Yash Khasbage, Daniel Alcaraz Carrión, Jennifer Hinnell, Frankie Robertson, Karan Singla, Peter Uhrig and Mark Turner*

The Red Hen Anonymizer and the Red Hen Protocol for de-identifying audiovisual recordings

https://doi.org/10.1515/lingvan-2022-0017
Received February 6, 2022; accepted September 7, 2022; published online December 26, 2022

Abstract: Scientists of multimodal communication have no established policy or default tool for sharing de-identified audiovisual recordings. Recently, new technology has been developed that enables researchers to de-identify voice and appearance. These software tools can produce output in JSON format that specifies bodypose and face and hand keypoints in numerical form, suitable for computer search, machine learning, and sharing. The Red Hen Anonymizer is a new tool for de-identification. This article presents the Red Hen Anonymizer and discusses guidelines for its use.

Keywords: data science; de-identification; gesture; IRB

1 Introduction

Advances in science often depend upon great ranges of data. In genomics, materials science, neuroscience, energy, and other fields, scientific advances often derive from new methods of amassing and sharing data. By contrast, researchers who study advanced higher-order human cognition and communication are often blocked from sharing data because of concerns about privacy. There are boilerplate mechanisms for sharing files securely, such as access control lists, as implemented in software such as Box.[1] There are cybersecurity regulations such as the USA Defense Federal Acquisition Regulation Supplement. There are standard routines for de-identifying textual spreadsheet data that result from, e.g., computer-mediated behavioral experiments or surveys: for such data, the researcher typically deletes any columns carrying identifying information before sharing the data file with other researchers. Articles in, e.g., behavioral economics typically state "Data available upon request," or even that the dataset is available from a repository.

But there is no established policy or default tool for de-identifying *audiovisual* recordings including people so that they may be shared with other researchers, even though recent technological advances have led to exponential growth in the availability of such multimodal data for research purposes. In this paper, we present new technology that allows researchers to de-identify voice and appearance in audiovisual recordings, using the Red

1 www.box.com.

***Corresponding author: Mark Turner**, Case Western Reserve University, Cleveland, USA, E-mail: mark.turner@case.edu. https://orcid.org/0000-0002-2089-5978, http://markturner.org
Yash Khasbage, Microsoft, Bangalore, India, E-mail: yashkhasbage25@gmail.com. https://orcid.org/0000-0002-6730-7846, https://www.linkedin.com/in/yash-khasbage/
Daniel Alcaraz Carrión, Fundación Séneca, University of Wisconsin, Madison, USA; and University of Murcia, Murcia, Spain, E-mail: daniel.alcaraz.carrion@gmail.com. https://orcid.org/0000-0001-5325-431X
Jennifer Hinnell, University of British Columbia, Vancouver, Canada, E-mail: jennifer.hinnell@ubc.ca. https://orcid.org/0000-0002-4809-3512
Frankie Robertson, University of Jyväskylä, Jyvaskyla, Finland, E-mail: frankie@robertson.name. https://orcid.org/0000-0001-8862-7032
Karan Singla, Interactions Corporation, Los Angeles, USA, E-mail: ksingla@interactions.com
Peter Uhrig, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, E-mail: peter.uhrig@fau.de. https://orcid.org/0000-0003-1237-7414

Hen Anonymizer (RHA). The RHA technology can also produce output in JavaScript Object Notation (JSON), a text-based format that specifies bodypose and face and hand keypoints in numerical form, rendering the data suitable for computer search, machine learning, and sharing. Here, we introduce the Red Hen Anonymizer as a new tool for de-identification of audiovisual recordings.

2 The international distributed little Red Hen Lab

Red Hen Lab[2] (Joo et al. 2017; Steen et al. 2018) is an international group of researchers whose mission is to advance the science of multimodal communication. Red Hen Lab researchers work in many fields in the humanities, social sciences, computer science, cognitive science, and data science. They share their resources and research findings. Red Hen Lab's first goal is to advance the theory of multimodal communication; its second is to develop computational, statistical, and technical tools to improve research possibilities; its third is pedagogy.[3] Red Hen Lab has no formal organization or staff. It is a cooperative, a bazaar of alternatives and possibilities, not a service. One of its primary resources is audiovisual data. At present, Red Hen Lab holds more than 520,000 h of audiovisual recordings in more than 25 languages. These comprise more than 7.5 billion words in metadata files, including time-aligned close-captioned text. The recordings occupy more than 150 TB of storage. Automatic capture and processing increases this dataset by about 150 h per day. This dataset is an official archive of both the UCLA Library and the CWRU Library. The Red Hen dataset is not publicly shared.

The largest single component of the Red Hen dataset is called "NewsScape." NewsScape consists of recordings of public news broadcasts. IRB approval is not needed to collect, analyze, or share NewsScape data with researchers. In the United States, gathering, processing, and loaning such data is explicitly governed and legitimated by Section 108 of the US Copyright Act.

Beyond working with NewsScape data, Red Hen researchers often use Red Hen's tools on data gathered by other means, including, e.g., videos recorded during behavioral experiments. Red Hen researchers can place these data inside a protected space in Red Hen and use Red Hen analytic tools to process and study the data. For example, researchers might want to tag a large dataset, and use the Red Hen Rapid Annotator to do so. The Red Hen Rapid Annotator presents in rapid succession examples to be tagged, along with a list of tags that can be selected. A researcher interested in timeline gestures (see, e.g., Alcaraz Carrión et al. 2020) might supervise research assistants who glance at candidate video clips or still photos and select from a list of tags: (1) timeline gesture, (2) no timeline gesture, (3) no people present, (4) gesture not visible (e.g. out of frame), (5) unsure – send to researcher. The very rapidly tagged dataset might then be fed to a machine learning system to create a gesture classifier through a supervised learning process to answer further research questions. But to share such data outside the research group, the researcher might need to de-identify the data. If privacy concerns block the sharing of the original data, then the tagged data cannot be built upon by other researchers. The Red Hen Anonymizer and the Red Hen Protocol are meant to make it possible to share data that otherwise could not be shared.

