International Journal of English Studies

IJES

UNIVERSITY OF MURCIA

www.um.es/engphil/ijes

Obstruent Voicing in English and Polish. A Pedagogical Perspective

WIKTOR GONET*
UMCS Lublin / WSU Kielce

ABSTRACT

'Voicing' in English voiced obstruents has been defined in terms of 'full' vs. 'partial'. When teaching English pronunciation to native speakers of Polish, where voiced sounds can be only fully voiced, it is diffcult to make the students aware of the phonation strategy to be used to obtain 'partially voiced' sounds, especially in plosives. The accessability of digital speech analysis computer software has made it possible to visualize the acoustic properties of speech sounds which can facilitate the teaching of English pronunciation to Poles, providing a visual feedback in class and at home. This is necessary for obtaining the correct phonation control that functions with utmost precision measured in centiseconds.

Yet speech visualisation for the purpose of teaching English phonetics in Poland is employed only at the author's institution, and the remaining hundreds of schools and universities do not take advantage of the possibilities modem technology offers. The 'pedagogical perspective' of the paper aims at exerting an encouraging impact both on teachers of phonetics and on students of English. The article also provides a description of Polish voicing rules and a detailed comparison of voicing in English and Polish obstruents based on the concept of Voice Onset Time.

KEYWORDS: acoustic phonetics, spectrographic analysis, obstruent voicing, VOT phonetic interference, teaching English pronunciation, speech timing, phonology.

Address forcorrespondence: **Viktor Gonet, Zaklad Fonetyki i Fonologii, Instytut Anglistyki, Uniwersytet Marii Curie-Sklodowskiej, Plac M. C. Sklodowskiej 4,** 20-031 **Lublin, Poland. E-rnail:** gonet@klio.umcs.lublin.pl

I. INTRODUCTION

Appropriate rendering of voicing belongs to relatively persistent pronunciation difficulties encountered by Poles when they leam English or German, where voicing control is governed by totally different implementation rules than those that are used in Polish (Gonet 1981:309, 315-317). Especially difficult to conceive and implement is the type of voicing often referred to as "partial", applied equally to English fricatives, plosives and affricates. When students rely on this term, they imagine "partial (de)voicing" as a segmental feature that characterizes sounds throughout their articulation'. Such an approach makes the acquisition of foreign voicing strategies very difficult if not totally impossible.

In modern perspective, however, voicing is associated with the timing of the vocal fold vibration relative to consonant constriction —narrowing for fricatives, and complete closure for stops (i.e. plosives and affricates). Possible phonation types used in English and Polish are such in which vocal fold vibration can (a) start simultaneously with the formation of the constriction and persist during its whole duration (Polish and English voiced obstruents between voiced sounds), (b) it can be delayed relative to the formation of the constriction (Polish word initial voiced obstruents), (c) it can cease prior to the release of the constriction (English word final "partially" voiced obstruents), (d) it can be simultaneous with the release of the constriction (English word initial partially voiced stops), (e) it can be little delayed relative to the release of the constriction (English and Polish voiceless stops), and **(f)** it can be further delayed relative to the release of the constriction (English aspirated voiceless plosives), schematically:

Occlusion	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx.....	xxxxxxxx.....	xxxxxxxxx.....
phonation	xxxxxxx	...xxxxx	xxxxx....	xxxxx	xxxxx	xxxxx
	(a)	(b)	(c)	(d)	(e)	(f)

***Figure* 1: Schematic presentation of the timing relations of voicing to occlusion**

Cases (b) and (d) through (f) are commonly referred to as Voice Onset Time, that is, the time interval that elapses between the release of closure and the initiation of vocal fold vibration[2]: negative for (b), also called prevoicing; simultaneous, or 'O', for (d), short positive for (e) and long positive for (f); case (c) is known as 'voicing into closure' (VIC). In literature, these terms are used in reference to stops (plosives and affricates); here we shall extend their application to fricatives, treating the term 'closure' as equivalent to 'constriction' (after all, closure is the extreme degree of constriction).

Correct rendering of these intricate timing relations is very hard to achieve, as it does not involve any specific shifts in the gross configuration of articulators, but rather synchronization of the two components shown in the diagram that requires accuracy of less than 30 ms. Our experience has shown that it is possible to facilitate this process of acquisition of foreign language articulation by the use of visual representation of articulation obtained through

computer based devices that present on the computer screen spectrograms and oscillograms in which acoustic correlates of individual articulation gestures can be found. The author's preliminary experiments carried out with Polish adult students of English (cf. Gonet et al 2001) have produced encouraging results which will add to the scarce literature on this subject (e.g. Chun 1988 on the use of visualisation in the teaching of intonation).

II. IMAGING: OVERVIEW

Presence of phonation is easily seen both in the spectrogram, in the form of a voicebar situated at the bottom of the spectrogram, and in the oscillographic waveform, as regular quasiperiodic vibrations (Figure 2):

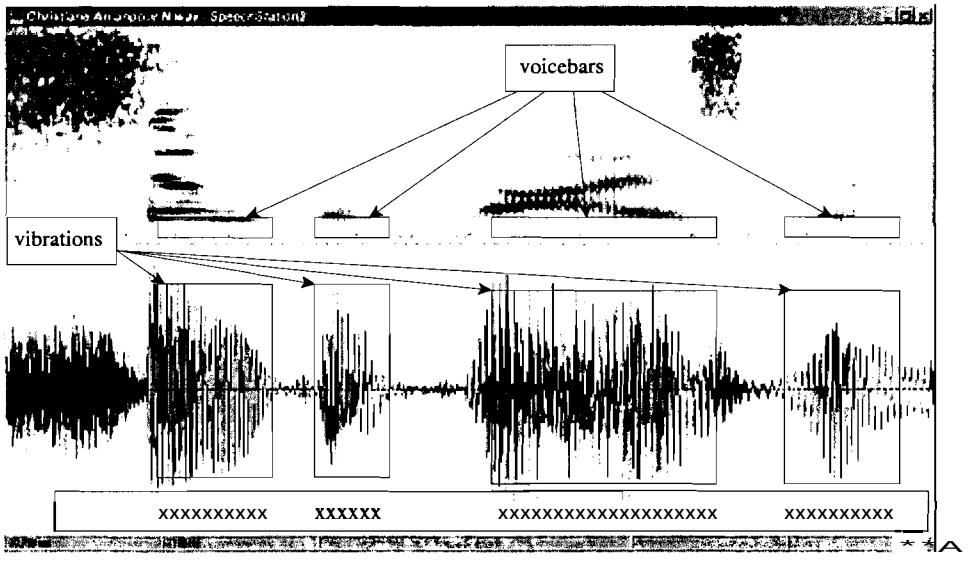


Figure 2: **Phonation marked 'xxx', as seen in a waveform and a spectrogram; clues indicated by rectangles. Utterance 'sympathize with'**

Because the analysis of acoustic images of phonation in fricatives presents a less complicated picture, it will be dealt with first; a discussion of stops will follow.

III. FRICATIVES

Polish word initial fricatives and affricates are either voiced or voiceless. Nowocień (2000:39) shows that in Polish about 20% of the duration of the phonation interval precedes the formation

of the constriction of fricatives, and 70%, in affricates.

