



THE UNIVERSITY OF
MELBOURNE

A preliminary investigation of the acoustic factors impacting decision-making in speaker attribution

Debbie Loakes¹, Helen Fraser¹ and Kirsty McDougall²



¹ Research Hub for Language in Forensic Evidence, The University of Melbourne; ² University of Cambridge

Context

Forensic Transcription Criminal cases can rely on the transcript of an indistinct conversation as evidence, e.g. covert recordings made by police (discussing illegal activity).

Transcripts are often made by police officers, and these are allowed in Australian courts, but this has reliability problems (Fraser & Loakes 2020).

A Call To Action from Australian linguists to the Australian judiciary asked for a review and reform of the handling of indistinct audio.

The Research Hub for Language in Forensic Evidence (Unimelb) was established to deal with this (Fraser 2020).

Speaker attribution?

The task of assigning an utterance to an individual speaker. High stakes – attributing an incriminating utterance to an individual aligns them with having committed a crime. This is under-researched, but some experimental studies exist:

- Love & Wright (2021): “closed-set” task with good quality audio – professional transcribers. Only 40% of turns were attributed to speakers with better-than chance agreement.
- Lublinskaja & Sappok (1996): Highly controlled conditions, transcribers regularly split one voice into two, or merged two voices.

Research shows that aptitude, familiarity, “distinctiveness” of a voice all contribute. Listeners seem to rely on a range of acoustic cues.

Aim

This study contributes to a longstanding question of how to ensure reliable attribution of individual utterances to specific speakers (Fraser 2018).

Research questions:

- 1) How well is speaker attribution done by listeners who have phonetic training?
- 2) What acoustic characteristics are important in listener decision-making?

Method

A 97 second audio clip from a television show (*The Panel*) was used as the stimuli. 6 speakers in conversation (4 M, 2 F). Spontaneous speech. **ASK ME AND I CAN PLAY IT FOR YOU!**

Appropriately challenging: multiple voices, some overlap and background noise. Speakers stationary, audio clear (broadcast).

Participants: 5 phonetically trained transcribers. Minimal instructions. Asked to provide a transcription, but focus primarily on speaker attribution. “Open-set”.

Feedback / debriefing session two weeks after task helped drive phonetic analysis:
Focus on **male speakers, F0 and intonation.**

Results

There were 6 speakers in the recording, transcribers identified between 5-7 speakers. Only one transcriber heard 6 speakers, but not reliably. This was the only listener with any familiarity.

Female speakers - relatively easy, Male speakers - multiple errors, differences depending on speaker. E.g. in non-overlapping speech, Speaker 1 = 67% correct attributions, Speaker 4 = 20%. Listeners reported various difficulties in the task.

“However long I spent on it. I couldn't do it”.

“[the] male speakers ... are particularly difficult to distinguish and it's probably something to do with their pitch and the way they use intonation. I thought it was just extremely difficult.”

“If I was approached to by a lawyer to try and do some work like this ... this helped me to realize that actually, I'm not able to do it. So this is actually extremely useful for me. Very valuable exercise to appreciate how difficult it is ... I thought I could do it, and I found out I couldn't.”

Results

F0: Majority of measurements fall within a narrow range (e.g. Fig. 1).

Typical distribution for Aus Eng, for this population.

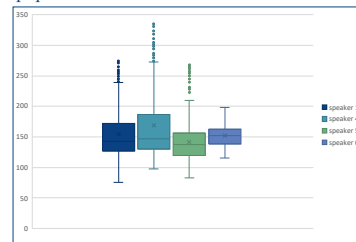


Figure 1. F0 distribution for 4 male speakers of Australian English (broadcast speech)

Outliers? Intonational variation rather than speaker-specific behaviour.

Intonation: Using a ToBI framework, it is clear that three male speakers use a very particular intonation pattern.

L+H* onsets, typical in forward looking dialogue acts and said to “influence addressee future action” (Fletcher et al. 2002), followed by downstepped accents and pitch compression.

Stylistic (TV speech), rather than speaker-specific. Aligns with another study which found intonation was used by listeners to discriminate speakers, but not reliably (Foulkes & Barron 2000).

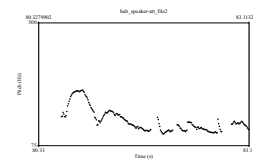


Figure 2. Speaker 1, L+H* onset
“I had something about that night when everyone died ...”

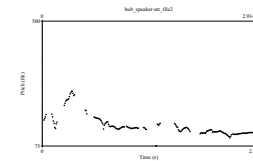


Figure 3. Speaker 4, L+H* onset
“I had more money than sense and I came to you and said I wanna ...”

Results

Summary: For both F0 and intonation, within-speaker variation is *greater* than between-speaker variation. Shared acoustic characteristics cause difficulty for listeners

Discussion and Conclusion

F0 and pitch not acoustically distinctive for this set of male speakers. Listeners use these cues, but not reliably. F0 helped distinguish female speakers (contra. Lublinskaja & Sappok 1996; 24.6% merged M-F pair).

Speaker attribution is regularly carried out in forensic linguistic contexts, but needs a managed process (Research Hub). Further difficulties in forensic context; indistinct audio, within-speaker variation, lack of accessible “ground truth”.

dloakes@unimelb.edu.au

@debbie_dloa

Find me on LinkedIn

Fletcher, J., L. Stirling, I. Mushin & R. Wales (2002) Intonational rises and dialog acts in the Australian English Map Task. *Language and Speech*. 45 (3): 229-253.

Fraser, H. (2018). Forensic Communication in Theory and Practice: A study of discourse analysis and transcription (review) *Language and Law / Linguagem e Direito* 5(1): 103-108.

Fraser, H. (2020). Introducing the Research Hub for Language in Forensic Evidence *Judicial Officers Bulletin* 32 (11): 117-118.

Foulkes, P. & A. Barron (2000) Telephone speaker recognition amongst members of a close social network. *International Journal of Speech, Language and the Law*. 7 (2): 180–198.

Lavan, N. et al. (2019). How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices. *British Journal of Psychology*. 110 (3): 576-592.

Love, R. (2020). *Overcoming challenges in corpus construction: The spoken British National Corpus 2020*. Routledge Advances in Corpus Linguistics. New York: Routledge.

Love, R. & D. Wright (2021) Specifying challenges in transcribing covert recordings: Implications for forensic transcription. *Frontiers in Communication* 6: 1-14. Article ID 797448.

Loakes, D. (2008). A forensic phonetic investigation into the speech patterns of identical and non-identical twins. *International Journal of Speech, Language and the Law* 15 (1) 97-100.

Lublinskaja, V. & C. Sappok (1996) Speaker attribution of successive utterances: The role of discontinuities in voice characteristics and prosody. *Speech communication*. 19, 145-159.