3 The benefits of sharing data and the need for an initial default protocol

Sharing data is fundamental to scientific advance. In domains that use recordings of human communicative performances, science would be served by the appropriate sharing of audiovisual data. Researchers would ideally share data in publications and conference presentations so that others can test the published analyses and inferences. Researchers would also ideally share data by creating open data sets for other researchers to use widely, including in machine learning, which requires large datasets. The creation, curation, and sharing of such

2 http://redhenlab.org.

3 https://sites.google.com/case.edu/techne-public-site/.

open data sets is the norm in many other areas of research in which de-identification is not an issue. One of the chief rationales for the establishment of known and widely-used datasets is so that various tools, especially computational and statistical tools, can be tested against identical datasets. Tools can be compared according to their performance on a known dataset.

While some data can be shared publicly, some fall under governmental restrictions whose purpose is to protect Human Subjects in behavioral research. The authority in the USA is the "Revised Common Rule."[4] There are analogous authorities in other nations. Researchers often wish to use Red Hen tools on data that might be regarded as subject to Human Subjects restrictions. A typical example of data that Red Hen would like to share is audiovisual recordings of face-to-face conversations including co-speech gestures between two subjects (Kendon 1972, 2004; McNeill 2005). A common design for this type of research deploys cartoon retelling (e.g. McNeill 1985, 1992). In this design, one subject is asked to watch a cartoon and to subsequently retell it to a second subject, usually a friend of the first subject, who has not seen the cartoon. The interaction is recorded in order to capture the speech and co-speech gesture of the first subject as that subject retells the content of the cartoon. Researchers would then use Red Hen analytic tools on the data.

In the spirit of open science, researchers would ideally be able to share data with new collaborators. They would also be able to create full open datasets of the recordings from the behavioral experiments. Such datasets might be useful for many different research projects. The researchers would anonymize, organize and publish the data as a resource for other scientific researchers, sharing not only the results of their analyses, but also the original data. The broader scientific community might also benefit from using such data as stimuli for other experiments, for brain imaging research, and so on.

In some cases, authorities will allow data to be shared with the explicit written permission of the experiment participants. However, this is not always the case, and furthermore, subjects might hesitate to grant permission in the absence of reasonable anonymization. The barrier to sharing the original data is often a requirement that it be de-identified. For textual data, there are standard techniques of de-identification, such as eliminating names, email addresses, institutional identifiers, identifying answers to a post-experiment questionnaire, etc. But there is no common tool and no common policy for de-identifying audiovisual recordings. The Red Hen Anonymizer and the Red Hen Protocol are steps to creating the tool and the policy.

4 The Red Hen Anonymizer

4.1 Background

The Red Hen Anonymizer was conceived in 2019. In its earliest phase, it deployed computational systems that transform a recording of live action into an animated sketch version. But this technique was inadequate for de-identification: the sketch versions are too good. It is remarkable how sketch artists over millennia have developed ways to capture a human being in a few swipes of charcoal that convey data sufficient for identifying the subject. Computer sketching systems inherit that power. In the intervening three years, spectacular success has been made in a different direction: anonymizing or switching faces in videos.

Since 2015, Red Hen Lab has participated as a managing organization in Google Summer of Code, an Open Source initiative. The Red Hen Anonymizer was offered as a project for the 2021 Red Hen Lab Google Summer of Code. Red Hen Lab accepted Yash Khasbage's proposal to work on the design specifications for an Anonymizer over the 10 week Google Summer of Code period. The mentor team consisted of Mark Turner, Daniel Alcaraz Carrión, Karan Singla, and Peter Uhrig. The face-hiding, face-swapping, and audio anonymization reported here were developed by Khasbage and the mentors. The OpenPose option was added by including functionality supplied by OpenPose, as adapted by Peter Uhrig and Frankie Robertson and installed by Yash Khasbage (Cao et al. 2019). Jennifer Hinnell provided expert guidance on the needed functionality and possible applications to

4 "Federal Policy for the Protection of Human Subjects" *Federal Register*, Vol. 82, No. 12, Thursday, January 19, 2017, Rules and Regulations; https://www.govinfo.gov/content/pkg/FR-2017-01-19/pdf/2017-01058.pdf.

linguistics and gesture research. The coding was performed inside Mark Turner's research group on the CWRU High Performance Computing Cluster, whom we thank for its support. The resulting tool is the Red Hen Anonymizer, which is available from a repository by that name in the RedHenLab GitHub.[5]

4.2 RHA: how it works

The Red Hen Anonymizer (RHA) can test a clip to determine whether it is audio-only, video-only, or audiovisual. It can then anonymize the voice in a variety of ways and the appearance of the subject in a variety of ways. RHA anonymizes audiovisual data by anonymizing both the voice and the subject's appearance and recombining them using the multimedia platform ffmpeg.

