Elliott Land discusses his study of naïve speaker identification

'I'd know that laugh anywhere'

"I can recognise your father's laugh. It's very distinctive", my mother said to me one day when I was telling her about my research into speaker identification using laughter. One of the reasons behind the research in the first place was my experience of hearing numerous people's similar claims that they had the ability to identify someone they knew well just from hearing that individual's laughter. The ability people are describing in such cases is known as 'naïve speaker identification'. 'Naïve' refers to the fact that the listeners are not trained experts in phonetics and 'speaker' is used regardless of whether the identified person is speaking or producing some other vocalisation like laughter or coughing.

While a fairly substantial amount of research suggests that naïve listeners perform speaker identification using speech to varying degrees of success, very little research has focused on naïve speaker identification using laughter. I also wondered whether the wealth of anecdotal evidence that I had heard would stand up to empirical testing. Investigating this would not just serve as an interesting test of the validity of people's intuitions about their speaker identification abilities, however. As I discuss at the end of this article, if people can recognise others on the basis of their laughter, this suggests laughter may be unique to individuals or, in other words, 'speaker-specific', which may have implications for Forensic Phonetics (introduced in Babel No1).

Method

To investigate naïve speaker identification using laughter, I had members of a close social network (i.e. a group of socially-close people) take part in a speaker identification task, which involved them listening to samples of each other's laughter and attempting to identify the speaker in each sample. The first step was to identify and recruit this close social network. While there are various criteria that can be used to define social networks, constraints on time and scope meant that I used some relatively simple criteria. I identified a group of seven female undergraduate university students on the basis that they had known each other for at least two years prior to the study and had spent several hours of academic and social time together every week. Another two female undergraduates who were not known to the members of the social network were also recruited to serve as so-called 'foils', whose presence would add an extra level of difficulty to the speaker identification task.

The next step was to elicit laughter from these nine speakers and thereby construct samples for the identification task. While elicitation is not the only possible means of gathering laughter (or in fact any) data, other methods were less favourable for the requirements of this study. For instance, the speakers could have been instructed to laugh on cue, but this would likely have caused unnatural laughter to be produced. The elicitation material consisted of several scenes from comedy films compiled into one video clip. Each speaker was audio recorded alone in a recording booth whilst watching this material. Though the choice to record them alone in an unfamiliar environment may have had its own shortcomings, like making the

participants feel uncomfortable, recording the network members together could have caused them to suspect a link between this elicitation session's participants and the speakers they would later be tasked with identifying. The listeners could have then made deductions on the basis of such prior suspicions rather than identifying the speakers on the basis of the laughter itself. (Of course, being a close group of friends, they may have discussed one another's involvement in the elicitation session, and therefore probably had some ideas of the speaker's identities anyway!).

Once the laughter had been elicited, the next steps were to (i) listen to each audio recording, (ii) identify and mark the points where laughter occurred to facilitate later construction of the samples, and (iii) note some phonetic properties of the laughter. The first two steps were achieved using audio software known as Praat, developed by Paul Boersma and David Weenink. Praat has a 'TextGrid' function that allows users to segment, and write text under, certain points in a waveform and spectrogram.

To complete step three, classifying the phonetic properties of the laughter, some descriptive terminology was needed. The terminology came from an extensive literature survey conducted in 2016 by Sarah Cosentino and colleagues, which considered laughter to be made up of individual 'calls', which form 'bouts', which in turn form 'episodes'. It can be useful (albeit a slight oversimplification) to explain this terminology using orthographic representations of laughter. One laughter call might be written as 'ha' or 'he' and two calls as 'haha' or 'hehe'. While this represents 'voiced' calls reasonably well,

Figure 1. Example of a TextGrid used to segment a six-call, one-bout, voiced episode of laughter in the waveform (above) and spectrogram (below): The vocalic portion of each call is clearly visible, being composed of several vertical striations that form a thick and dark vertical line down the length of the spectrogram. Aspiration occurs before each call and is most clearly visible before the first call of this example.

in that it suggests there is some aspiration followed by a vocalic segment wherein the vocal folds are vibrating, it misrepresents 'voiceless' calls, which mostly consist of aspiration and no vocalic segment with vocal fold vibration. One or more calls make up a bout, and one or more bouts make up an episode, meaning that, for instance, a two-bout laughter episode in which a one-call bout is followed by a three-call bout could be written as 'ha hahaha'. Applying this descriptive terminology to my data, each episode of laughter was classified perceptually as being either a voiced or unvoiced episode depending on whether there was a majority of voiced or unvoiced calls. If an episode did not appear to have a clear majority of either type of call, it was classified as a 'mixed' episode of laughter.

