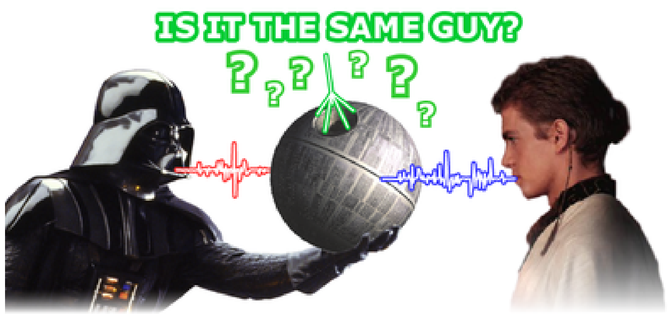
Designing Speaker Verification Systems Robust to Telephone Codecs

Ján Profant <xprofa00@stud.fit.vutbr.cz>

15

Excel@FIT

Task



Speaker Verification System

Compressed audio plays a significant role in mobile communications, Voice Over Internet Protocol (VOIP), archival audio storage, gaming communications and also internet streaming audio. In most of these tasks, there is a widespread use of lossy speech coders.

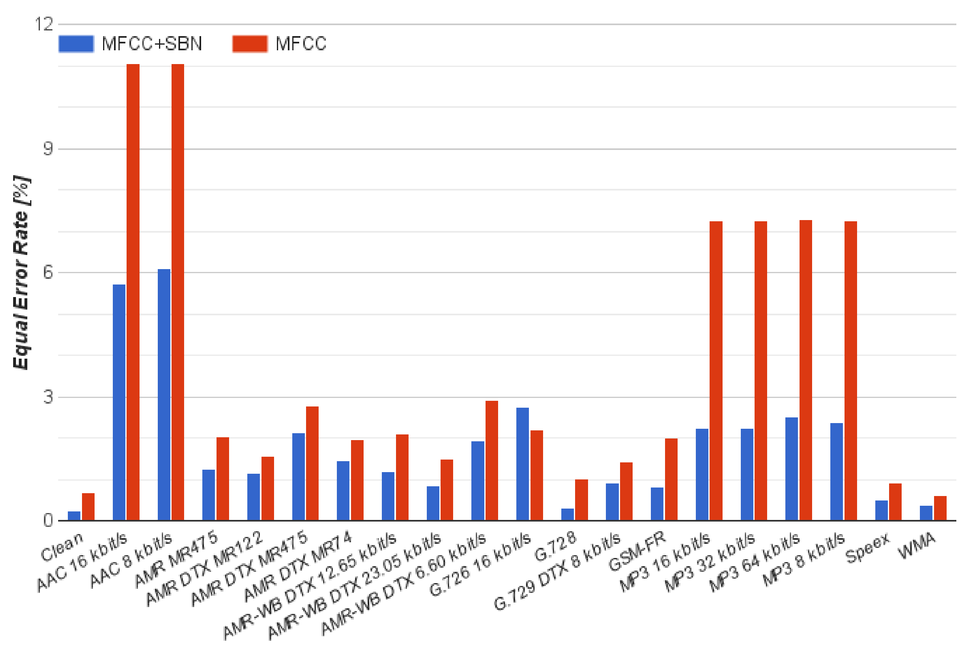
In Speaker Recognition, the distortion introduced by speech coders may have a significant negative impact on the performance. Of interest, therefore, is the analysis of codec-related degradation and the development of robust techniques against this degradation. Ideal speaker verification system should achieve same results on any codec variation.

In our work, we also analyzed the effect of codec distortion on the PLDA compensation module with the state-of-art system developed by *Speech@FIT*. Later, we used a Within-Class Covariance Correction (WCC) technique for Linear Discriminant Analysis (LDA) to analyze impact of codec-degraded speech on speaker verification system, we used WCC to improve performance of our system on codec-degraded speech. Finally we used both techniques together to improve performance of our system on codec-degraded speech.

We compare scenarios where PLDA is trained only on clean data, then system where we add also noise and reverberant data, and at last, codec degraded speech. We evaluate the systems on the matched conditions and also mismatched conditions. We can see clear benefit of adding transcoded data to PLDA or WCC (with approximately same gain) for both tested conditions (matched and mismatched).

Baseline System

The first system can be considered as the Baseline. Here the PLDA model was trained only with clean speech data. Results are presented in following Figure in terms of Equal Error Rate (EER) when speech is degraded using each codec.



We can conclude, that several codecs result in significant EER degradation relative to clean conditions. Specifically, EER of AAC codecs is higher on both our systems. Also, *BN+MFCC* performance is much better, as we expected. Average EER on all codecs conditions is ***1.94%*** for *BN+MFCC* and ***3.94%*** for *MFCC* respectively.

PLDA Robust System

Here, we explore the case where the PLDA model is exposed to all available codec-degraded speech along with noisy and reverberate speech.