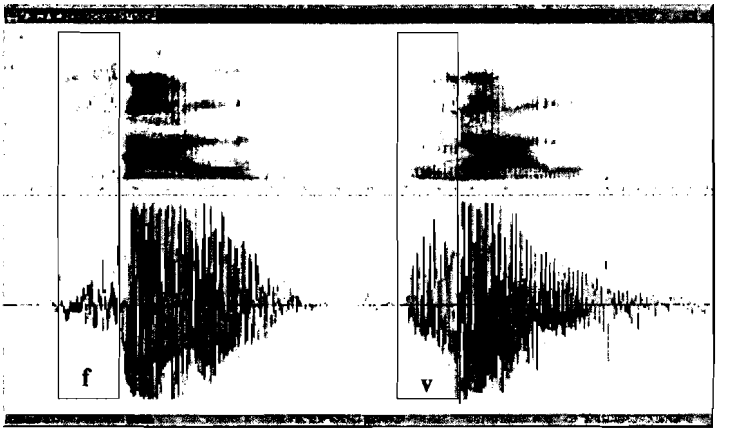


Figure 3: Polish [fal] ('waves', gen. pl.), [wal] ('hit' voc.). Polish is transcribed using the SAMPA ASCII phonetic transcription; cf. http:llwww.phon.ucl.ac.uWhomelsampalpolish.htm

In English, word initial fricatives are 'partially (de)voiced' which, in terms of our typology, represents Case (b) in Fig. 1, i.e. the 'negative VOT' (prevoicing) or, in our terminology, 'voicing from closure'. In Nowocien (2000:40), about 40% of the fricative constriction was voiceless, and 60%, voiced, while in English affricates, voiceless and voiced intervals were of equal duration. Consider the following example coming from the author's database:

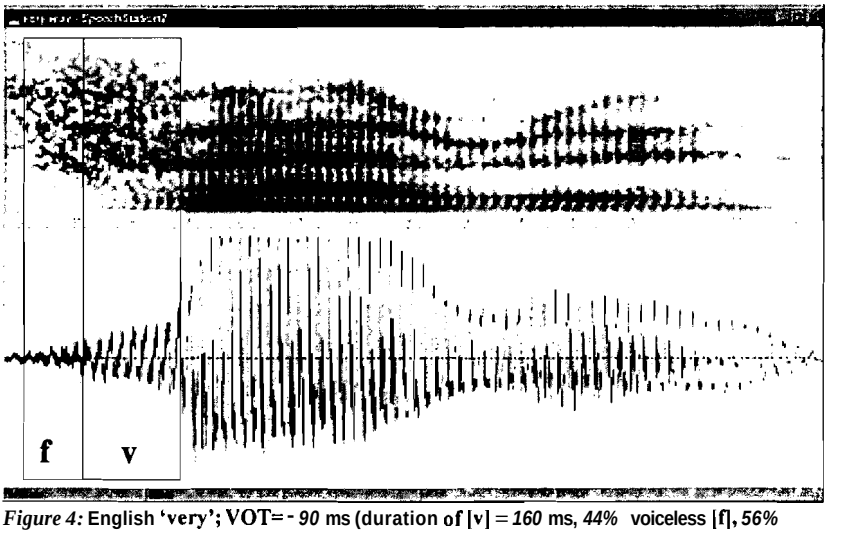


Figure 4: English 'very'; VOT= - *90* ms (duration of |v| = *160* ms, *44%* voiceless |f|, *56%* voiced |v|

In word final position Polish admits only voiceless fricatives, irrespective of the morphophonological status of the consonant:

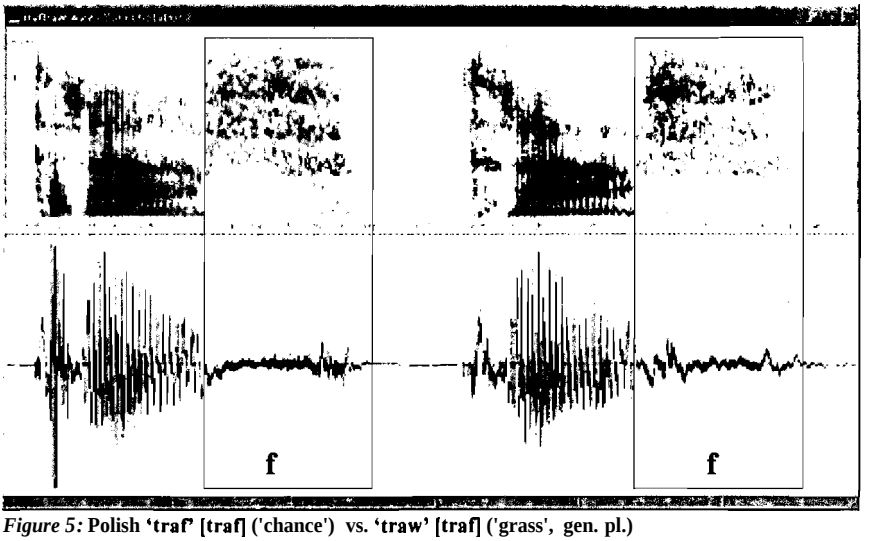


Figure 5: **Polish 'traf' [traf] ('chance') vs. 'traw' [traf] ('grass', gen. pl.)**

In English, in an analogical environment, articulation requires the use of Case (c) from Fig. 1, i.e. 'voicing into closure'; this, however, results in the formation of a transition segment, in which noise is superimposed on the quasi-periodic vibration[3]; cf. Figures ***6*** and 7:

