Perceptual voice similarity of related speakers:
telephone and microphone recordings

Hanna S. Feiser, Christoph Draxler
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität,
Munich, Germany
{feiser|draxler}@phonetik.uni-muenchen.de

Introduction

In forensic casework it is necessary to deal with similar male voices on a daily basis. Thereby some voices are more similar than others (e.g. Jessen 2012, Rose 2002). A perception experiment carried out by Feiser (2012) showed that naïve listeners are able to distinguish between related and non-related speakers. The aim of the present perception study is to test whether naïve listeners are able to identify voices of speaker pairs from telephone and microphone recordings and whether voices from related speakers are confused more often.

Methods

Recordings were obtained from ten pairs of brothers between the age of 19 and 31 and all were speakers of central Austro-Bavarian. The speakers read the *German 100 Berlin sentences* in a sound-attenuated booth. Recordings ran over two Nokia mobile phones and two *Neumann TLM 103* microphones at the same time. The perception experiment was conducted using *Percy web-experiments* and took about 20 minutes per subject. A voice identification task with no repetitions was presented to 122 listeners (64 female, 58 male) between the age of 19 and 64. 30 listeners participated in the phonetic lab and the remaining 92 attended the experiment online from different locations. Ten speaker pairs were presented to the listeners in ten separate blocks. Firstly, every block contained a training period where four stimuli (two from each speaker) were represented and the listener had to memorize a coloured symbol for each of the two speakers. Secondly, in the test period 16 stimuli were represented – one at a time from each speaker. Afterwards, listeners had to decide which speaker spoke each stimulus by clicking on the matching coloured symbol.

Results

Correct identification of the twenty speakers by their voices was significantly above chance. Listeners correctly identified the speakers in 88% of all instances (for details see Figure 1). A general linear mixed model with correct identification as the dependent variable, relation and recording type as independent variables plus listener and speaker as random factors showed that stimuli over microphone were identified significantly better than stimuli over telephone. Additionally, the number of false identifications was greater for brother pairs than for pairs with two unrelated speakers.

Discussion and conclusion

The findings clearly show that naïve listeners are able to identify speakers after a short familiarization. Results indicate that siblings' voices and voices over telephone are more often confused. Therefore, it seems that the voices of family members are perceptually more similar than those of unrelated speakers (e.g. Nolan 2009). In the present study listeners had the opportunity to use all acoustic information available in the speech signal. Results suggest that when listeners could not rely on dialect features (speakers came from the same dialect area), they had to use different features. Previous acoustic analysis of mean F0, vowel formants and articulation rate of the same 20 speakers revealed that those features seemed not to be responsible for the perceptual similarity. This raises the question of what is responsible for the similarity.
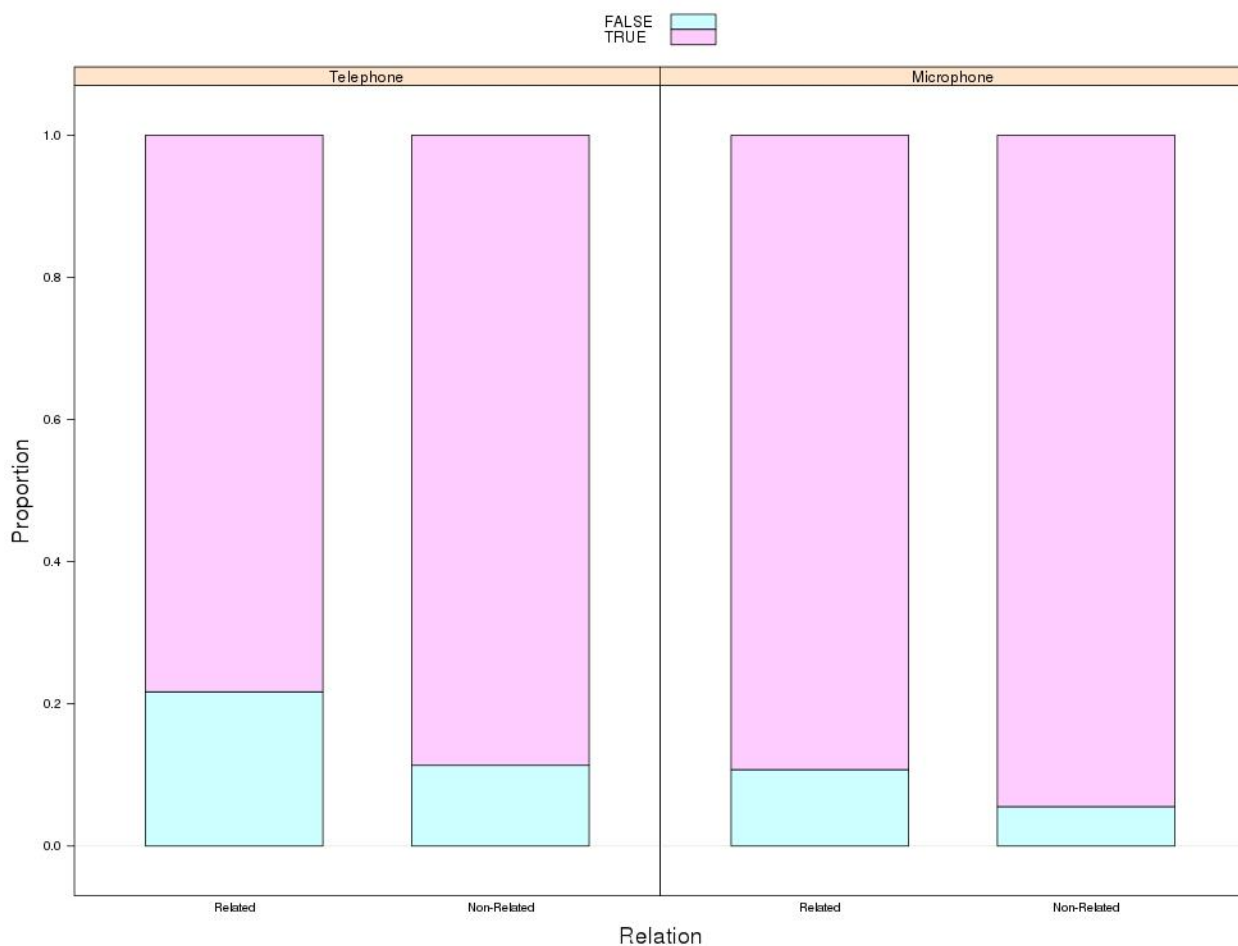
Figure 1 Proportion of true (pink) and false (turquoise) identifications separately for recording type (left box: telephone, right box: microphone) and separately for relation in each box (left: related speaker pairs, right: non-related speaker pairs).

References

Feiser, H.S. & Kleber, F. (2012). Voice similarity among brothers: evidence from a perception experiment. In Proc. 21st Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA) 2012, Santander, ES.

Jessen, M. (2012). Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs. Munich: LINCOM Studies in Phonetics.

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.

Percy Web-experiments: http://webapp.phonetik.uni-muenchen.de/WebExperiment/

Rose, P. (2002). *Forensic speaker identification.* London: Taylor & Francis.