / by Japanese learners of English. *Language Learning, 62*(2), 595–633.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta–analysis. *Language Learning, 69*(3), 652–708.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*. *38*(4), 439–462. doi:10.1093/applin/amv047.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta–analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224.

Saran, M., Seferoglu G., & Cagiltay, K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research, 34*, 97–114.

Saylag, R. (2013). Facebook as a tool in fostering EFL teachers' establishment of interpersonal relations with students through self–disclosure. *Procedia–Social and Behavioral Sciences, 82*, 680–685.

Schmerling, S. (1976). *Aspects of English sentence stress*. University of Texas Press.

Schmider, E., Ziegler, M., Danay, E., Beyer, Y., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology: *European Journal of Research Methods for the Behavioral and Social Sciences, 6*(4), 147–151.

Schmidt, A. M. (2012). Effects of EPG treatment for English consonant contrasts on L2 perception and production. *Clinical Linguistics & Phonetics, 26*, 909–925.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schwartz, A. H. (1996). *Technology enhanced accent modification (TEAM) for international teaching assistants (ITAs)*. Cleveland, OH: Department of Speech and Hearing, Cleveland State University.
Retrieved from: http://eric.ed.gov/?id=ED414836.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19*, 245–253.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Newbury House.

Seferoglu, G. (2005). Improving students' pronunciation through accent reduction software, *British Journal of Educational Technology, 36*(2), 303–316.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–388). Blackwell.

Seilstad, B. (2012). Using tailor–made YouTube videos as a preteaching strategy for English language learners in Morocco: Towards a hybrid language learning course. *Teaching English with Technology, 12*(4), 31–47.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209– 231.

Selwyn, N. (2009). Faceworking: Exploring Students' Education–Related Use of Facebook. *Learning, Media and Technology, 34*(2), 157–174.

Setter, J. (2008). Theories and approaches in English Pronunciation. In R. Monroy, & A. Sánchez (Eds.), 25 Years of Applied Linguistics in Spain: Milestones and Challenges (447–457). Murcia, Spain: Universidad de Murcia, Servicio de Publicaciones.

Setter, J., & Jenkins, J. (2005). Pronunciation. *Language Teaching, 38*, 1–17.

Sharwood Smith, M. (1981). Consciousness–raising and the second language learner. *Applied Linguistics 2*, 159–68.

Sharwood Smith, M. (1993). Input enhancement in instruction SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165–179.

Sharwood Smith, M., & Truscott, J. (2014). Explaining input enhancement: a MOGUL perspective. *International Review of Applied Linguistics 52*(3), 253–281.

Shih, R. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology, 27*(5), 829–845.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics, 43*, 269–286.

Smith, J. (2011). The YouTube revolution: Engagement, perception, identity. *Presentation at the 17th Congress of Phonetic Sciences*, Hong Kong, China.

Smith, J. L. (1893). The Gouin Method, *American Annals of the Deaf,* 38(3), 177–185.

Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal, 10*(3), 19–30.

Spaai, G. W., Hermes, D. J., Derksen, E. S., & Kaufholz, P. A. (1994). The Effectiveness of the Intonation Meter for Teaching Intonation to Deaf Persons. *Lecture Notes for Computer Science, 860*, (355–62). Springer.

Stenson, N., Downing, B., Smith, J., & Smith, K. (1992). The Effectiveness of Computer–Assisted Pronunciation Training. *CALICO Journal, 9*(4), 5–19.

Stevenson, M. P., & Liu, M. (2010). Learning a language with Web 2.0: Exploring the use of social networking features of foreign language learning websites. *CALICO Journal, 27*(2), 233–259.

Stockwell, G., & Hubbard, P. (2013). Some emerging principles for mobile–assisted language learning. *The International Research Foundation for English Language Education* (1–15). Monterey, CA.

Stockwell, G. (2007). A review of technology choice for teaching language skills in the CALL literature. *ReCALL, 19*(2), 105–120.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception and Psychophysics, 36*, 131–145.

Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution? *Chest, 140*, 16–18.

Suter, R. (1976). Predictors of pronunciation accuracy in second–language learning. *Language Leaning 26*, 233–53.

Swan, M., & Smith, B. (2001). *Learner Language: A teacher's guide to interference and other problems*. Cambridge university Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & M. Madden (Eds.), *Input in Second Language Acquisition* (235–253). Newbury House.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371–391.

Sweet, H. (1906). *A Primer of Phonetics*. Clarendon Press.

Sweet, H. (1899). *The practical study of languages.* Oxford University Press.

Taniguchi, M., & Abberton, E. (1999). Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language: Work in Progress, 11,* 76–89.

Tanner, M. W., & Landon, M. M. (2009). The effects of computer–assisted pronunciation on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology, 13*(3), 51–65.

Thomson, R. I. (2011). Computer assisted pronunciation Training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal, 28*(3), 744–65.

Thomson, R. I. (2012). Demystifying pronunciation research to inform practice. In H. M. McGarrell, & R. Courchêne (Eds.), *Special Research Symposium Issue of CONTACT*, *38*(2), 63–75.

Thomson, R. I. (2016). Does training to perceive L2 English vowels in one phonetic context transfer to other phonetic contexts? *Canadian Acoustics, 44*(3), 198–199.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36,* 326–344.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (88–97). Ames: Iowa State University.