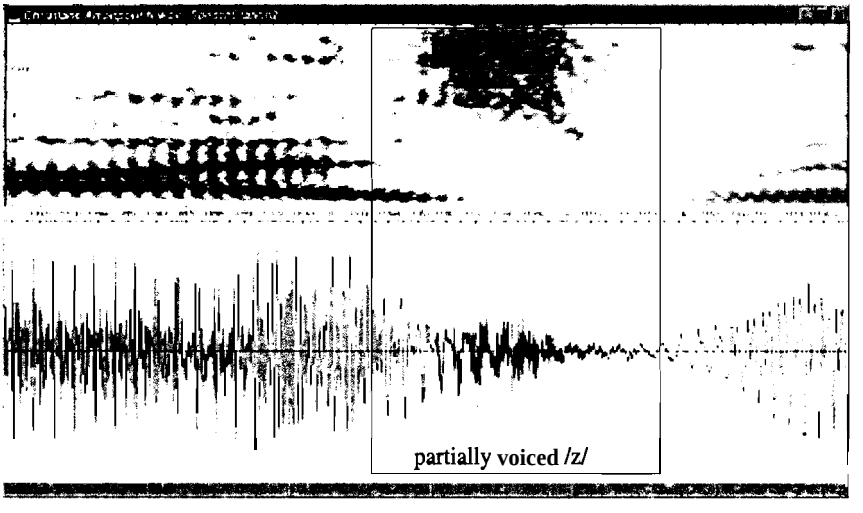


Figure 6: **English [aiz wi] – part of 'sympathize with'; cf. also Fig. 7**

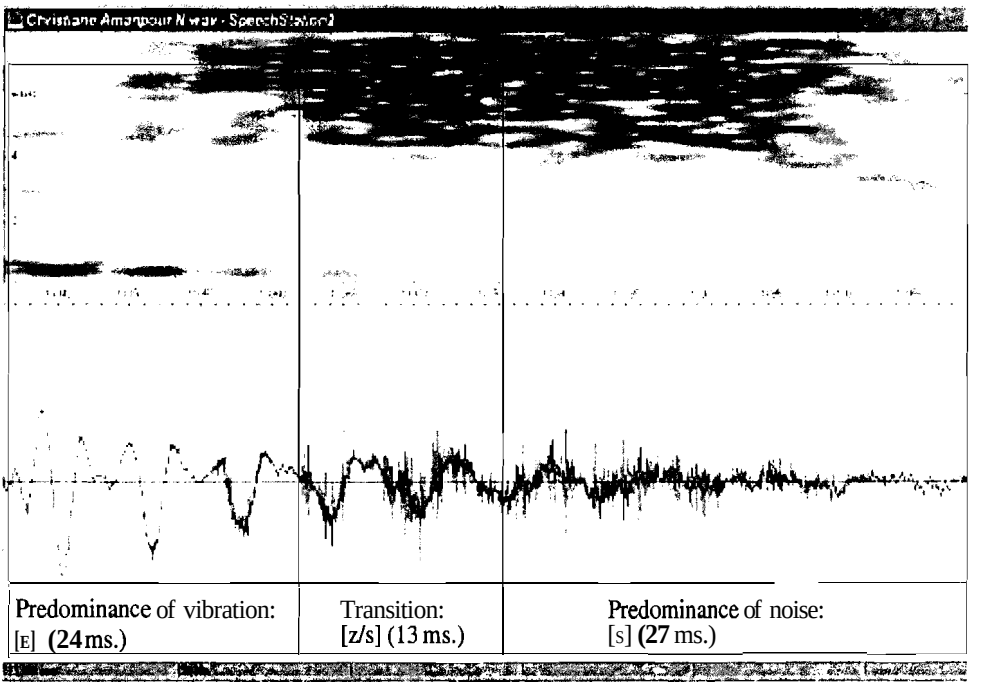


Figure 7: **Magnification of partially devoiced [z]: voiced part 38%, transition 20%, and voiceless 42% of duration of [z]**

On the basis of data presented elsewhere (Gonet, in preparation), it was found out that the amount of voicing in word final fricatives significantly depends on their place of articulation: the longest voiced part of the fricative (80% of its duration) appears in the interdental fricative; shorter (60% of the duration) voicing is associated with the labio-dental fricative, and the shortest (50%), with the alveolar /s/. This regularity can be explained by referring to perceptual strength of these sounds: the less conspicuous the fricative is, the more care is taken by the speakers to distinguish the lenes from the fortes by prolonging the duration of the voiced interval. Thus, the mellow interdental fricative is usually voiced during almost all of its duration, the more conspicuous [v] is pronounced with a shorter voicing interval, while the most strident [z] is very often voiced in not more than 50% of its duration. This less careful rendering of voicing can also be the result of the interplay of morphology when the voicing value can be predicted in inflectional endings on the basis of the voicing of the segment it precedes.

In English, word medially, partially devoiced fricatives occur adjacent to voiceless sounds (Figure 8):

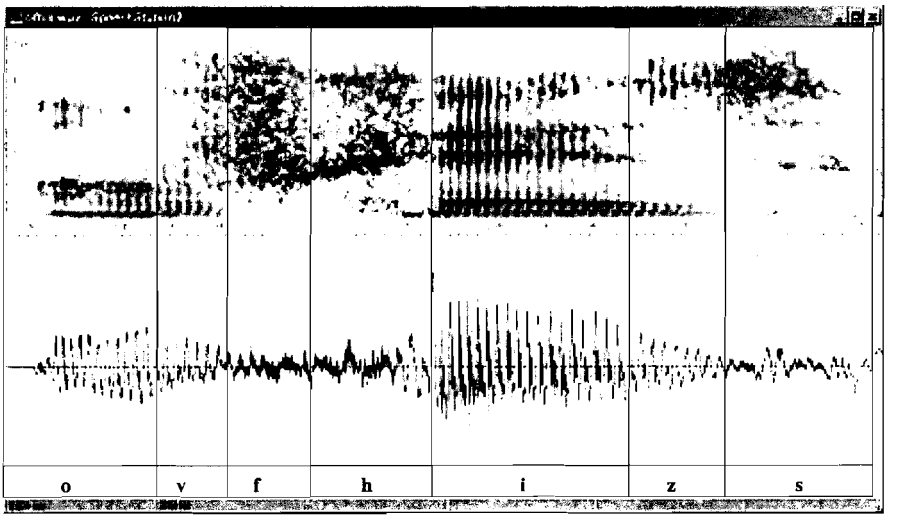


Figure 8: **English 'of his'; voicing of [v] is preserved partially (transcription shows successive segrnents)**

In Polish, the choice is limited to fully voiced vs. voiceless consonants, in that voiced fricatives occur in voiced environments. Since members of Polish consonantal clusters have to agree in voicing, and the adjustment always goes in the direction of devoicing, preceding or following voiceless obstruents devoice voiced fricatives (Figure 9):

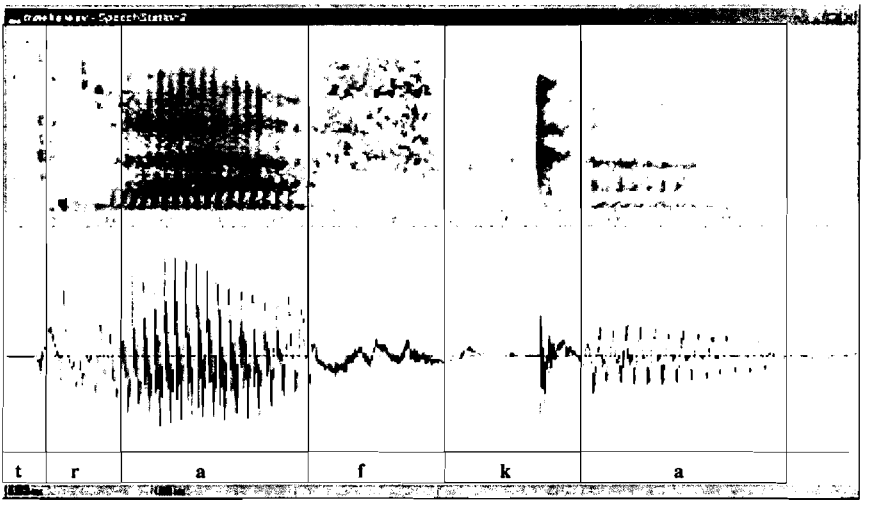


Figure 9: **Polish 'trawka' [trafka] ('grass', dirn.), with fully devoiced /v/**

To sum up this part of the discussion, let us state that for Polish learners of English the correct rendering of voicing in both individual voiceless and fully voiced fricatives is not problematic, as the same sound variants occur in their native language. Problems that appear in other contexts are of two kinds: (1) substitution of English partially devoiced fricatives with Polish voiceless fricatives in clusters with voiceless obstruents and word finally, and (2) correct rendering of "partial (de)voicing".

IV. STOPS

Regarding the realization of voicing in English plosives by Polish learners of English, the crucial contexts are: i) word initial, ii) word medial after a voiceless sound (including clusters with –s), iii) word medial before a voiceless sound, and iv) word final.

