

Effects of telephone transmission on the validity of formant-trajectory-based forensic-voice-comparison systems

Cuiling Zhang^{1,2}, Ewald Enzinger^{2*}, Geoffrey Stewart Morrison², Felipe Ochoa²

¹Department of Forensic Science & Technology, China Criminal Police University, Shenyang, China

²Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney, Australia

*e.enzinger@student.unsw.edu.au

In forensic-voice-comparison casework a common scenario is that the suspect's voice is recorded directly using a microphone in an interview room but the offender's voice is recorded via a telephone system. Acoustic-phonetic approaches to forensic voice comparison often include analysis of vowel formants, and the second formant is often assumed to be relatively robust to telephone-transmission effects.

This presentation reports on a study which assessed the effects of telephone transmission on the performance of formant-trajectory-based forensic-voice-comparison systems. The effectiveness of both human-supervised and fully-automatic formant tracking were investigated. Human-supervised formant tracking is generally considered to be more accurate and reliable but requires a substantial investment of human labour.

Measurements were made of the formant trajectories of /iau/ tokens in a database of recordings of 60 female speakers of Chinese using one human-supervised and five fully-automatic formant trackers. Measurements were made under high quality, landline-to-landline, mobile-to-mobile, and mobile-to-landline conditions. Telephone-transmission degraded recordings were made by passing the original high-quality recordings through telephone systems, training and test data were otherwise exactly the same. High-quality recordings were treated as suspect samples and telephone-transmitted recordings as offender samples (see Table 1).

Table 1: Telephone-transmission conditions tested.

suspect condition	offender condition
high-quality	high-quality
high-quality	landline-to-landline
high-quality	mobile-to-mobile
high-quality	mobile-to-landline

Discrete cosine transforms (DCT) were fitted to the formant trajectories and likelihood ratios were calculated on the basis of the DCT coefficients. For each telephone-transmission condition the formant-trajectory system was fused with a baseline mel-frequency cepstral-coefficient (MFCC) system, and performance was assessed relative to the baseline system.

System performance was assessed using C_{lr} and Tippett plots.

The human-supervised systems always outperformed the fully-automatic formant-measurement systems, but in conditions involving a mobile telephone the addition of human-supervised formant-trajectory measurements resulted in little or no meaningful improvement over the baseline system. Even in the other conditions it is debatable whether the cost of human-supervised formant-trajectory measurement is warranted.

This presentation is based on Zhang et al. (2012).

Reference

Zhang, C., Morrison, G. S., Enzinger, E., & Ochoa, F. (2012). *Effects of telephone transmission on the performance of formant-trajectory-based forensic voice comparison – female voices*. Manuscript submitted for publication.