Having classified the laughter, it became apparent that some speakers had laughed a lot and some had laughed hardly at all, with the net laughter ranging from approximately 4 seconds to 60 seconds. Presenting listeners with differently-sized samples from each speaker could have reduced the validity of the results, as the listeners would have had different amounts of information about the speakers' laughter with which to make their identifications. To create similar-sized samples, the episodes of elicited laughter were randomly sampled to be representative of what each speaker would produce in terms of voiced, unvoiced and mixed laughter within 4 seconds.

Finally came the point at which the samples could be used to create the speaker identification task. To do this, the nine speaker samples were arranged into seven randomly-ordered sets, one set for each of the seven listeners. Each speaker sample appeared only once in each set, though the listeners were told at the time of testing that they may hear the same

sample more than once. The listeners were not told who the speakers might be, but were told that they may or may not know the speakers. If the listeners were given any prior indications of who the speakers might be, they could then deduce the speakers' identities rather than rely solely on the information in the samples to identify them.

Results and discussion

The listeners' answers were classed as correct if a network member speaker was correctly identified or a foil speaker was correctly rejected. Answers were classed as incorrect if the listeners mistook one network member for another or mistook a network member for someone outside the identified network. When no answer was provided, this was classed as an instance of the speaker being left unidentified.

The listeners' results overall were poor. One network member's answers were excluded as they appeared to have misunderstood the task, leaving the results of six listeners' answers to nine speakers. Each listener identified only one speaker from the network correctly, resulting in a 14% correct identification rate. In terms of the speakers themselves, the total six correct identifications were unevenly distributed across the seven network members' speaker samples. One speaker was identified three times, three speakers were identified once each, and three speakers were never identified at all. Five out of seven speakers were misidentified at least twice. Additionally, both of the foils were mistaken for members of the network and were never correctly rejected as unknown speakers.

Our initial conclusion on the basis of these results might be that listeners do not seem to be able to easily perform naïve speaker identification using samples of laughter. However, things may actually be more complex than this, with identification depending on the samples themselves. Before exploring this complexity, it is interesting to see just how poor their performance was by comparing these results to those of a similar study conducted in 2000 by Paul Foulkes and Anthony Barron, which investigated naïve speaker identification of speech in a close social network of male university students. Their study consisted of ten speakers and nine listeners. All except one listener made at least six out of ten correct identifications. In addition, misidentifications occurred far less frequently than in my study, with only two out of ten speakers being misidentified at least twice. Foulkes and Barron largely attribute the successful (as well as the unsuccessful) identifications to individual speakers' fundamental frequency (or F0), which is the acoustic correlate of what listeners perceive as pitch. Listeners most frequently identified those speakers whose F0 range and mean values were most extreme, whilst those speakers whose values were less extreme were more frequently unidentified or misidentified for one another. I will return to the significance of their explanation in terms of my results shortly.

The introduction to this article may have given the impression that no previous research has examined naïve speaker identification of laughter. However, there exists one notable study conducted by Axelle Philippon and colleagues in 2013, which exposed listeners to a sample of voiced laughter and then, after a short break and a filler task, presented them with laughter samples from several foil speakers and the previously-heard speaker. The listeners in that study were more successful in identifying the target speaker than this one, achieving a correct identification rate of 25%. Some methodological differences between the respective studies' tasks may account for this difference in results. For example, the listeners in the previous study knew they had to identify only one speaker from a series of samples, whereas the listeners in my study were not told how many of the speakers were known to them. However, a more compelling account of the results might be found in the fact that there are quantitative and qualitative differences in the respective studies' laughter samples. Whereas the present study used four-second samples of voiced, unvoiced and mixed laughter, Philippon and colleagues used much longer fifteen-second samples of only voiced laughter. This may suggest that voiced laughter but not voiceless laughter facilitates naïve speaker identification. In other words, rather than concluding based on my results that listeners do not perform well when identifying speakers using laughter, it may be more accurate to suggest that listeners do not perform well when faced with voiceless laughter or small amounts of voiced laughter.