4.2.1 RHA: audio anonymization

For any given research project, the researcher will need to retain some features of the voice of the speaker but not others. RHA extracts audio from video using ffmpeg and saves it as a .wav file. The .wav file is anonymized using SoX (SoundExchange) transforms. The Red Hen Anonymizer presents the researcher with a menu of voice features; the researcher chooses to change some features that are unrelated to the object of the study. SoX offers 61 audio parameters, called effects, most with several possible values. The Red Hen Anonymizer can in principle use all SoX effects, which can be combined. Even if only 20 parameters were available, each with, say, only three possible values (where one of those values omits that effect), the possible combinations would number about 3.5 billion. The most frequently used effect is *–pitch*. Red Hen researchers often also use *–echo* and *–distortion*. To make reversibility difficult, audio anonymization often uses a "vocoder" process. The reason vodocoding makes reversibility difficult is that "the original speech is never passed through to the output. Instead, information is extracted from the original audio, then used to generate a new signal. The process is inherently lossy enough to make it fairly prohibitive to reverse."[6]

There are many free, open-source vocoder packages available. SoX works as a vocoder: one uses the SoX "synth" effect to create a modulator of some sort, and then uses SoX to multiply the original voice by the modulator. Since RHA uses SoX, RHA automatically includes a vocoder. RHA's default setting for its vocoder deploys a 300 Hz sinusoidal wave modulator, to produce a "robotic" voice. The commands to make RHA substitute the robotic voice for the original are on Red Hen's GitHub.[7] The RHA vocoder process is sufficiently destructive to make reversibility prohibitively difficult, but it still preserves intelligibility and many communicative features of the speech. The original is Figure 22, below.

An example of the robotic voice is at http://go.redhenlab.org/rha/29, or click the QR code.

There are many audio morphing tools other than SoX, such as the Automatic Speech Morphing System.[8] In principle, the Red Hen Anonymizer could be extended to incorporate any of the functionality available in any of these platforms.

4.2.2 RHA video anonymization

For anonymizing appearance, RHA offers three options: (1) Body pose estimation with hand and face keypoint extraction, (2) Face-hiding, and (3) Face-swapping.

5 https://github.com/RedHenLab/RedHenAnonymizer.

6 See https://security.stackexchange.com/questions/227146/is-changing-pitch-enough-for-anonymizing-a-persons-voice for a detailed explanation.

7 https://github.com/RedHenLab/RedHenAnonymizer/issues/8.

8 Available at http://language.sakura.ne.jp/icnale/download.html.

4.2.2.1 Body pose estimation with hand and face keypoint extraction

Software of this sort performs automatic recognition of position and movement. That is, it makes its best guess at the location of joints and facial and manual keypoints. For example, Figures 1–3 represent schematizations of

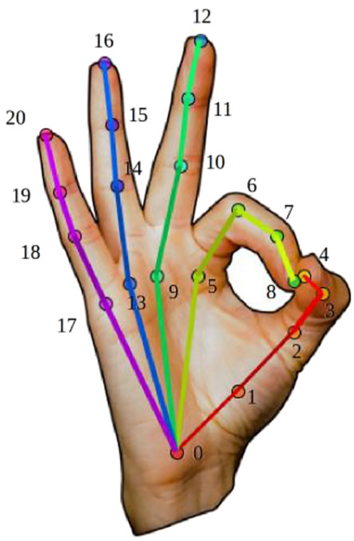

Figure 1: Schematization of standard hand keypoints [click link: http://go.redhenlab.org/rha/01 (or) scan QR code].

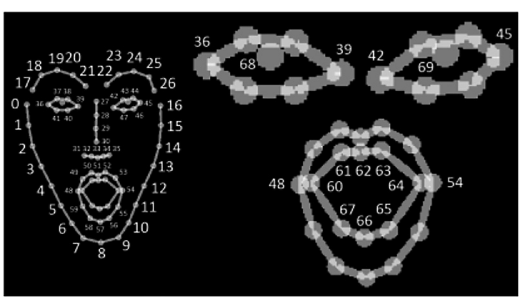

Figure 2: Schematization of standard facial keypoints [click link: http://go.redhenlab.org/rha/02 (or) scan QR code].

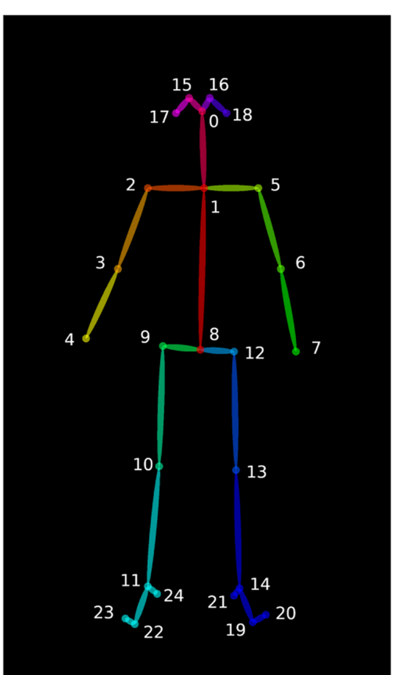

Figure 3: Schematization of standard body pose keypoints [click link: http://go.redhenlab.org/rha/03 (or) scan QR code].

some standard keypoints, as explained in the OpenPose Manual.[9] OpenPose and its manual are "freely available for free non-commercial use, and may be redistributed under these conditions."

The software determines keypoints purely from the audiovisual recording, without any interaction with the subjects, and without any supervision from the researcher. Indeed, one can apply these techniques to audiovisual recordings from any source, era, and country. They can be applied to early silent films, television shows, Hollywood blockbusters, home movies, etc. The computational analysis determines keypoints and expresses them as a set of time-stamped vectors. These vectors are represented in one of several standard computer formats, such as JSON. The most widely-used extraction software at the moment is OpenPose (Cao et al. 2019). The OpenPose Manual gives a thorough description of the kind of output provided by the software.