IV.1. Word initial position

Similarly to fricatives, word initial stops in Polish are either fully voiced or voiceless, while —as described in numerous textbooks — English requires here a contrast between partially (de)voiced and voiceless aspirated plosives. The control of English voicing by speakers of Polish requires a reshuffling of the timing relations measured with regard to the initiation of voicing and the release of the closure. More specifically, Polish voiced word initial plosives are usually produced with negative VOT (Case b in Fig. 1):

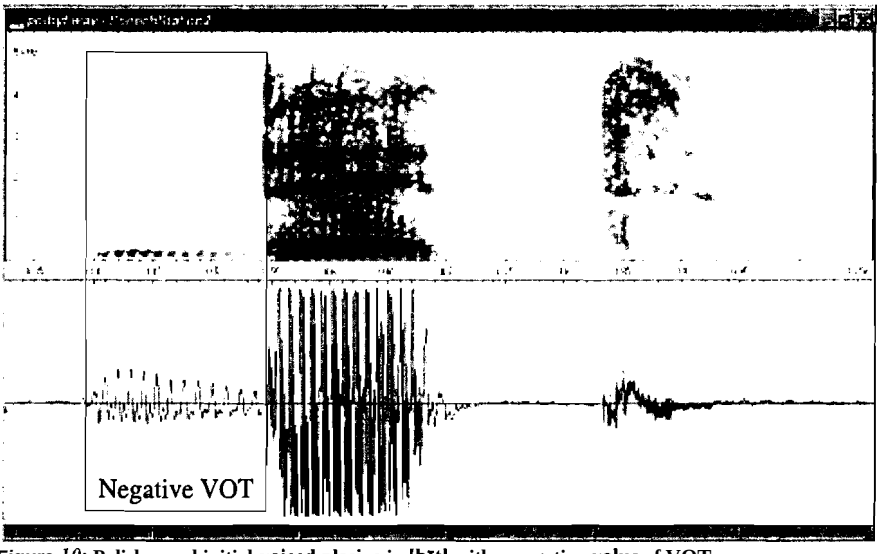


Figure 10: **Polish word initial voiced plosive in [bIt] with a negative value of VOT**

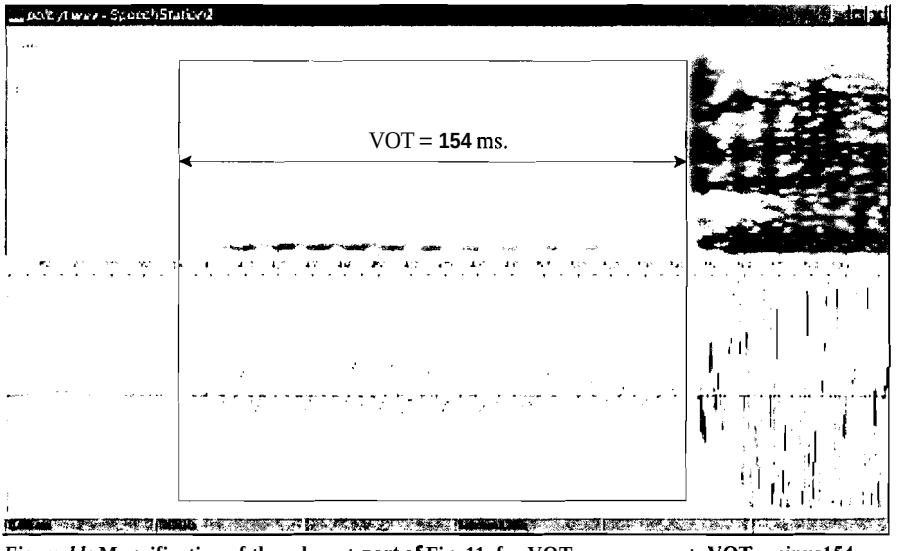


Figure 11: **Magnification of the relevant part of Fig. 11. for VOT measurement; VOT= minus154 ms.**

Nowocien (2000:41) shows that the mean duration of the pre-plosive and pre-affricate glottal pulsing (prevoicing) in Polish constitutes 70% of the duration of the whole voiced segment associated with the obstruent. Both Gonet (1989) and Nowocien (2000) found native speakers of English who use a similar voicing strategy for English.

Consider now a Polish voiceless plosive (Figures 12 and 13):

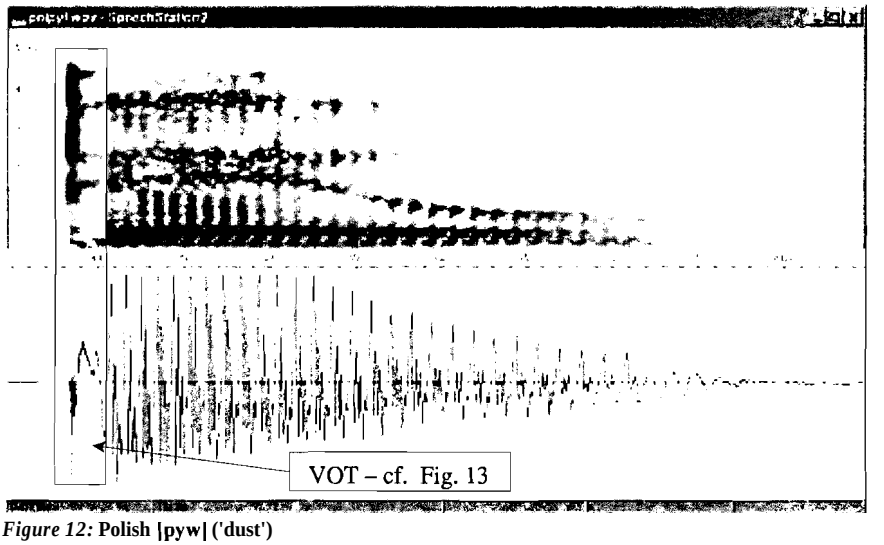


Figure 12: **Polish |pyw| ('dust')**

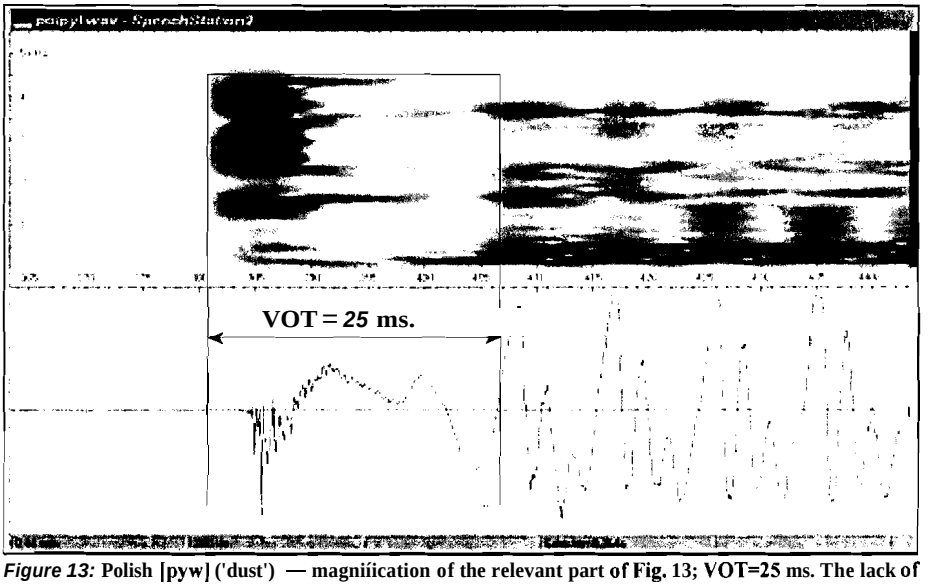


Figure 13: **Polish [pyw] ('dust') — magniiication of the relevant part of Fig. 13; VOT=25 ms. The lack of complete correspondence between the spectrogram and the oscillogram is an artefact due to large magniiication; both images are complementary.**

Typical "textbook" English voiced word initial plosives require VOT ranging around 0 (from shori negative through 0 to shori positive):

