Effect of the Double-Filtering effect on Automatic Voice Comparison

Jonas Lindh[1], *and Joel Åkesson*[1]
[1]*Voxalys AB, Gothenburg, Sweden.*
{jonas|joel}@voxalys.se

In forensic casework today it is not uncommon to receive material recorded with mobile phones or other handheld recording devices. From experience we know most people do not treat recordings with as much care as a person well versed in audio technology. Especially given the varying circumstances under which the material can be recorded. Thus it is important we learn more about what sort of acoustic effects take place under particular conditions and how these effects can influence Automatic Voice Comparison (AVC). The current study aims at evaluating the effects of recording material consisting of what could be described as 'double-filtered' sound, henceforth referred to as DF, e.g. when a phone call is recorded using a handheld recorder placed in the vicinity of the mobile device. This filtering effect constitutes sound transmitted via GSM communication (1st filter) which then passes an indeterminable distance through the air before being captured by another recording device, such as a mobile phone or handheld recorder's microphone (2nd filter). This effect affects the energy in the signal. The energy decreases in both the low and the high frequencies, while the middle frequencies are boosted.

In this study we have used a database consisting of 150 female speakers of Swedish, all students of speech and language pathology. The recordings were made in a sound treated recording booth using a set-up of one computer equipped with an internal M-Audio soundcard and a high quality headset microphone. Each recording consists of solicited spontaneous speech together with read speech material (Swedish standard reading passage called 'Ett svårt fall'). Each speaker is informed and encouraged to finish the task at their own pace. Mean duration of the full recording among the speakers was 69.3 seconds (std 16 seconds).

Figure 1 Re-recording with double filtering in studio.

The DF effects have been evaluated using two AVC systems applying two different techniques, Batvox 4.1, (developed by Agnitio), a so called iVector system (Dehak et al.,

2009) and Vocalise (Oxford Wave Research) applying the so called UBM-GMM approach (Reynolds, 1992). Each recording in the database was split so that the read passage could be used as training material, while the spontaneous passage would be used for testing. For Batvox 100 speakers were used for testing, 50 speakers for score normalisation (30 speakers for T-norm and 20 speakers for Z-norm) (Barras and Gauvain, 2003). For Vocalise the same 100 speakers were used for testing and 50 speakers for the UBM.

The results show that normalisation techniques decreases the effect of the double filter.

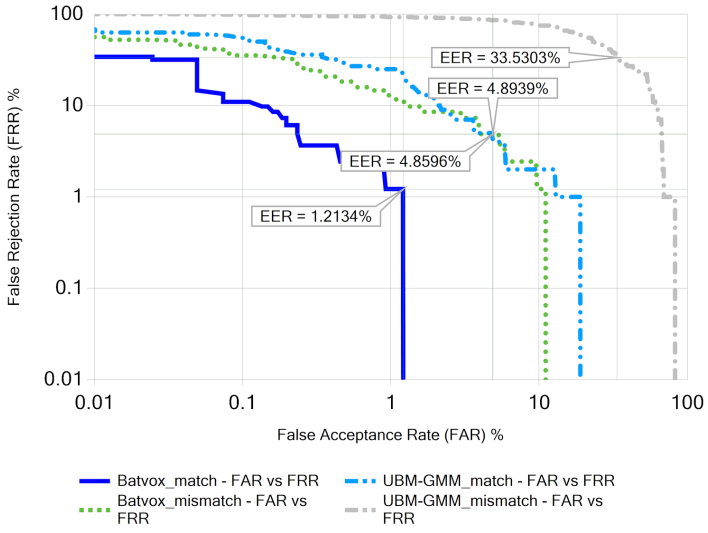

Figure 2 Log EER from the test results for both systems with both mismatched and matched training and test recordings.

In the next phase an error-check will be made to see whether the same mistakes are made by the two systems and between conditions. After that the material will be double-filtered using different recording distances to see how that affects the result.

References

Barras, C. and Gauvain, J.L. Feature and score normalization for speaker verification of cellular data, *Proc. of ICASSP*, April 2003.

Dehak, N., Dehak, R., Kenny, P., Brummer, N., Ouellet, P. and Dumouchel, P. Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification, in INTERSPEECH, Brighton, UK, Sept 2009.

Reynolds, D. A. A gaussian mixture modeling approach to text-independent speaker identification, Ph.D. thesis, Georgia Institute of Technology, August 1992.