For the purposes of capturing the nature of the data, the software can render those vectors graphically as color-coded, moving stick figures. These stick figures, being only renderings of the vector data, have no identifiers.

Figure 4 shows an example of a blend of the original recording and the superimposed stick figures:

Figure 4: Blend of original recording with superimposed stick figure (This is an excerpt from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.) [click link: http://go.redhenlab.org/rha/04 (or) scan QR code].

But in cases where the identity of the individual people in the scene is to be protected, it is possible to share *just the skeleton and just the vectors*. Obviously, it is impossible to reverse engineer from these vectors or stick figures to recover the original data. Figure 5 shows an example of the stick-figure anonymization of an original video with additional audio anonymization:

Figure 5: The same as Figure 4, but without the underlying video and with anonymized audio [click link: http://go.redhenlab.org/rha/05 (or) scan QR code].

JSON files containing the numerical metadata can be used by the researcher for various purposes: computational search, comparison, detection of patterns, machine learning, and so on. These metadata are inherently utterly de-identified. A JSON file of metadata for a particular single frame of the original video for Figures 4 and 5 starts as follows and goes on in the same vein for 35 lines:

{"version":1.3,"people":[{"person_id":[-1],"pose_keypoints_2d":[827.225,259.156,0.84002,805.727,400.101,0.691057,646.981,390.353,0.483986,617.594,660.752,0.583629,868.442,686.174,0.604898,964.436,411.909,0.537032,1027.07,631.312,0.588458,1195.55,692.078,0.287076,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,788.032,223.85,0.872891,854.789,216.057,0.845437,729.395,243.383,0.873423,897.809,229.738,0.822803,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],"face_keypoints_2d":[734.99,222.455,0.758396, . . .

Suppose a researcher would like to search for cases where someone is looking directly at the camera. In principle, since the numbers indicate the position of the eyes, the nose ridge, and the tip of the nose, the code could check whether the nose line segment is strictly perpendicular to the line segment connecting the eyes. Alternatively, suppose a researcher would like to search for cases where someone is looking directly at the camera and pointing to the right (or performing a Palm Up Open Hand gesture to the right, etc.) In principle, all such cases are simple linear algebra calculations. Data anonymized in this way could be assembled into open datasets, could be used for machine learning, and, crucially, could be shared with other researchers and presented in publications. One example of such sharing—in which the data had to be anonymized owing to copyright restrictions—is the open

9 https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html.

dataset by Tong and Uhrig (2022), with the additional machine learning baselines developed in Swadesh Jana's Google Summer of Code project in 2021.[10]

4.2.2.2 Face-hiding

Scientific articles frequently anonymize stills of lab subjects by placing an oval over the face. Additionally, in gesture research, a sketch of the gesture is often superimposed on the anonymized photo, as in Figure 6.

("Example of Abstract Spatial Gesture," from Parrill, Fey & Kashmiri Stec. 2017. Gestures of the abstract. *Pragmatics & Cognition* 24(1). 33–61. DOI:10.1075/pc.17006.par).

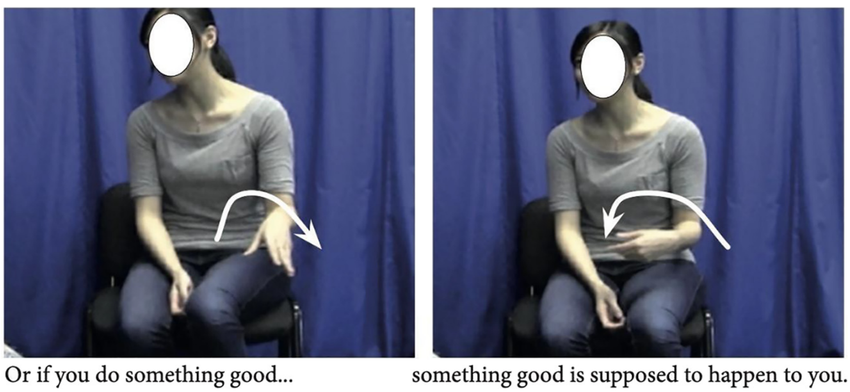

Figure 6: Example of anonymization in a published article using still image, face-hiding, and sketch suggesting the nature (here, trajectory) of the gesture [click link: http://go.redhenlab.org/rha/06 (or) scan QR code].

But such a de-identification technique for sharing data is inadequate for many purposes. What other researchers need is typically not an interpretation of what the original researcher saw (e.g. a circular arrow) but, rather, the actual dynamic gestures, and their combination with other gestures. RHA can reliably replace *all* faces with a rectangle (or other shape, such as an oval or a circle). No matter how low the confidence that a face has been recognized, the anonymizer replaces it with the blank shape. Any recognition mistake (e.g. recognizing a face in a poster on the wall and so superimposing a shape on it) might be inconvenient for the researcher, but it does not diminish anonymization. So far, we have found that such mistakes occur only in exceedingly rare cases, and only for complicated video from TV or movies, in which there are complicated backgrounds, with decorated vases, baroque wallpaper, framed photos on a mantle, and so on. Moreover, some of these mistakes are not really mistakes: the recognized object is a face—perhaps a sketch on a poster, or a decoration on a vase. In principle, the Red Hen Anonymizer can provide very many different shapes for replacing the face. At present, it offers blocking with either a white square or a white oval. Here (Figure 7) is a video where the face is anonymized with the face-hiding option of the Red Hen Anonymizer but the original audio is preserved:

Figure 7: A lab assistant demonstrating a gesture, with the face hidden by RHA [click link: http://go.redhenlab.org/rha/07 (or) scan QR code].