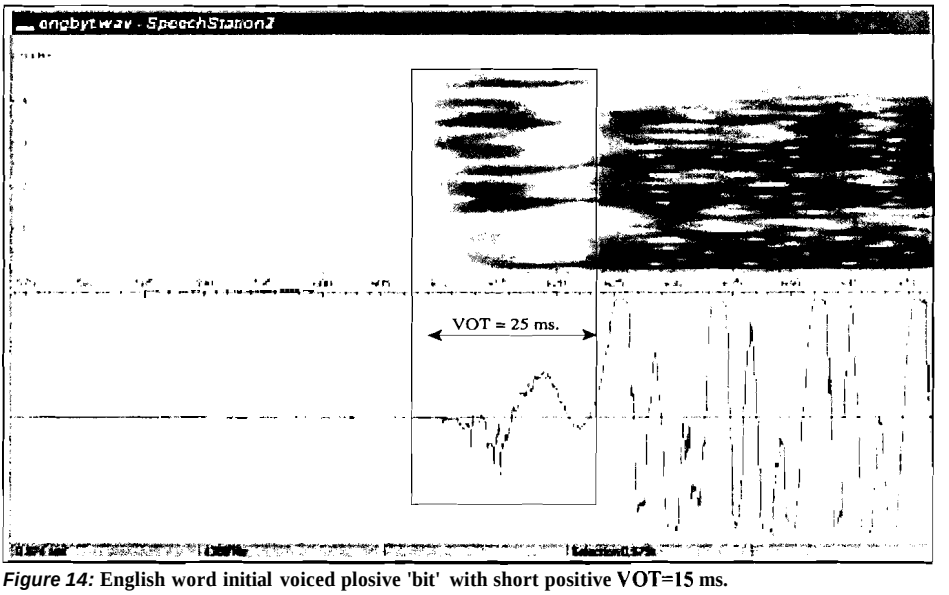


Figure 14: **English word initial voiced plosive 'bit' with short positive VOT=15 ms.**

Such timing of phonation with respect to the release of the closure produces a perceptual effect of "partial voicing", in which serially ordered phenomena are perceived as stable characteristics of sounds. The key to correct rendering of this type of pronunciation by a foreigner lies in the comprehension of the nature of the phenomena involved, and in training backed up by visual feedback provided by the use of speech analysis software; Gonet and Święciński (2001) presenta review of programs useful for such a purpose.

English syllable initial fortes that stand before a strongly stressed vowel require 'aspiration' —customarily defined as 'a puff of air'. This definition is harmful to the foreign learner of English as it wrongly leads him to practicing that 'puff of air', thereby making the effect much too strong. Consider a spectrographic and oscillographic image of English 'cow':

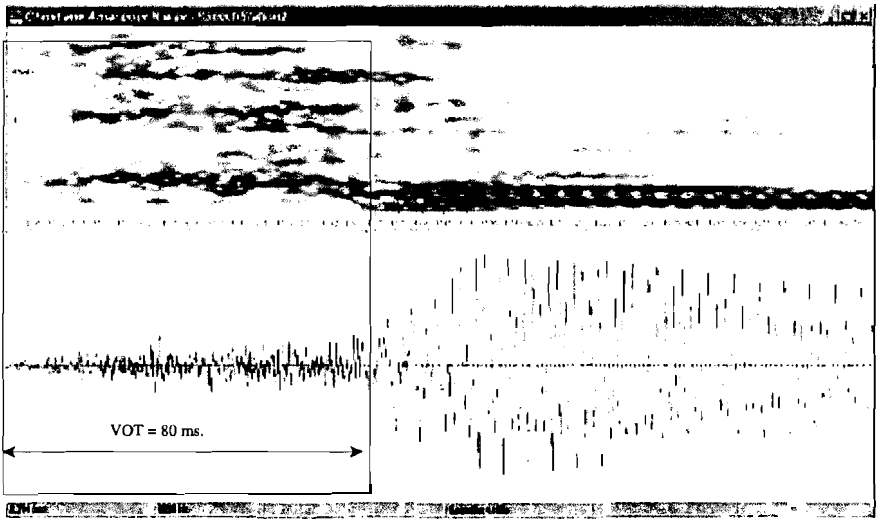


Figure 15: **English 'cow'; VOT=80 ms., interpreted as 'aspiration'**

The description of aspiration by means of the concept of VOT places it on a par with voicing, indicating that voicing and aspiration function on one axis as various degrees of a property that is used in making a perceptual distinction between the broad categories of 'voiced' and 'voiceless'. One can use VOT measurements to see how strongly 'voiced' sounds are differentiated from 'voiceless' in a given language. Thus, for Polish, a VOT value for a voiceless plosive was –154 ms. (cf. Fig. 11), and the one for a voiced plosive was 25 ms. (cf. fig. 13). Hence the perceptual distance pd defined on the VOT axis equals (minus 154)+25=179 ms. For English, the corresponding values are: 15 ms. (cf. Fig. 14) and 80 ms. (cf. Fig. 15); hence the pd value for English, calculated on these two examples, equals 80-25=55 ms.; cf. the following diagram[4]:

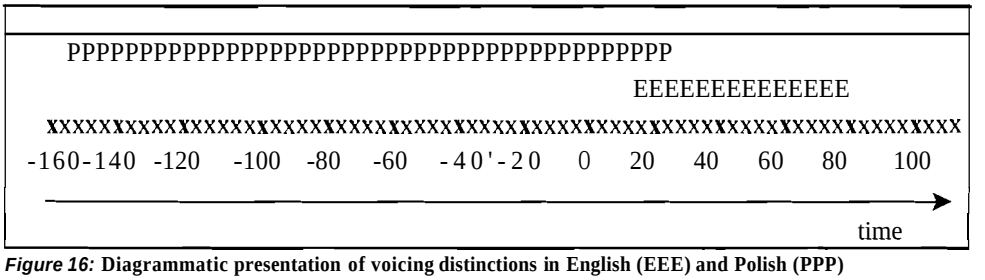


Figure 16: **Diagrammatic presentation of voicing distinctions in English (EEE) and Polish (PPP)**

The diagram in Fig. 16 shows two important observations: (1) that the ranges of VOT difference drastically differ in the two languages studied: for Polish, the pd difference is large and it reaches 200 ms., while for English it is about one-fourth of this value. This observation prompts explanation of the importance of aspiration in English if the voiced plosive is only partially voiced: aspiration enhances the pd low spectral frequency voicing contrast by adding noise placed in the upper part of the spectrum. (2) It also explains why it is often the case that voiced plosives pronounced by native speakers of English are often mistakenly taken by Polish leamers of English for voiceless sounds: VOT values for Polish voiceless and English partially voiced plosives overlap (cf. the diagram in Fig. 16).

