

Speaker Comparison Evaluation Using a New Corpus of Urban Speech

Julia Forsberg¹, Johan Gross¹, Jonas Lindh¹² and Joel Åkesson¹²

¹*University of Gothenburg, Sweden*

{julia.forsberg|johan.gross}@gu.se

²*Voxalys AB, Sweden.*

{jonas|joel}@voxalys.se

Urban speech is forensically highly relevant, as most cases deal with these varieties. In Sweden research has been focused on rural dialects (with some exceptions) and their geographical variation (Lindh and Eriksson, 2009; Leinonen, 2010) creating a gap in the knowledge of Swedish spoken in the largest cities.

The project *Språkbruk i Stockholm och Göteborg* (SSG, ‘language use in Stockholm and Gothenburg’) was initiated in 2014, with the purpose of mapping young urban Swedish. The material comprises a total of 111 informants between the ages of 16-19, from 4 schools: one in each city centre, and one in a suburb to each. There was a total of 51 boys and 60 girls, and 63 informants with at least one foreign-born parent. Each informant was recorded on two occasions: in a sociolinguistic interview; and in a map-task (in Swedish and English) together with a peer. This combination allows for a mixture of speech styles (speaking to a researcher vs. speaking to a peer), a time lapse (of varied length) between the recordings, and the use of more than one language from each informant. This enables us to evaluate forensic speaker comparison methods using this urban youth speech data.

To begin with, speech was extracted from the male speakers in the database and silence automatically removed using a dB-threshold. To investigate how well an automatic voice comparison system can differentiate between the voices, the software Vocalise (Jessen et al., 2014), a UBM-GMM (Reynolds et al., 2000) system, was used. Initially, a subset of 20 speakers were selected for the UBM and a test run for all the other voices without any quality check. That test produced an EER of 22.44%. After a quick quality check (always performed for forensic casework) many different recording problems were discovered in the database. For a second test all recordings with electronic noise, environmental noise or too much clipping were removed. Where only minor clipping was discovered, it was edited out. In the cases where there was only one recording, that one was used in the UBM together with those recordings where only one was of suitable quality out of the two. All recordings were then normalised to 75dB using a Praat script and a new test was performed. In the new test 32 recordings ended up in the UBM while only 17 pairs of voices were found suitable for testing. For this test the EER dramatically dropped to 0.8716%.

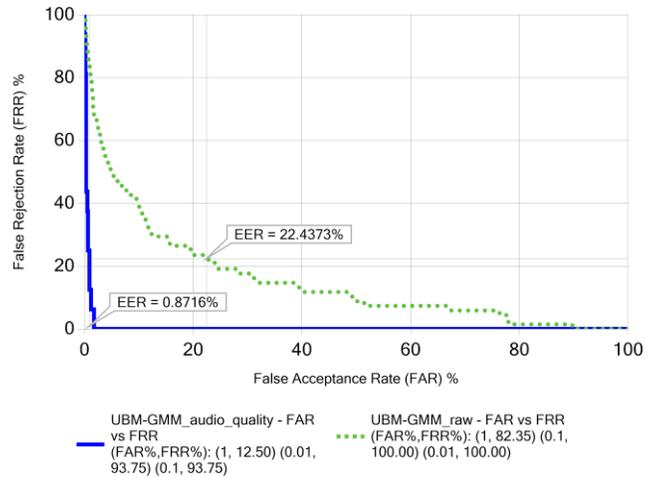


Figure 1 DET plot showing results from testing the raw recordings before and after audio quality checking.

It was obvious from the quick perceptual analysis that many of the pairs could be compared with an analytical phonetic approach even though the quality did not allow for proper acoustic automatic comparison. This indicates how important it is to reject audio that is not suitable for automatic acoustic comparison (Schwarz et al., 2011). The recordings in this database present the opportunity to further investigate urban youth speech in Sweden, with benefits to both forensic and sociophonetic areas of research and practices.

References

- Jessen, M., A. Alexander and O. Forth (2014). Forensic voice comparisons in German with phonetic and automatic features using VOCALISE software. *Proceedings of the Audio Engineering Society 54th International Conference*; London, June 12–14, pp. 28–35.
- Leinonen, T. N. (2010). An acoustic analysis of vowel pronunciation in Swedish dialects. (Dissertation), University of Groningen, Groningen
- Lindh, J., & Eriksson, A. (2009). The SweDat project and Swedia database for phonetic and acoustic research. In *e-Science, 2009. e-Science'09. Fifth IEEE International Conference on* (pp. 45-49) IEEE
- Reynolds, Douglas A., Thomas F. Quatieri, and Robert B. Dunn. "Speaker Verification Using Adapted Gaussian Mixture Models." *Digital Signal Processing* 10.1-3 (2000): 19–41.
- Schwartz, Reva, Joseph P. Campbell, and Wade Shen. "When To Punt on Speaker Comparison?" *J. Acoust. Soc. Am. The Journal of the Acoustical Society of America* 130.4 (2011): 2547.
- Taeldelman, J., 2005. The influence of urban centres on the spatial diffusion of dialect phenomena. In: Auer, P., Hinskens, F., Kerswill, P. (Eds.), *Dialect Change: Convergence and Divergence in European Languages*. Cambridge University Press, Cambridge, pp. 263–283