Characterizing related speakers using mean fundamental frequency and vowel formant frequency information

Hanna S. Feiser[1,2]
[1]*Department of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany*
[2]*Bavarian State Criminal Office, Munich, Germany*
feiser@phonetik.uni-muenchen.de

Introduction

Forensic case work can involve identifying voices of male relatives who are not twins. Similar voices could belong to suspects who are, for example, father and son or two brothers. Information regarding speaker identification among family members and related speakers has hardly been investigated. This has already been confirmed in an earlier investigation concerning siblings' voices (Feiser, 2009). As "...research into the acoustic properties contributing to the perceived similarity of voices has been lacking" (Nolan, 2009) and acoustic analysis has been very important for voice comparison in forensic phonetics for quite some time now, the present study will focus on acoustic investigations of family members' voices. In doing so, the mean fundamental frequency and vowel formants F1 and F2 are analyzed.

Methods

The investigation focuses on voice similarities and voice differences among 14 brothers (7 brother pairs). Data from the seven pairs were acquired under forensically realistic conditions by means of read and spontaneous utterances over the telephone and in studio quality. In this way, data can be analyzed by comparing different recordings made under different circumstances, e.g., a telephone conversation of an unknown speaker compared with a police interview or read recording with a known speaker. There were two reading tasks and one spontaneous speech task. Recordings took about one hour per brother pair. The subjects first had to read 80 minimal pairs and then 100 sentences of the so-called "Berliner Sätze". While one brother was reading, the other watched a video sequence of the German TV series "Tatort". When both finished reading they talked about the different video sequences they saw via mobile phone while sitting in two different rooms. This yielded four conditions: reading over microphone (RM), reading over telephone (RT), spontaneous speech over microphone (SM), and spontaneous speech over telephone (ST). Microphone speech and telephone speech were recorded simultaneously. Hereafter, the evaluation of one RM task is described. Minimal pairs from the first reading task contained 20 words with four repetitions in which every word was embedded into one of the following German carrier phrases: "Anna/Timo hat … gesagt." ("Anna/Timo said ..."). The results of this study focus on mean fundamental frequency, which was measured and averaged across the entire utterance, and the first and second formants of long vowels /i:/, /a:/, and /u:/ in the German words "bieten", "baten" and "bucht".

Results and discussion

The results for mean f0 (see fig.1) show that the difference in f0 between related brothers was significant, while the difference between any one speaker and the mean f0 of all unrelated speakers was not. Therefore, this parameter does not seem to reveal the similarity in brothers' voices. However, for the vowel formants (see fig. 2) there are partially quite different results. The Euclidean distances between the mean values of F1 and F2 of four vowel tokens of /i:/, /a:/, and /u:/ were measured separately for each speaker and his brother, and additionally between the means of one speaker and the means of all 12 other unrelated speakers.

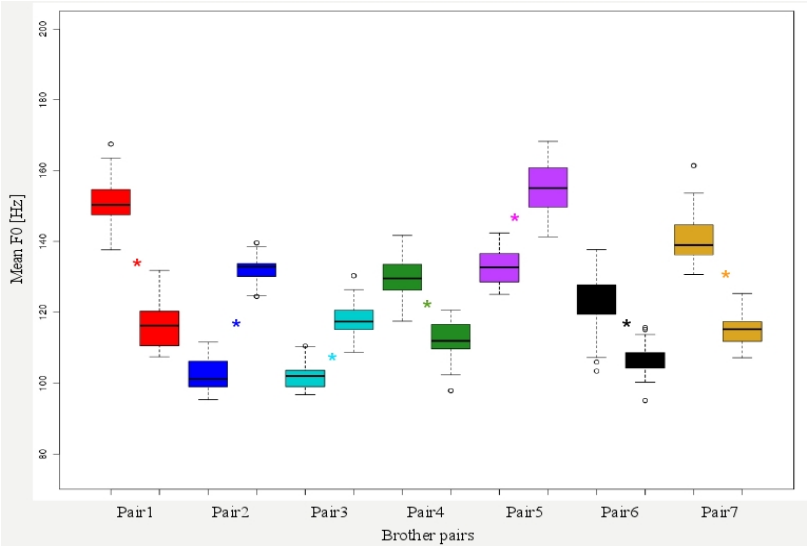

Figure 1 Mean fundamental frequency of each speaker (same color = same family).

The Euclidean distances in fig.2 for the vowel /i:/ in the middle and /u:/ on the right side were not significantly lower for the brothers (in the condition "same"), although there is a tendency in this direction. For /a:/ (left-most pair), the Euclidean distance to the brother (yellow) is significantly lower than the one between unrelated speakers (blue). This might be due to the significantly larger (non-sibling) inter-speaker variation in F1 of /a:/ compared with /i:/ and /u:/, which appear to be slightly more robust against speaker-dependent variation.

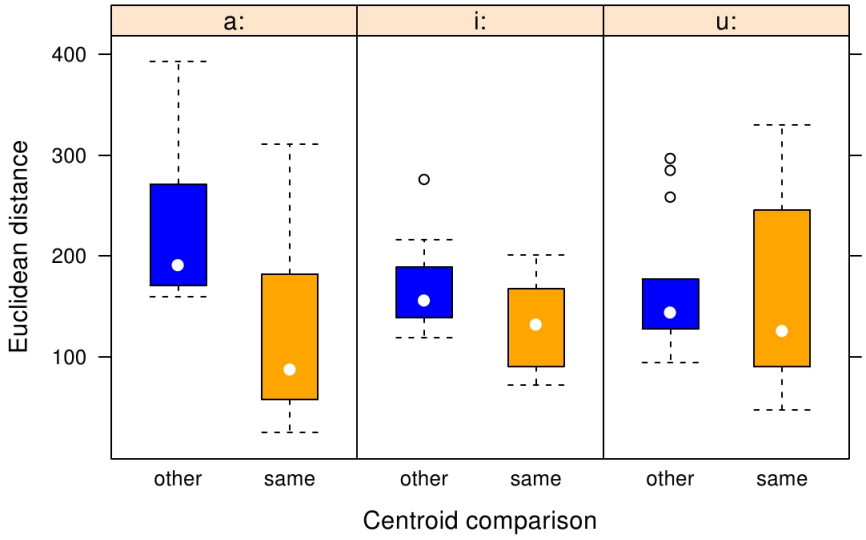

Figure 2 Euclidean distance of F1 and F2 of three vowels (same = comparison to speaker's brother, other = comparison to all 12 unrelated speakers

The next step will be a perception experiment in which about 20 naive listeners judge the similarity of speakers on a scale from "very similar" to "very different". Speakers will be presented to the listeners in random order over telephone and in full-bandwidth quality.

References

Feiser, H.S. (2009). Acoustic similarities and differences in the voices of same-sex siblings. In *Proceedings of 18th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA) 2009*, Cambridge, UK, p. 12.

Nolan, F. et al. (2009). Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence. *Full Research Report, ESRC End of Award Report, RES-000-22-2582*. Swindon: ESRC.