This suggestion is somewhat supported by comparing the identifications of speakers in this study with the proportions of voiced laughter in their samples. The speaker who was identified three times had by far the largest proportion of voiced laughter and the speaker who produced

no voiced laughter was never correctly identified. However, the middle of the range does not show as clear a pattern, with speakers who had smaller proportions of laughter being identified more than those with larger proportions and vice versa. This unclear pattern may suggest that simply presenting listeners with large amounts of voiced laughter will not necessarily result in higher identification rates. Therefore, certain properties of voiced laughter might influence the success of identification. I noted above that one previous study suggested that fundamental frequency is an influential property in naïve speaker identification using speech. Perhaps, then, the perceived pitch of a speaker's laughter, including the pitch variability and range, may influence the rate at which that speaker is identified. It was not possible to investigate this possibility during my study, but fundamental frequency is one feature of voiced laughter that is worthy of investigation in future work.

What, then, should we make of the mismatch between what people, in my experience at least, often claim about their abilities to identify people using laughter and the results of this study? Well, it is important to recognise the potential for numerous differences between the idealised, controlled conditions and subjects used in this experiment and the highly variable conditions of everyday situations upon which people might base their intuitions. For instance, listeners might find themselves identifying people to whom they are much closer (and thus more familiar with) than university friends, such as people with whom they have lived since birth. One notable aspect of this study's experimental condition is the attempt I made to limit the amount of contextual knowledge that the listeners had at the point of testing. Speaking again purely from my own experience, it seems to me that in most of the anecdotes I hear, people overlook the influence that their own knowledge of the context of the situation might have had on the success of their identifications. For instance, to return to the case of my mother, brilliant though she is, she went on to tell me that she once recognised that it was my father and not myself who was laughing in another room of our house. In that particular instance, it seems likely to me that she, however (un)consciously, may have relied on her knowledge of the members of our household to deduce that only my father and I were possible candidates for the identity of the person laughing. Such a deduction could have helped her to, at the very least, narrow down the possibilities. Further testing would be required to see if the influence of context is indeed so strong, but at present it seems to be one of the most credible explanations for the difference between listeners' intuitions and the results of this experiment.

To conclude, I would like to note some of the real-world applications of the results of this study. As I suggested at the start of this article, the results may have implications for Forensic Phonetics, an application of phonetics that deals with issues of speaker identity. One area of research in Forensic Phonetics is concerned with investigating potentially speaker-specific properties of speech and non-linguistic behaviours like laughter for use in forensic speaker comparisons, which involve the comparison of samples from unknown speakers and known suspects. As naïve listeners seem to be able to identify speakers based on samples of voiced laughter, it is possible that voiced laughter is therefore speaker-specific. Future research could therefore focus on understanding, among other things, which properties of voiced laughter, such as fundamental frequency, might contribute to its speaker-specificity. Such research would be essential before laughter could be properly utilised in forensic speaker comparisons. Laughter, however, may eventually prove to be a useful addition to the collection of features analysed in forensic phonetic casework. ¶

Elliott Land is a former student of the University of Huddersfield, where he studied a BA (Hons) in English Language and Linguistics. He is currently studying for an MSc in Forensic Speech Science at the University of York.

Find out more

Online

Paul Boersma and David Weenink (2017) Praat – software available at fon.hum.uva.nl/praat/r.

Articles

Sarah Cosentino, Salvatore Sessa and Atsuo Takanishi (2016) 'Quantitative laughter detection, measurement, and classification – A critical survey', in *IEEE Reviews in Biomedical Engineering* 9(1).

Paul Foulkes and Anthony Barron (2000) 'Telephone speaker recognition amongst members of a close social network', in *The International Journal of Speech, Language and the Law* 7(2).

Axelle C. Philippon, Liane M. Randall and Julie Cherryman (2013) 'The impact of laughter in earwitness identification performance', in *Psychiatry, Psychology and Law*, 20(6).