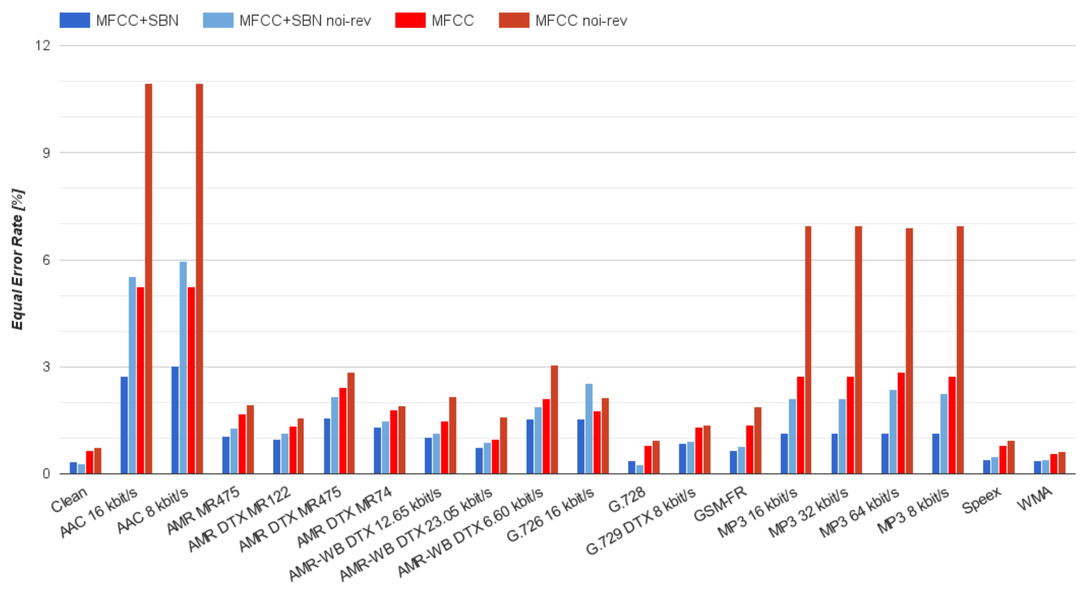


Figure indicates that including speech compressed with the same codec as used for model enrollment and testing in the PLDA training data set significantly lowered EERs. Average EER on all codecs conditions is ***1.19%*** which is ***36.5%*** relative improvement comparing to Baseline 2 for *BN+MFCC* and ***2.11%*** with relative improvement ***44.9%*** for *MFCC* respectively.

Comparing Techniques for Improving Robustness

We compared various techniques for improving performance of our system. We can see that adding different variety of transcoded speech helps also for the unseen codec degraded data. Generally WCC technique yields slightly better results than exposing codec degraded speech data to the PLDA. There is a slight gain if we use both techniques together.

	MFCC			BN+MFCC		
	EER [%]	DCF^{min}_{new}	DCF^{min}_{old}	EER [%]	DCF^{min}_{new}	DCF^{min}_{old}
Baseline 2	3.82	0.3437	0.1365	1.87	0.2496	0.0771
PLDA Unseen Codec	3.78	0.3466	0.1352	1.63	0.2434	0.0710
WCC Unseen Codec	3.73	0.3419	0.1348	1.61	0.2353	0.0695
WCC and PLDA Unseen Codec	3.75	0.3425	0.1333	1.57	0.2425	0.0704
PLDA Using All Codecs	2.11	0.2038	0.0531	1.19	0.2503	0.0833
WCC Using All Codecs	2.16	0.2627	0.0893	1.27	0.2089	0.0563
WCC and PLDA Using All Codecs	2.13	0.2459	0.0828	1.21	0.2038	0.0536

GSM-FR Analysis

We compared the results from the system tuned for the GSM-FR codec and general system adapted to the all codecs.

System Description	EER [%]	DCF^{min}_{new}	DCF^{min}_{old}
Baseline	0.8278	0.1612	0.0324
Baseline 2 + Codecs in PLDA (without GSM-FR)	0.7894	0.1487	0.0360
Baseline 2 + Codecs in PLDA and WCC (without GSM-FR)	0.7802	0.1491	0.0356
Baseline 2 + GSM-FR in WCC	0.7674	0.1450	0.0317
Baseline 2	0.7557	0.1473	0.0309
Baseline 2 + Codecs in WCC (without GSM-FR)	0.7231	0.1355	0.0326
Baseline 2 + Codecs in PLDA	0.6683	0.1527	0.0352
Baseline 2 + Codecs in PLDA and WCC	0.6677	0.1525	0.0343
Baseline 2 + GSM-FR in PLDA and WCC	0.6073	0.1355	0.0326
Baseline 2 + GSM-FR in PLDA	0.6051	0.1484	0.0271
Baseline 2 + Codecs in WCC	0.6044	0.1335	0.0288
System trained with GSM-FR data - Baseline	0.4094	0.1248	0.0215
System trained with GSM-FR data - Baseline 2	0.3709	0.1199	0.0243