IV.2. Word medial after a voiceless sound (English: s-)

In both languages a word initial s- cluster requires a following voiceless obstruent; while Polish uses the 'normal' voiceless plosive, English requires an untypical for English unaspirated voiceless plosive. Disregarding strength of articulation, in this context Polish learners of English can use the Polish VOT strategy[5]:

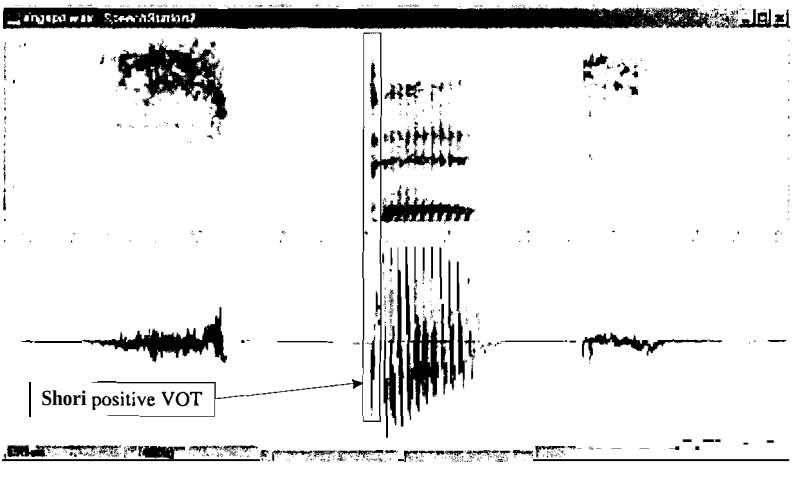


Figure 17: **English 'split': short VOT**

It is now easy to explain why there is no aspiration following a fortis plosive if it is preceded by s-. There have been numerous attempts trying to explain this restriction by claiming that so much effort is expended on the articulation of s- that not much energy is left for aspiration. In fact explanation should be based on a claim that articulatory effort is reduced where it is not necessary[6]. In this case, since s- (which happens to be the only possible first element of a word initial cluster) can be followed only by a voiceless obstruent, and cannot be followed by a voiced obstruent, there is no need to provide an additional cue of 'aspiration' that otherwise serves the function of distinguishing between English 'voiced' and 'voiceless' sounds. Therefore, in the position in which the voicing contrast is suspended (i.e. after s-), /p, t and k/ are not aspirated. Experimental findings show that unaspirated voiceless plosives, when spliced from their natural context and replayed, sound like partially voiced plosives, which strongly supports the observations shown in the diagram in Fig. 16.

IV.3. Word medial voiced plosives appearing before a voiceless sound

In this context English plosives retain their phonemic voicing characteristics, while in Polish they undergo regressive assimilation in voicing whereby the whole cluster becomes voiceless; in fact the process is more general and concerns all obstruents (cf. part 2 above for a description of its implementation on fricatives).

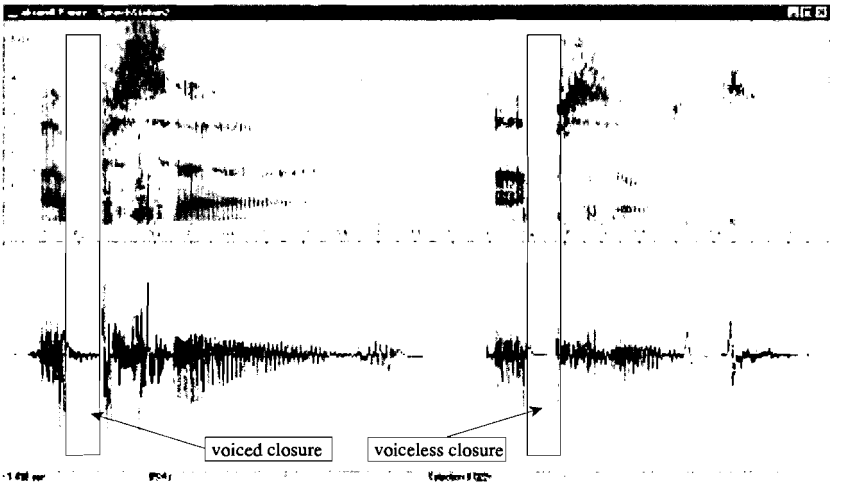


***Figure* 18: English and Polish pronunciation of 'absurd'**

Hence Poles have to learn to retain voicing in clusters if a voiced plosive is followed by a voiceless stop. Access to spectrographic imaging facilitates the acquisition of this initially difficult timing strategy.

IV.4. Word final voiced stops

In the pre-pausal position (i.e. absolute word final position), Polish voiced obstruents lose their voicing, while English voiced plosives retain part of voicing (voicing into closure); in textbook terms, English obstruents in word final position are partially (de)voiced:

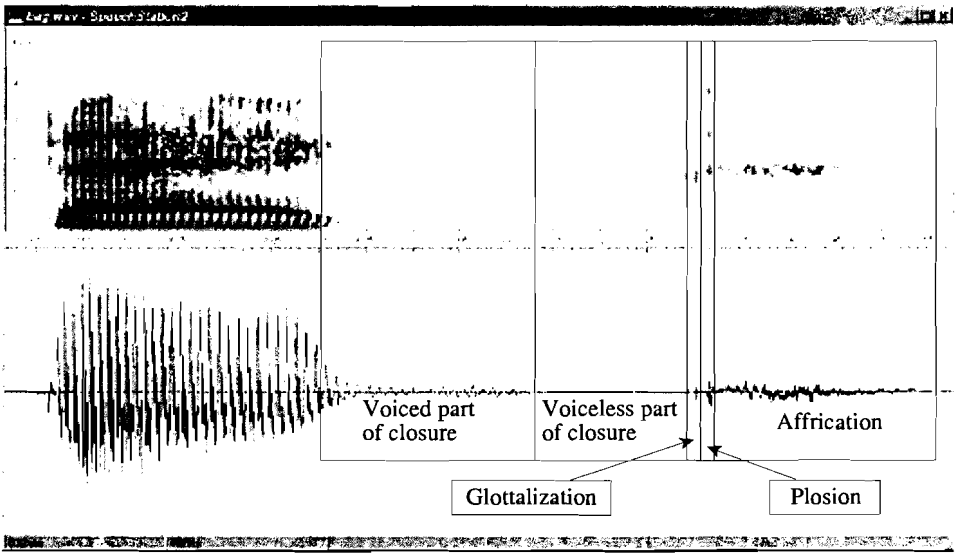
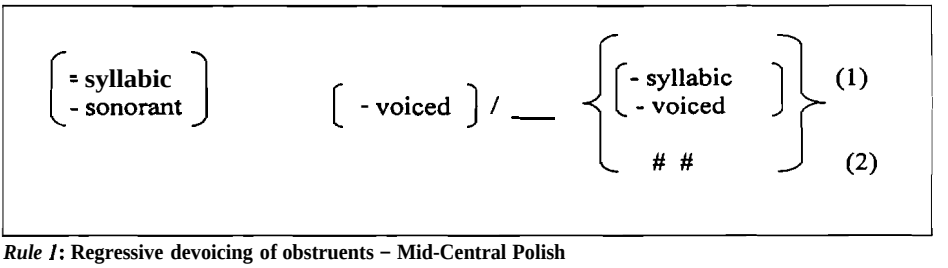


Figure 19: **English 'bag' with a partially voiced [g]; VIC = 219 ms., voiceless part of closure − 165 ms.**

The voicing state of Polish obstruents in this position depends on the geographical accent. In Mid-Central Poland (Warsaw), word final voicing rule is quite general:'

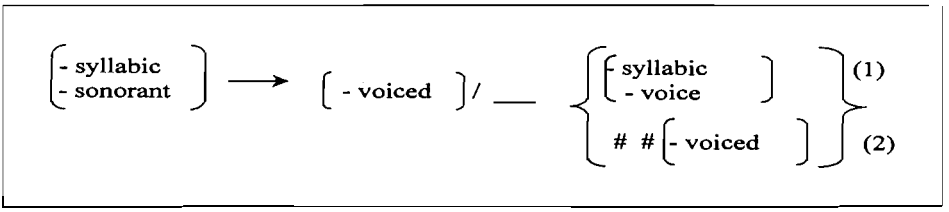


***Rule 1*: Regressive devoicing of obstruents − Mid-Central Polish**

Context (2) describes devoicing of all voiced obstruents before a pause (thus, e.g. the word 'zjazd' ('meeting') is rendered as [zjast]. According to Context 1, a voiced obstruent is

devoiced before a voiceless consonant immediately following it in the same word, e.g. /v/ in 'wstęp' ('entry') is realized as [f]: [fstemp], or /z/ in 'zjazd' is realized as [s] in [zjast].