And here is a different original video (Figure 8) and a version where RHA has hidden the face and also anonymized the voice (Figures 9, 10, 12, 14, 16, 18, 20, 21, 23–26):

Figure 8: Original video [click link: http://go.redhenlab.org/rha/08 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.)

10 https://sites.google.com/site/distributedlittleredhen/summer-of-code/red-hen-lab-gsoc-2021-projects.

Figure 9: The video from Figure 8 but now with anonymized voice and hidden face [click link: http://go.redhenlab.org/rha/09 (or) scan QR code].

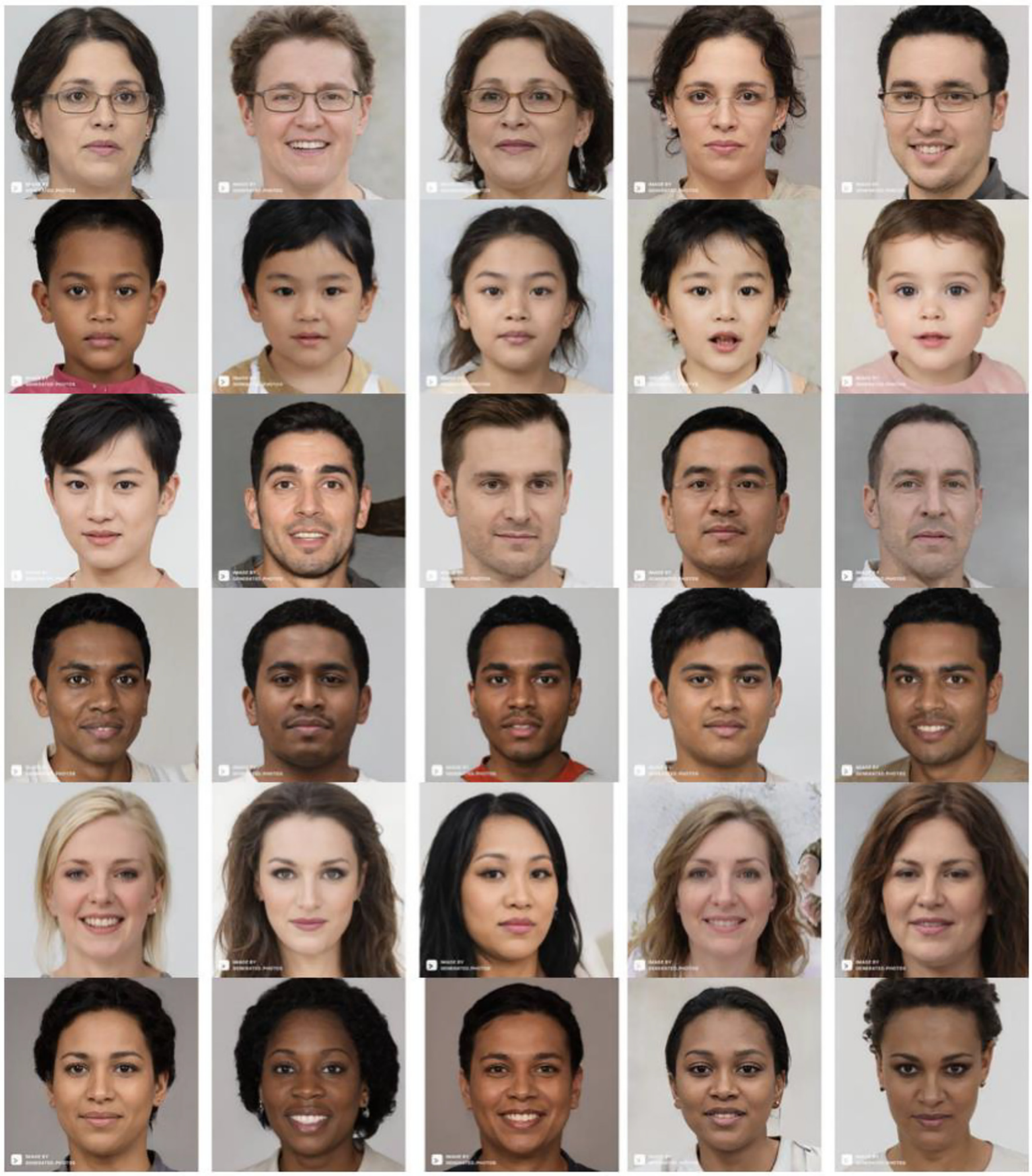

Figure 10: Computer-generated faces that belong to no one [click link: http://go.redhenlab.org/rha/10 (or) scan QR code].

4.2.2.3 Face-swapping

At present, gesture analysis relies on manual tagging. Manual tagging is a routine, widespread, painstaking method. Manual tagging is often facilitated by the use of specialized computer software, such as ELAN (Wittenburg et al. 2006). ELAN allows videos to be tagged frame by frame, with high temporal resolution, for maximum accuracy. No automatic tagging system at present can begin to compare with the quality of manual tagging done by experts. Manual tagging depends upon *gathering* data, by finding it or making it, and such gathering may in certain cases be subject to IRB review. Once it is gathered, it must be shared with the researcher's assistants, who do the tagging. Subsequently, it would be best if that data could be *shared* with researchers in the community. This sharing is indispensable if we want the analysis to be reproducible (see the *Nature* Special[11] on the problem of reproducibility). Keypoint extraction and face-hiding are two forms of de-identification to make sharing possible, but for many purposes, they lose too much information.

