HOW VOICE ANALYSIS CAN HELP SOLVE CRIMES

Amelia Gully[1], Philip Harrison[1,2], Vincent Hughes[1], Richard Rhodes[1,2] and Jessica Wormald[1,2]*

[1] Department of Language and Linguistic Science, University of York, York, United Kingdom

[2] J P French Associates, York, United Kingdom

YOUNG REVIEWERS:

RYAN
AGE: 15

WILLOW
AGE: 13

Imagine you are a police officer investigating a crime in which somebody has made a threatening phone call. There may not be any physical evidence like DNA or fingerprints, but the person's voice is captured on a recording. Experts can analyze the voice recording and compare it with the voice of a known suspect to find out how likely it is that they are the same person. This method is called forensic voice comparison. It works because people's voices contain lots of information about them. Your voice tells a story about where you grew up and learned to speak, and your voice also depends on your individual biological make-up. In this article, we explain how this information is used to compare recordings of voices in criminal cases.

HOW YOUR VOICE WORKS

Imagine that a criminal has made a threatening phone call, and the police have a recording of the call but do not know who made it. The voice in the recording can be compared to the voices of suspects,

to try to identify the criminal. This process is called **forensic voice comparison** [1].

To understand forensic voice comparison, we need to understand how sound and voices work. All sound is made up of vibrations in the air, and the **frequency** of a sound is how fast it makes the air vibrate. We hear low frequencies as low pitches and high frequencies as high pitches, but most natural sounds contain lots of frequencies that give us additional information—for example, a violin and a piano playing the same note sound different because they contain other frequencies as well as the pitch.

When we speak, sound is produced by the vocal cords, which are two small, vibrating muscles in your throat. The vocal cords make a buzzing sound, and how fast they vibrate determines the pitch of your voice. To turn the buzzing sound into speech, you also need the vocal tract, which is the name for the flexible tube between your vocal cords and your lips. To make specific speech sounds, you change the shape of this tube by moving what are called your articulators: your tongue, jaw, and other parts of your mouth. This changes the frequencies that are present in your voice [2].

Try it now: start by humming and place your hand on your throat. You should feel a vibration, caused by your vocal cords. While you are humming, your mouth is closed—your articulators are in the position used to produce the sound "m." Now, while still humming at the same pitch, open your mouth. It will sound like "ma" as your articulators move from the position for "m" to the position for "ah." You did not change the pitch of your voice, but by changing the shape of the vocal tract, you changed the other frequencies that were present, resulting in different speech sounds.

People's voices are different from each other for two reasons: *biological* differences in the size and shape of their vocal cords and vocal tracts, and *behavioral* differences in how they have learned to use their voices. For example, larger people will have bigger vocal cords, giving them lower-pitched voices. They did not learn this, it is just a result of their size. Alternatively, people with different accents might pronounce words differently. An English person might say "down" or "daan," but a Scottish person might pronounce it "doon." These people have *learned* to move their articulators differently, based on where they learned to speak.

The same person's voice can also be different in different situations. You probably know how different someone's voice sounds when they have a cold. This is because the nose is blocked, changing how air flows around the vocal tract. There are lots of reasons that your voice might change, including who you are speaking to, your mood, and even the time of day. Even if someone said the same thing one hundred times, no two repetitions would be identical. It is important

to know about these differences when we compare voices from various situations.

COMPARING VOICES

Due to the biological and behavioral differences in how people speak, voices contain information that is specific to each person. This allows people to be recognized from their voices.

To compare two voices and judge whether they belong to the same person, we need to consider how *similar* the voices are. If they are not very similar, this might suggest that the suspect is not the criminal. We also need to consider how distinctive the voice is. Two people may have similar voices just because they grew up in similar places. Therefore, forensic analysts explore whether the voices contain distinctive features that would not be shared by many people from a similar background. Voices that share a greater number of distinctive features provide stronger evidence that the suspect and the criminal are the same person, because fewer other people's voices would share those features.

It is difficult to say with 100% certainty that a voice in a recording belongs to a suspect because the human voice is so complex. Whereas, our DNA or fingerprints remain fixed throughout our lives, voices change for all sorts of reasons. This means that the same person could sound very different in two recordings made in different situations. In our example, the recording of the criminal's voice could be compared with a suspect's police interview recording. The situation of making a threatening telephone call is very different from the situation of being interviewed by the police, so a person's voice is likely to differ between these two settings. It is also possible that a person will change their voice when committing a crime, to disguise their own voice or sound like somebody else. This may make it very difficult to compare voices, but sometimes features can be detected despite the disguise.