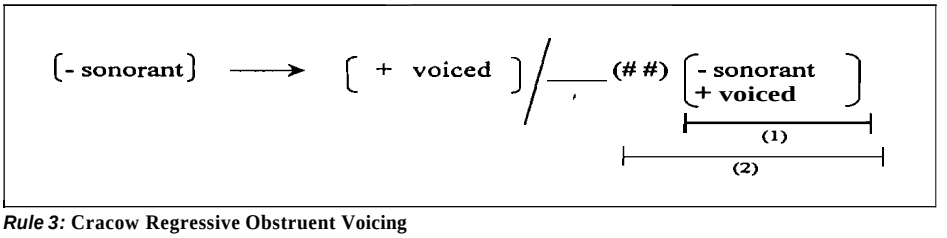
In Southem Poland (Cracow-Poznan) the word final obstruent retains its voicing if the following word's initial sound is voiced; the form of the rule is more restrictive than Rule 1:



Rule 2: **Cracow Voicing Retention**

As can be seen in Rule 2, Context (2) becomes more restricted by confining devoicing only to such cases of connected speech in which the following word begins with a voiceless consonant, e.g. 'wróg ciotki' ('aunt's enemy') is realized as [vrukc'otki], 'grób kolegi' ('colleague's grave') is pronounced as [gmpkolegi], 'sad sąsiada' ('neighbour's orchard'), as [satsow–s'ada], etc. Rule 2 implicitly assumes the lack of devoicing of voiced obstments in other connected speech contexts, i.e. before vowels or before sonorant consonants: 'wróg wujka' pronounced as [vmgvujka] ('uncle's enemy'), 'grób dziadka', as [grubdz'atka] ('grandfather's grave'), 'sad wiśniowy' as [sadvis'n'ovy] ('cherry orchard'), etc.

The occurence of voiced consonants in contexts that usually promote voicelesness can be also due to another Southem Polish derivation mechanism that is shown in Rule 3 below:



Rule 3: **Cracow Regressive Obstruent Voicing**

Rule 3 describes voicing taking place in two contexts: (1) in which word boundary '##' is omitted, and (2), in which it *is* taken into account. According to the rule, part (1) causes voicing of the type 'byliśmy' ('we were') realized as [bIliźmy], while part (2) refers to a context that occurs in the next word, e.g. 'gdzieś z waszego' ('somewhere from your...') realizowane jako [gdz'eźzvaSego]; word final /s'/ becomes voiced under the influence ofthe word initial /z/ in the next word. Spectrogram in Fig. 20 shows a continuity of voicing:

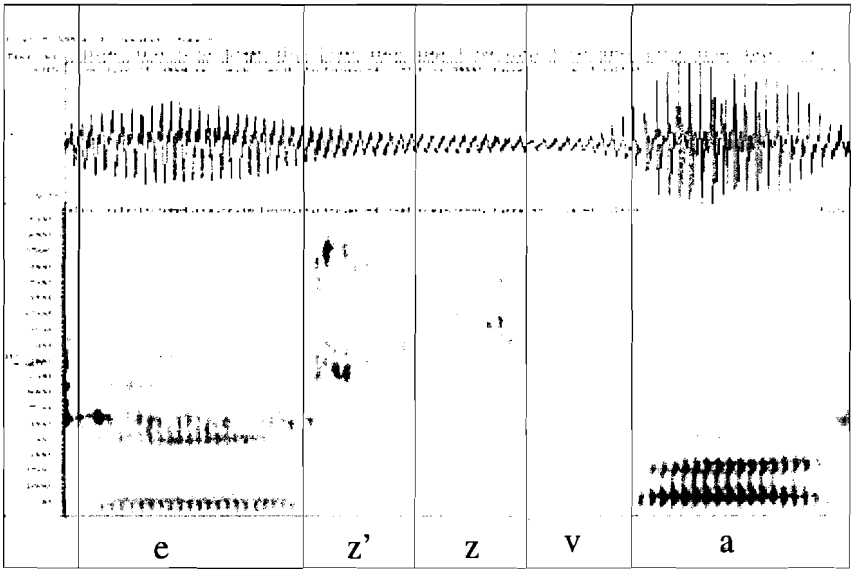


Figure 20: **'gdzieś z waszego' realized as |gdz'ez'zvaSego|. Only the non-parenthesized part of the word is shown on the spectrogram**

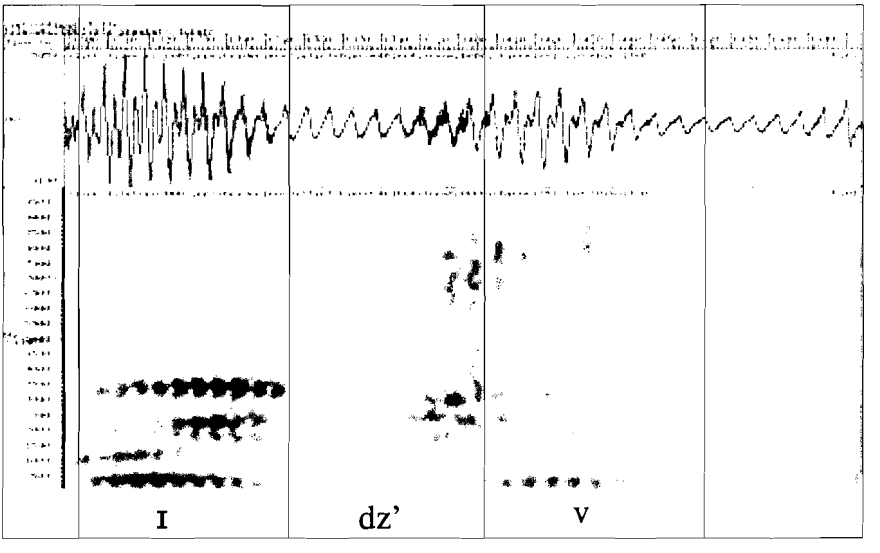


Figure 21: **'być w Łodzi' ('be in Lodz') realized as |bIdźvw(odzi)|**

If in the word final position of the first word there is a cluster of voiceless consonants, and the following word starts with a voiced sound, then voicing will concern the whole cluster (through voicing and successive assimilations):

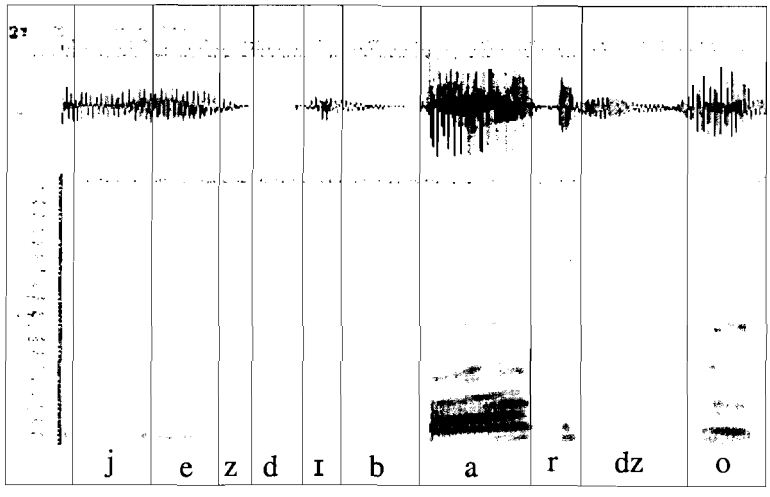


Figure 22: **'jest bardzo' ('is very') realized as [jezdɪbardzo]**

As the result of such voicing, the sounds that appear are in fact members of other phonemes; hence it can be said that in this accent variant of Polish, the context of a following voiced consonant causes a neutralization of voicing of word final consonant(s) occurring in the preceding word. The situation can become even more drastic as it can cause the origination of a sound that is not a part of the phonemic inventory of Polish: a voiceless /x/ is voiced to a voiced velar fricative /G/:

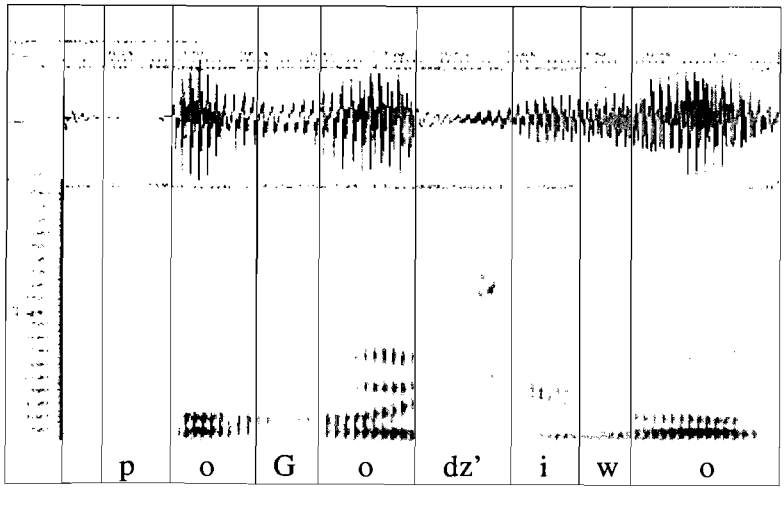


Figure 23: **'pochodziio' realized as [poGodziło]**

Summing up the question of correct rendering of English voicing in Polish learners, there are a number of points which can be regarded as positive phonetic/phonological interference:

(1) Between voiced sounds, English and Polish voiced plosives are fully voiced:
(2) Word final voiceless plosives are voiceless in both languages; optionally, English plosives can be aspirated.
(3) Disregarding force of articulation, an English unaspirated voiceless plosive is identical with a typical Polish voiceless plosive.
(4) If appropriate, Cracow Voicing Retention rule (Rule 2) helps to maintain voicing in word final obstruents.

Negative interference from Polish concerns the following situations:

i) Retention of voicing in word final position;
ii) Retention of voicing before voiceless obstruents;
iii) Correct pronunciation of partial (de)voicing in all appropriate positions;
iv) Cracow Regressive Obstruent Voicing rule (Rule 3).

V. CONCLUSION

Polish admits a two-way contrast, i.e. between fully voiced and fully voiceless obstruents, while English has the following contextual variants of voiced plosives: (i) initially devoiced, (ii) voiceless unaspirated, (iii) voiceless aspirated, (iv) fully voiced, and (v) finally devoiced.

The realization of these facts and practice enhanced by the use of visualization techniques greatly facilitate the acquisition of new pronunciation habits by foreign learners of English. The technique suggested in the present paper can certainly be applied to teaching English pronunciation to learners coming from other language backgrounds. It should be noted that not more than basic knowledge of speech visualization is sufficient to appreciate its pedagogical role.

NOTES:

1. The tenns 'partial devoicing' are equivalent: the former emphasizes the result, while the latter, the direction of the process.

2. Cf. Lisker and Abramson (1964), Ladefoged (1975:124), Port and Rotunno (1975: 654), Cruítenden (2001: 152-153); more references in Gonet (1989: 44-47).

3. According to Jassem (1970), quasi-periodic vibration, noise and the superposition of the latteron the former are three of the four basic types of acoustic events used in speech; the fourth is impulse corresponding to a plosion.

4. The values given here are only illustrative, based on individual measurements. A more extensive study of perceptual distance, based on a large number of examples and evaluated with statistical infcrence, is under way (Gonet, in preparation).

5. Cf. also Cruttenden (2001:152).

6. For an approach based on interaction between articulatory and perceptual drives consult Boersma (1998).

7. Rule 1 is formulated in the convention of Chomsky and Halle (1968): the slash '/' divides the description of the *change* on its left from the specification of the *context* on its right. The change (or: process) is specified by means of distinctive features that uniquely define the class of sounds undergoing it (non-syllabic non-sonorants à hence obstruents) and its operation (devoicing) The description of the input class, according to the economy convention, is devoid of predictable (redundant) elements; therefore the class of obstruents is not defined here as [+voiced], as devoicing must concern [+voiced] sounds. The context in which the change takes place is indicated by an underscore '___'; in Rule 1, it takes place *before* the specified elements of the context that are disjunctive: either before a consonant (Context 1) or in absolute word final position (i.e. before a pause): '##' (Context 2).

REFERENCES

Boersma, Paul (1998). *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives.* Netherlands Graduate Schoolo of Linguistics.

Chun, D. M. (1998) Signal analysis software for teaching discourse intonation. *Language Learning & Technology,* Vol. *2,* No. 1, July 1998, pp. 61-77.

Cruttenden, A. (2001) Gimson's Pronunciation of English. Sixth Edition. London: Arnold.

Gonet, W. (1981) "Legato speech", or the psychology of teaching phonetic fluency. *Lubelskie Materiały Neofilologiczne,* pp. 307-319.

Gonet, W. (1989) *Factorial Analysis of the Duration of R. P. Monophthongs in Monosyllabic Words.* Unpublished Doctoral Dissertation. Lublin: UMCS.

Gonet, W. and G. Pietroń. (1999). Semi-Automatic Continuous Speech Sgmentation and Labelling Procedure. *Speech and Language Technology.* Vol. 3. Poznań: Polish Phonetic Association.

Gonet, W., A. Trochymiuk, M. Goworek and L. Kowalczyk (2001). Protrain© – An lnteractive Visual Feedback Pronunciation Training System For Foreign Language And Hearing lmpaired Learners. In: Bőttinger, K., S. Dőnninghaus and R. Marzari (eds). Beiträge der Europaischen Slavistischen Linguistik (POLYSLAV 4). München, (Die Welt der Slaven. Sammelbande).

Gonet, W. and R. Święciński. (in press). Speech Lab @ Work and @ Home. *Speech and Language Technology*, Vol. VI. Poznań: Polish Phonetic Society.

Gonet, W. (in preparation). *Articulation andperception ofvoicing.*

Jassem, W. (1970). *Podręcznik wymowy angielskiej*. Warszawa: PWN.

Ladefoged (1975). *A Course in Phonetics*. New York, Chicago, San Francisco, Atlanta: Harcouri Brace Jovanovich, Inc.

Lisker, L. and Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical Measurements. *Word* 20,384-422.

Nowocien, P. (2000). *Voicing of Polish and English initial lenis obstruents in isolated words.* Unpublished M.A. thesis written under the supervision of Dr. Wiktor Gonet. Lublin: UMCS.

Port, R. F. and R. Rotunno (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America* **66**, 654-62.