Thorne, S. L., & Payne, J. S. (2005). Evolutionary Trajectories, Internet–mediated Expression, and Language Education. *CALICO Journal, 22*(3), 371– 397.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21*(3), 217–228.

Townshend, B., Bernstein, J., Todic, O., & Warren, E. (1998). Estimation of spoken language proficiency. *Proceedings of the ESCA Workshop Speech Technology in Language Learning* (179–182). Stockholm.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in L2 Spanish word processing: Implications for pronunciation instruction. *Modern Language Journal 90*, 519–535.

Trubetzkoy, N. (1971). *Principles of phonology*. University of California Press.

Trudgill, P. (2001). *Sociolinguistic variation and change*. Edinburgh University Press.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing 16*, 255–272.

Underhill, A. (2010). Four reasons to teach pronunciation. Available at:

https://www.adrianunderhill.com/2010/11/23/four–reasons–to–teach–pronunciation/.

van Mulken, S., André, E. & Müller, J. (1998). The Persona Effect: How Substantial is it? In *Proceedings of Human–Computer Interactions* (53–66). Sheffield.

van Patten, B. (1994). Evaluating the role of consciousness in SLA Terms, linguistic features, and research methodology. *AILA Review 11*, 27–36.

Vardanian, R. (1964). Teaching English intonation through oscilloscope displays. *Language Learning, 14*, 109–118.

Vitevitch, M. S., Luce, P. A., Charles–Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47–62.

Wachowicz, K., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal 16*, 253–276.

Walker, N., Trofimovich, P., Cedergren, H., & Gatbonton, E. (2011). Using ASR technology in language training for specific purposes: A perspective from Quebec, Canada. *CALICO Journal, 28*(3), 721–743.

Walker, R. (2005). Using student–produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly, 39*, 550–557.

Walker, R. (2010). Teaching the pronunciation of English as a lingua franca. Oxford University Press.

Wang, J., & Munro, M. J. (2004). Computer–based training for learning English vowel contrasts. *System, 32*, 539–552.

Wang, S., & Vásquez, C. (2012). Web 2.0 and second language learning: What does the research tell us? *CALICO Journal, 29*(3), 412–430.

Wang, Q., Woo, H. L., Quek, C. L., Yang, Y., & Liu, M. (2012). Using the Facebook group as a learning management system: An exploratory study. British *Journal of Educational Technology, 43*(3), 428–438.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113*(2), 1033–1043.

Wang, Y., Spence, M. M., Jongman, A. & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649–3658.

Warschauer, M. & Liaw, M. (2011). Emerging technologies for autonomous language learning. *Studies in Self–Access Learning Journal*, *2*(3), 107–118.

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL Classroom. *Research in Pedagogy, 7*(1), 21-29.

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge University Press.

Weltens, B., & de Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech, 27*, 79–88.

Werker, J. (1989). Becoming a native listener. *American Scientist 77*, 54–59.

Wichern, P. U. M., & Boves, L. (1980). Visual feedback of F0 curves as an aid in learning intonation–contours. *Proceedings, Institute of Phonetics 4*, (53–63). Nijmegen.

Wik, P., & Engwall, O. (2008). Can visualization of internal articulators support speech perception. In *Proceedings of Interspeech 2008* (2627–2630). Brisbane, Queensland, Australia.

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication, 51*(10), 1024–1037.

Wilson, I. (2008). Using Praat and Moodle for teaching segmental and suprasegmental pronunciation. In *Proceedings of the 3rd international*

WorldCALL Conference: Using Technologies for Language Learning (112–115). Fukuoka, Japan.

Witt, S., & Young, S. J. (1997). Language learning based on non–native speech recognition. In *5th European Conference on Speech Communication and Technology* (22–25). Rhodes, Greece.

Wong, J. W. S. (2012). Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology.* Sydney, Australia.

Wong, J. W. S. (2014). The Effects of High and Low Variability Phonetic Training on the Perception and Production of English Vowels /e/–/æ/ by Cantonese ESL Learners with High and Low L2 Proficiency Levels. In H. Li, & P. Ching (Eds.), *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (524–528). Singapore.

Wong, J. W. S. (2015). The impact of L2 proficiency in vowel training. In J. A. Mompeán, & J. Fouz-González (Eds.), *Investigating English pronunciation: Current Trends and Directions* (219–239). Palgrave Macmillan.

Xodabande, I. (2017). The effectiveness of social media network telegram in teaching English language pronunciation to Iranian EFL learners. *Cogent Education, 4*(1).
doi:10.1080/2331186X.2017.1347081.

Xu, Y., (2019). Prosody, Tone and Intonation. In W. F. Katz, & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (314–356). Routledge.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics, 52*, 376–392.

Yavas, M. (2011). *Applied English phonology*. Wiley–Blackwell.

Yoshida, H. (2004). *An analytic instrument for assessing EFL pronunciation.* Unpublished Ph.D. dissertation. Philadelphia, PA: Temple University.

Yunus, M. M., & Salehi, H. (2012). The effectiveness of Facebook groups on Teaching and Improving Writing: Students' perceptions. *Journal of Education and Information Technologies, 1*(6), 87–96.

Zielinski, B. (2008). The listener: no longer the silent partner in reduced intelligibility. *System 36*(1), 69–84.

Zielinski, B. (2015). The segmental/suprasegmental debate. In: M. Reed & J. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 397–412). John Wiley & Sons.

Zimmerman, D. W. (1998). Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions. *The Journal of Experimental Education, 67*(1), 55–68.