LINGUISTIC ANALYSIS

Voices are analyzed by experts trained in **linguistics**, which is the scientific study of language and speech. Linguists examine various linguistic features in the threatening call and compare them with similar features in the police interview [3]. They break the speech down into various parts and analyze each part by carefully listening to specific sounds. This is called **auditory analysis**. They also examine speech using computer software to look at images of speech sounds, called **spectrograms**, such as those in Figure 1. This process is called **acoustic analysis**.

LINGUISTICS

The scientific study of how languages are formed, used, and understood.

AUDITORY ANALYSIS

A form of forensic speech analysis performed by experts, involving careful listening to different parts of voice recordings.

SPECTROGRAM

A visual representation of a sound signal, such as speech, which shows the frequencies that are present and how they change over time.

ACOUSTIC ANALYSIS

A form of forensic speech analysis performed by experts, making use of spectrograms, and other quantifiable measurements to determine the specific frequencies present in different parts of voice recordings.

Figure 1

DISFLUENCIES

Speech habits that interrupt the regular flow of speech, such as pauses, stutters, hesitations, and using filler words like "um" and "er."
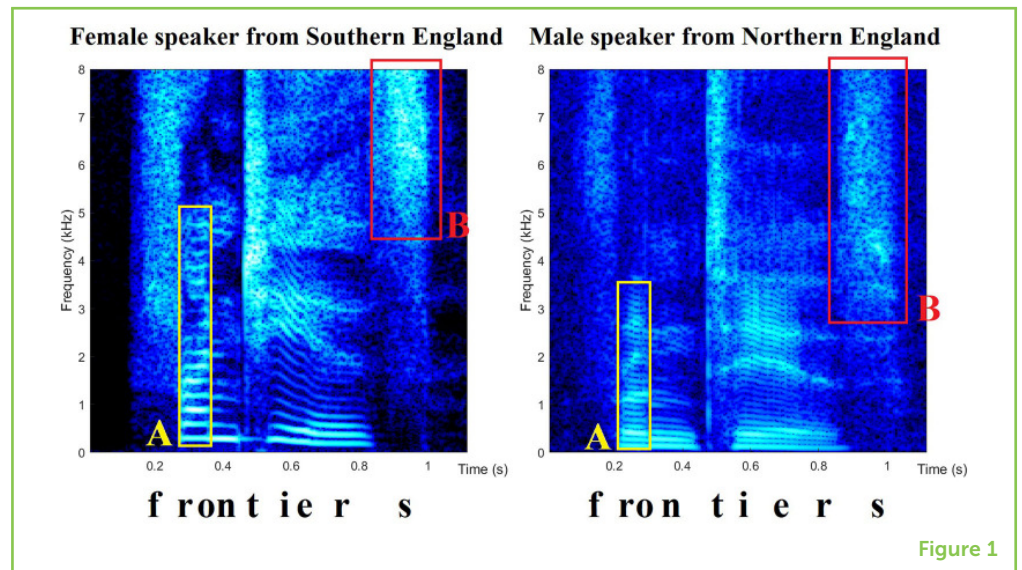
Linguistic features that are commonly analyzed include:

Voice quality: The overall sound of someone's voice. Some people's voices sound rough or harsh, some sound creaky or croaky, and some sound like they have a cold. Voice quality is assessed and compared across a range of categories like this.

Pitch: How high- or low-pitched a voice sounds. Pitch changes due to differences in the shape and size of the vocal cords, but also from the way speakers use their vocal cords, for example to express emotion when they speak.

Vowel and consonant sounds: The pronunciation of specific vowels and consonants can be analyzed to show the speaker's behavior, which might help to define the accent a speaker has. There may also be additional unusual features, like a lisp, or producing "r" to sound like "w."

Timing and rhythm: How slow or fast a person speaks, and how they place stress on different syllables and words.

Level of fluency: When speakers pause, hesitate, interrupt, or repeat themselves, these are called **disfluencies**. Some speakers use lots of disfluencies and others use very few; they may also use distinct types of disfluencies.

Wording and grammar: Speakers might repeatedly use distinctive words or phrases, or they might use unusual sentences in a similar way.

Some linguistic features will be related to the speaker's accent, and some will be more individual. By analyzing a range of features, a full

"profile" of each speaker's voice can be generated and compared with other profiles. The forensic analyst can then produce a conclusion regarding how strong the evidence is that the criminal and the suspect are the same person.