It is, however, possible to anonymize the appearance of an individual by mapping onto it the appearance of a different individual. Indeed, "deep fake" techniques allow us to map a recording of anyone onto an image of a celebrity: the result looks and sounds like the celebrity. Of course, the mapping can impose an appearance that in fact does not correspond to an existing personal identity. Such computer-generated faces are available to researchers and businesses free of charge. The Red Hen Lab Anonymizer routinely uses computer faces.[12] Here is a tiny selection of such computer-generated faces:

For example, we present below a still image taken from an original video alongside an anonymized version of that still image. The face-swapping is performed by open source software FSGAN (Nirkin et al. 2019).

In this example, we have hand-selected a computer-generated target face highly similar to the original, but face-swapping can easily produce dramatic differences. Human beings have a strong bias toward perceiving similarity, so if someone sees the original and the de-identified version, the viewer will typically grossly over-estimate the statistical similarity. The question for de-identification is: if the viewer sees only the de-identified version, what is the probability of their being able to recognize that the de-identified version derives from a particular person? Naturally, if a parent with three children is asked to identify the child behind a particular de-identified version, the parent is almost certain to be able to do so, because of deep knowledge of the individuals in the pool of candidates. But the number of, e.g., white beardless males in their early thirties who could be the source of the original for our example is huge. Unless the viewer can use other cues aside from face and voice to pick out the original, it is statistically quite acceptable to de-identify by face-swapping to any computer-generated face, even to a computer-generated face that matches the original on major demographic categories. In the case of celebrities—where there might also be distinctive backgrounds, highly distinctive verbal content that would dramatically narrow the pool of possibilities, and so on—the probability of identifying the original speaker increases. But even then, it might take great motivation and elite computational resources to identify the person behind the anonymized version. Such issues have always been part of de-identification for purposes of IRB approval. Researchers know to de-identify textual data by eliminating information about the university, the timestamp of the experiment, and so on, and they know to instruct subjects not to wear distinctive clothing, and so on. All such pre-existing constraints remain and are untouched by the possibilities offered by the Red Hen Anonymizer. The question for researchers is instead: if all other standard de-identification has been performed but there is an impediment because the face is identifiable, does de-identification through face-swapping with a computer-generated face remove the impediment to sharing the data?

11 https://www.nature.com/collections/prbfkwmwvz/.

12 Faces taken from https://generated.photos/faces.

Figure 11: Still from original video [click link: http://go.redhenlab.org/rha/11 (or) scan QR code]. (This is a still from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.)

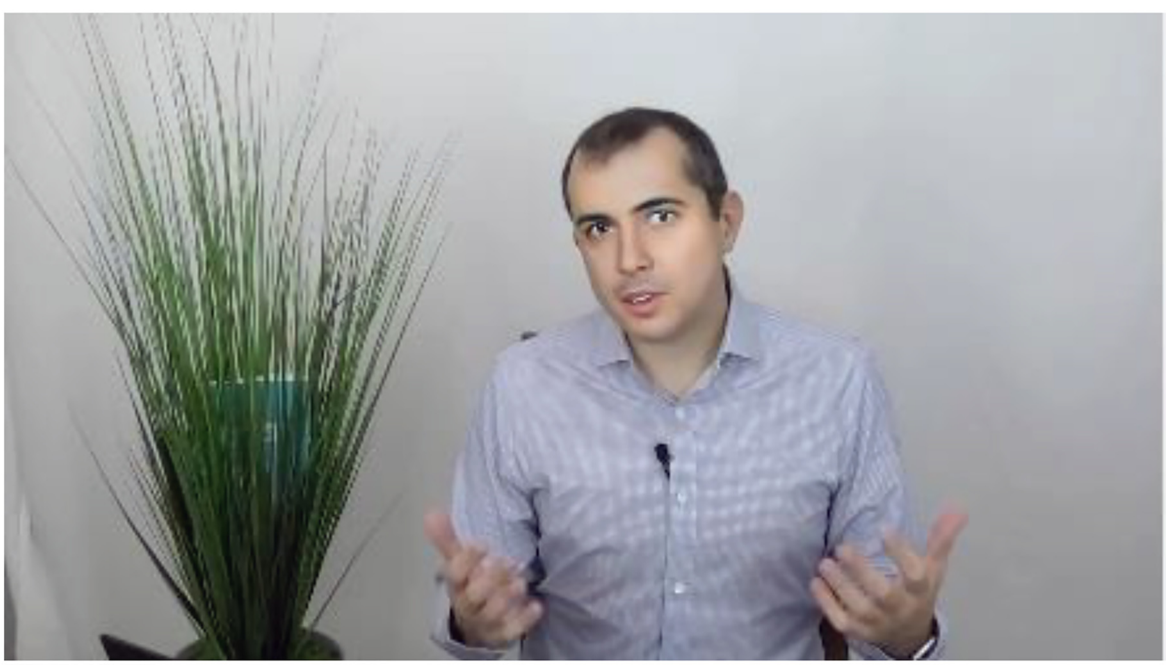

Figure 12: The still from Figure 11 with a swapped face [click link: http://go.redhenlab.org/rha/12 (or) scan QR code].

