Simulating casework recording conditions for forensic voice comparison

Felipe Ochoa

Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunication,
University of New South Wales, UNSW Sydney, NSW 2052, Australia
f.ochoa@student.unsw.edu.au

In forensic-voice-comparison, recordings of caseworks are affected by distortion of phone transmission, background noise, reverberation and other factors which degrade the quality of their signals. In order to estimate the validity and reliability of a forensic-voice-comparison system under specific circumstances, the recordings used for construct the background, development and test databases should reflect as close as possible the real conditions. This presentation explains three procedures employed to simulate such conditions from a database of high quality recordings.

The first procedure describes a software [1] which simulate the acoustics of a room. This program estimates the impulse response of a simple parallelepiped room based on its dimensions, relative position of source and receptor, and acoustic absorption of its walls. The practical implementation of this procedure is presented as follow:

- Initially, if the room where the original recording was taken is known (likely in the case of the suspect recording), its dimensions and position of both, suspect speaker and recording microphone can be measured. The construction materials of walls, ceiling and floor will determine the acoustic absorption characteristics of the room (from tables). Other less important parameters of the software such as humidity (in %) and temperature (in Celsius) can be set as typical values (70% and 24 ^{o}C).
- Based on those parameters, the program calculates the room impulse response which defines the reverberant environment.
- This impulse response is then convolved with the database recordings to incorporate reverberation effect to the speech.
- Finally, the recordings with synthetic reverberation are evaluated using other software [2] to estimate the reverberation time (T60) and comparing it against the T60 of the original suspect/offender recording.

The second procedure explains the addition of background noise to recordings of the database. The workflow can be described as follows:

- The procedure starts with the estimation of the signal to noise ratio (SNR) of the actual suspect/offender recording.
- Then, noise is extracted from non-speech portions of the actual suspect/offender recordings.
- This noise is then added to the high quality recordings from the database. The high quality signal is scaled in an attempt to match the original SNR.
- The final recording is a speaker speech from the database with additive noise taken from the actual casework recording, and with a SNR as close as the actual recording.

The third procedure described is the process of degradation of recordings due to phone transmission. For instance, if the original recording is a conversation over a phone, then the high quality recordings should be degraded using either simulation algorithms [3] to emulate signal compression and filtering, or by passing the database speakers through real phones transmissions and record the output directly from the phone device. The latter procedure is described in this presentation. The steps to simulate this effect are summarised below:

- The high quality recordings are played using a computer and soundcard through a flat-response loudspeaker.
- A phone device is located in the vicinity of the loudspeaker. This phone make a call to other device (receiver phone) which is connected to a recording system (in this case the same computer with an external soundcard)
- The degradation of the recording because of phone transmission (GSM compression, landline filtering, switch station, etc.) is registered in the receiver phone and recorded.
- Due to a delay of the transmission process, a correlation between the original and the degraded recordings is performed to align the time of the latter.
- Notice that a software has been implemented to automatize those steps.

References

[1] D. Campbell, ROOMSIM toolbox [online], http://media.paisley.ac.uk/~campbell/Roomsim (last viewed October 2012)

[2] H. W. Lollmann et al (2010),"An improved algorithm for blind reverberation time estimation", *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp 1-4.

[3] B. J. Guillemin, C. Watson (2008),"Impact of GSM mobile phone network on the speech signal: some preliminary findings", *Journal of Speech, Language and the Law*, Vol. 15.2, pp 192-218.