AUTOMATED COMPUTER ANALYSIS

Computers can also be used to compare voices automatically, using a process called **automatic speaker recognition** (**ASR**). This is similar to technology used by banks, in which you can use your voice as a password. Rather than analyzing the linguistic aspects of speech, the computer analyzes the voice recording as a whole, identifying the biological as well as behavioral characteristics of a speaker. The computer software makes thousands or millions of measurements from each voice recording and uses complicated statistical models to turn these measurements into a speaker profile. The software also analyzes the profiles of hundreds of other speakers so it can determine whether the profiles are distinctive or not. So, in our case, a sample of speech from a threatening phone call would be analyzed and compared with the police interview. This type of analysis can be used on its own but, in most cases, it is used alongside linguistic experts [4]. One of the benefits of using a computer is that it can analyze hundreds or even thousands of samples, because it works much faster than a human could.

CONCLUSION

There are lots of differences between people's voices, and some differences are biological while others are behavioral. By comparing many voice characteristics, especially those that are known to vary a lot between individuals, experts can decide how likely it is that two voice recordings come from the same person. This type of voice comparison evidence can be provided for the police or for a court. This evidence might help law enforcement decide whether they have the right suspect, or used in a trial to help decide if a person is guilty of a crime. Now that almost everyone has a smartphone that can easily record audio and video, voice evidence is becoming more common and can be critical in solving crimes.

ACKNOWLEDGMENTS

REFERENCES

1. Jessen, M. 2008. Forensic phonetics. *Lang. Linguist. Compass*. 2:671–711. doi: 10.1111/j.1749-818X.2008.00066.x
2. Lieberman, P., and Blumstein, S. E. 1988. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge: Cambridge University Press.
3. Foulkes, P., and French, P. 2012. "Forensic speaker comparison: a linguistic–acoustic perspective," in *The Oxford Handbook of Language and Law*, eds L. M. Solan, and P. M. Tiersma (Oxford: Oxford University Press). p. 418–21.
4. Parliamentary Office of Science and Technology. 2015. *Forensic Language Analysis: POSTnote*. Parliamentary Office of Science and Technology. Available online at: https://post.parliament.uk/research-briefings/post-pn-0509/

YOUNG REVIEWERS

RYAN, AGE: 15
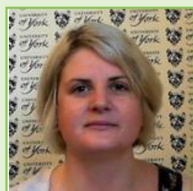I really enjoy coding and I love Rubik's cubes. I also really love playing Minecraft.
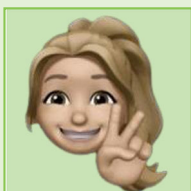
WILLOW, AGE: 13
My name is Willow (named after the character from Buffy the vampire slayer), and I am an 8th grader at a local middle school. I love playing soccer and kicking

around a ball with my dog, Suki. In my free time, I enjoy baking and hanging out with friends.

AUTHORS

AMELIA GULLY

Dr. Amelia Gully is a British Academy Postdoctoral Fellow in the Department of Language and Linguistic Science, University of York. Her work focuses on the acoustics of the voice, and particularly how differences in people's vocal physiology affect the sound of their voices. She has a B.Sc., in audio technology, a M.Sc., in digital signal processing, and a Ph.D., in electronic engineering. She is interested in all aspects of the human voice, from production to perception. *amelia.gully@york.ac.uk

PHILIP HARRISON

Dr. Philip Harrison is a lecturer in the Department of Language and Linguistic Science at the University of York, where he teaches, supervises, and conducts research in forensic speech science. He worked as a forensic scientist for over 20 years at J P French Associates, an independent forensic speech and acoustics laboratory in York, UK, where he worked since 1997.

VINCENT HUGHES

Dr. Vincent Hughes is a lecturer in forensic speech science in the Department of Language and Linguistic Science at the University of York. His research interests lie in forensic speech science, as well as more general questions about how speech sounds are produced and vary among the population. He teaches at both the undergraduate and postgraduate levels and supervises student projects.

RICHARD RHODES

Dr. Richard Rhodes is a senior forensic consultant at J P French Associates, an independent forensic speech and acoustics laboratory in York, UK. He acts as an expert witness in (mainly) criminal proceedings in court and produces reports for voice comparison, transcription, and other types of forensic speech analysis casework. He is also an associate lecturer in the Department of Language and Linguistic Science at the University of York and teaches forensic speech science.

JESSICA WORMALD

Dr. Jessica Wormald is a forensic consultant at J P French Associates, an independent forensic speech and acoustics laboratory in York, UK. She acts as an expert witness and works primarily on voice comparison and transcription cases. She has carried out research into new and developing accents in different locations in the UK, using used forensic methods as well as ultrasound and automatic methods. She is also a courtesy associate lecturer in the Department of Language and Linguistic Science at the University of York.