Below are links to other videos that have been anonymized with face-swapping and voice de-identification. In all cases, the face used for swapping is a computer-generated face: it belongs to no one. Some of these anonymizations are sensational and are therefore unlikely to be used by a researcher, for various reasons of research design. We include them to give a sense of the power and scope of RHA. Extraordinary dissimilarity is as easy for RHA as statistically acceptable de-identification, and of course, no one would be able to gauge the degree of similarity or dissimilarity from the de-identified version if they did not also have the original data. RHA can turn a young person of one race and gender into an old person of a different race and different gender. In brief, any computer-generated face of any sort—there are millions—can be used as the target face for anonymization. We expect that the researcher will weigh as needed how different a target face should be from the original to serve the purpose of de-identification. This assessment of sufficient dissimilarity will vary as other cues for identification vary in the data. Statistically, the possibility of reliable identification from the de-identified version is negligible in the absence of other identifier cues of the sort that all researchers already know to either keep out of the original data or remove during the process of de-identification.

Figure 13: An original video [click link: http://go.redhenlab.org/rha/13 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.)

Figure 14: The video from Figure 13 but with face-swapping anonymization [click the link: http://go.redhenlab.org/rha/14 (or) scan QR code].

Figure 15: An original video [click the link: http://go.redhenlab.org/rha/15 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.)

Figure 16: The video from Figure 15 but with face-swapping anonymization [click the link: http://go.redhenlab.org/rha/16 (or) scan QR code].

Figure 17: An original video [click the link: http://go.redhenlab.org/rha/17 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=XCSfoiD8wUA; licensed under the Creative Commons Attribution License.)

Figure 18: The video from Figure 17 but with face-swapping anonymization [click the link: http://go.redhenlab.org/rha/18 (or) scan QR code].

Figure 19: An original video [click the link: http://go.redhenlab.org/rha/19 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=sXb0S1Mz2g0; licensed under the Creative Commons Attribution License.)

Figure 20: The video from Figure 19 but with face-swapping anonymization [click the link: http://go.redhenlab.org/rha/20 (or) scan QR code].

Figure 21: The video from Figure 19 but with a second face-swapping anonymization [click the link: http://go.redhenlab.org/rha/21 (or) scan QR code].

Figure 22: An original video [click the link: http://go.redhenlab.org/rha/22 (or) scan QR code]. (This is an excerpt from https://www.youtube.com/watch?v=X_IorAF7FqQ; licensed under the Creative Commons Attribution License.)

Figure 23: The video from Figure 22 but with a first face-swapping anonymization and with speech anonymization to a robotic voice, produced by RHA Sox vocoder transform sine 300 [click the link: http://go.redhenlab.org/rha/23 (or) scan QR code].

Figure 24: The video from Figure 22 but with a second face-swapping anonymization [click the link: http://go.redhenlab.org/rha/24 (or) scan QR code].

Figure 25: The video from Figure 22 but with a third face-swapping anonymization [click the link: http://go.redhenlab.org/rha/25 (or) scan QR code].

Figure 26: The video from Figure 22 but with a fourth face-swapping anonymization [click the link: http://go.redhenlab.org/rha/26 (or) scan QR code].

Given an input video or still image to be anonymized, the Face Swapper preserves some features of the input, and some are borrowed from the target face. The following are the features that are preserved or not preserved while anonymizing an input:

Preserved	Not preserved
Lip movements	Lip features
E.g. the lip movements of speaking “Hello there”	E.g. the color and appearance of lips
Eye gaze	The appearance of the eyes
E.g. the direction where a person is looking	E.g. color
Eyebrow movements	Eyebrow features
E.g. raising or knitting the eyebrows	E.g. color and thickness of eyebrow
Face angles	Face features
E.g., the directions in which the face points	E.g. skin color of face, wrinkles of skin

The current state of RHA does not support changing the hair. The eyeglasses are handled by FSGAN variably, and so may be preserved variably, depending on several factors, such as the color of the glasses, the nature of the overlap with eyes and eyebrows, and the light reflected from the glasses.

In short, RHA preserves the position of keypoints on the face. Accordingly eye-gaze, face angle, and face keypoint movement are preserved, but other features are replaced. Body pose is of course also preserved for the parts of the body other than the face, since only the face is swapped.

5 De-identifying experimental stimuli

De-identification is similarly required for purposes aside from sharing data with researchers. Many experiments use data as *stimuli* that are subject to privacy constraints. For example, at present, a video used as a stimulus in an experiment might be blurred or partially blurred so as to obscure the identity of the recorded performer. An article reporting such an experiment might publish a snapshot from the video stimulus to suggest to the reader the nature of the stimulus. Figure 27 shows an example of such a snapshot from an article that reports results from using blurred videos as stimuli.

Figure 27: Example of anonymization in experimental stimuli with obscured face, head, and shoulders (Hinnell and Parrill 2020; Parrill et al. 2022) [click the link: http://go.redhenlab.org/rha/27 (or) scan QR code to see the snapshot].

[and click this link: http://go.redhenlab.org/rha/28 or scan the following QR code to see the blurred video from which the snapshot is taken].

In this experiment design, the shoulders and face were blurred to obscure the fact that the audio and video in the stimuli had been edited across conditions (i.e., the audio did not always match mouth movements). However, regardless of the reason for de-identifying experimental stimuli, the losses to science here are severe. First, much of the interesting performance in the original version of the stimulus is lost under blurring. Face-swapping or even face-hiding would lose less of, or at least different aspects of, the original performance. Even skeleton videos would convey facial data in a way the blurred video cannot. Second, the reader of the article receives an extremely limited version of even that blurred stimulus: all the reader sees is a snapshot. De-identification through more advanced techniques would improve both the experiment and its reporting.

6 Limitations

There are limitations on the de-identified data produced by RHA. Researchers will need to consider those limitations carefully. Some limitations result from the nature of the recording: the recording is not the performance; the recording will capture only some of the performance, in only some ways. For example, audiovisual recordings are two-dimensional and have a frame outside of which data are not captured. In addition, any of the de-identification transformations wrought by RHA further lessens the quality of the data, since they all destroy or transform aspects of the data in order to achieve de-identification. Bodypose and keypoint transformation eliminates all of the visual data and substitutes an imperfect, partial skeletal representation. A thorough review of limitations resulting from recording and from bodypose and keypoint transformation is provided in Pham (2022), which also proposes machine learning methods for compensating for the data limitations resulting from the nature of the recording. The face-hiding technique loses less of the recorded data for body and hands but loses all of the data for the face. The face-swapping technique keeps a great deal of the facial data but, like keypoint de-identification, is imperfect because the facial performance is driven by only a limited number of keypoints, as summarized in Nirkin et al. (2019). The individual researcher will need to consider the extent to which data loss produced by de-identification is acceptable for the purposes of any particular study.

7 Guidelines for de-identifying audiovisual recordings

Standard procedures were developed over time for de-identifying textual data resulting from behavioral experiments. A similar development is to be expected for de-identifying audiovisual data. To launch that development, Red Hen Lab presented RHA to the Case Western Reserve University Institutional Review Board in September, 2021. The CWRU IRB considered relevant issues for several months and, in October 2022, provided an initial set of guidelines for the use of RHA in Human Subjects research. These guidelines are based on the Revised Common Rule (2017). Red Hen Lab would be most grateful to learn of subsequent responses by similar authorities worldwide, to be added to this webpage.

Those guidelines, referred to as “the Red Hen Protocol on De-identifying Audiovisual Recordings,” or just “the Red Hen Protocol,” are available at http://go.redhenlab.org/rha/irb, or by scanning the QR code.

8 Conclusion

Technology for producing multimodal communication in media—in film and video games, for example—is astoundingly powerful and costly. None of this technology seems to be available to researchers in linguistics, gesture studies, and related fields dedicated to the scientific study of multimodal communication. The entry of such technology into our research methods is likely to be quick and thorough, establishing default practices as automatic to the researcher as are current methods of de-identifying textual data. The Red Hen Anonymizer[13] is a beginning point, one that will no doubt be forked and improved rapidly. The Red Hen Protocol is an attempt to establish not a global policy but instead a useful point of departure for deliberations about such policy. It is an efficient statement of expectations that can be cited in discussions with IRBs and analogous authorities worldwide. Red Hen predicts rapid change in the methods, technology, and policies for studying multimodal communication. This article is meant to spark that transition.

Acknowledgments: We gratefully acknowledge the funding of the software development by Google through their *Google Summer of Code 2021* program. Some of the research was carried out within the project *Talking Hands,* funded by the Competence Network for Scientific High Performance Computing in Bavaria (KONWIHR). Some of this work was supported by an Anneliese Maier Research Award, from the Alexander von Humboldt Foundation. Some of this work was supported by 21250/PD/19 Fundación Séneca. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

References

Alcaraz Carrión, Daniel, Cristóbal Pagán Cánovas & Javier Valenzuela. 2020. Enaction through co-speech gesture: The rhetorical handing of the mental timeline. *Zeitschrift für Anglistik und Amerikanistik* 68(4). 411–431.

Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei & Yaser Sheikh. 2019. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. arXiv.

Hinnell, Jennifer & Fey Parrill. 2020. Gesture influences resolution of ambiguous statements of neutral and moral preferences. *Frontiers in Psychology – Cognition* 11. 587129.

Joo, Jungseock, Francis Steen & Mark Turner. 2017. Red Hen Lab: Dataset and tools for multimodal human communication research. *KI – Künstliche Intelligenz* 31. 357–361.

Kendon, Adam. 1972. Some relationships between body motion and speech. In A. W. Siegman & B. Pope (eds.), *Studies in dyadic communication*, 177–216. New York: Pergamon.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review* 92(3). 350–371.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, David. 2005. *Gesture and thought*. Chicago: University of Chicago Press.

Nirkin, Yuval, Yosi Keller & Tal Hassner. 2019. FSGAN: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE international conference on computer vision*, 7184–7193.

Parrill, Fey & Kashmiri Stec. 2017. Gestures of the abstract. *Pragmatics and Cognition* 24(1). 33–61.

Parrill, Fey, Jennifer Hinnell, Grace Moran, Hannah Boylan, Ishita Gupta & Aisha Zamir. 2022. Observers use gesture to disambiguate contrastive expressions of preference. *Language and Cognition* 14(19). 46–28.

Pham, Kyle. 2022. *A temporal approach to recover full human pose in occluded videos through generating transformation options*. Cleveland, Ohio: Department of Computer and Data Science, Case Western Reserve University Unpublished Master of Science thesis. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4130581.

Revised Common Rule. 2017. Federal policy for the protection of human subjects. *Federal Register* 82(12). Rules and regulations.

Steen, Francis F., Anders Hougaard, Jungseock Joo, Inés Olza, Cristóbal Pagán Cánovas, Anna Pleshakova, Soumya Ray, Peter Uhrig, Javier Valenzuela, Jacek Woźny & Mark Turner. 2018. Toward an infrastructure for data-driven multimodal communication research. *Linguistics Vanguard* 4(1). 1–9.

13 Freely available on the RedHenLab Github at https://github.com/RedHenLab/RedHenAnonymizer.

Tong, Yao & Peter Uhrig. 2022. The Red Hen interview gesture collection. Available at: http://go.redhenlab.org/rhigc.
Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings of LREC 2006, fifth international conference on language resources and evaluation*. Genoa. Available at: https://hdl.handle.net/11858/00-001M-0000-0013-1E80-B.