Speaking under cover: The impact of face-concealing garments on the acoustics of fricatives.

Natalie Fecher
Department of Language and Linguistic Science, University of York, York, United Kingdom
natalie.fecher@york.ac.uk

Introduction

The study reported here is part of a larger project which examines multimodal speech and speaker recognition in a forensic context, both from a human and an automatic perspective. It builds on pilot work by Llamas et al. (2009) and Watt et al. (2010) and investigates the forensic implications of audio-visual integration in speech intelligibility and speaker identification. For this purpose, the speaker's visual appearance is changed by introducing different types of headgear and face-concealing garments which are typically worn for occupational, recreational or religious reasons, or for the commission of crimes such as armed robberies and assaults. They can lead to miscommunication and complaints based on degraded speech intelligibility, or to a mismatch between authentic forensic speech samples and reference material. So far, comparatively little is known about the effects the face coverings have on speech production, acoustics and perception. The general difference from research on voice disguise is that the face coverings are not worn to deliberately disguise the voice with the intention of concealing identity (Zhang&Tan, 2008), but mainly to prevent a talker's face from being recognised. All of the presented face coverings obscure various parts of the talker's articulators. This hinders or eliminates the mapping between visual (facial) cues and auditory percepts on the part of the observer and as a consequence seems likely to appreciably degrade speech intelligibility ("visual speech"; see e.g. Schwartz et al., 2004). To a greater or lesser degree the face coverings also get in the way of the normal functioning of the articulators and will thus cause modifications to the acoustic (and consequently the auditory) signal.

Material and Method

This talk focuses on the acoustic analysis of the voiceless fricatives /s, ʃ, f, θ/ taken from high-quality recordings of six native British English speakers reading phonetically-controlled stimuli under various face disguise conditions (for details see Fecher, 2011). On the basis of relevant previous studies (e.g. Jongman et al., 2000) five features capturing intensity and spectral properties of the frication noise were chosen. A repeated-measures ANOVA was applied to investigate main and interaction effects of the *place of articulation*, *syllable position* and *disguise condition* on the *intensity*, *peak frequency*, and the first four statistical moments of the averaged FFT power spectrum, i.e. the *centre of gravity*, *variance*, *skewness* and *kurtosis*.

Figure 1. Control condition and face coverings, incl. relevant materials that cover the mouth/nose. Selection criteria were *forensic relevance*, *material* and *parts of the face covered*.

Results and Discussion

The choice of fricatives was motivated by their high perceptual confusability (see e.g. Lovitt&Allen, 2006), their relevance as consonantal features in forensic phonetics, and an anticipated larger attenuation by certain face-concealing garments of energy in higher frequency bands that are particularly discriminative for this phoneme class. The acoustic measures capturing the spectral properties of the frication noise were demonstrated to be significantly affected when the speakers wore different kinds of forensically-relevant face coverings. The shifts in the spectral patterns may, on the one hand, be caused by *acoustic damping effects* of certain mask materials, leading to acoustic energy being absorbed especially of higher frequencies. When face-concealing garments or other headgear obstruct the talker's face, they may also interfere with speech production, i.e. constrain articulation due to *physiological and somatosensory effects* (e.g. restricted jaw elevation, lip/nose contact, skin stretching) or lead to *articulatory compensation*, e.g. an increased vocal effort (Fuchs et al., 2010; Haley et al., 2010; Sluijter et al., 1997). Experimental work to investigate aspects of these phenomena is ongoing. Gaining more expertise in this area will be beneficial for forensic-phonetic practitioners in regard to speaker comparison, transcription work or the evaluation of the reliability of earwitness testimony.

Acknowledgement

This project is supported by the FP7 Marie Curie Initial Training Network "Bayesian Biometrics for Forensics" (http://bbfor2.net). Many thanks to Dom Watt for his feedback, and to all participants for allowing me to put tape on their mouth.

References

Fecher, N. 2011. Spectral properties of fricatives: a forensic approach. *Proceedings of the fourth ISCA Tutorial and Research Workshop on Experimental Linguistics*, May 25-27, Paris, France (accepted).

Fuchs, S., Weirich, M., Kroos, C., Fecher, N., Pape, D., Koppetsch, S. 2010. Time for a shave? Does facial hair interfere with visual speech intelligibility? In: Fuchs, S., Hoole, P., Mooshammer, C., Zygis, M. (eds.). *Between the regular and the particular in speech and language*, Frankfurt: Lang, 247-264.

Haley, K. L., Seelinger, E., Mandulak, K. C., Zajac, D. J. 2010. Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *J. Phonetics* **38** (4), 548-554.

Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am*. **108** (3), 1252-63.

Llamas, C., Harrison, P., Donnelly, D., Watt, D. 2009. Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers in Linguistics* (Series 2) **9**, 80-104.

Lovitt, A., Allen, J. B. 2006. 50 years late: repeating Miller-Nicely 1955. *Interspeech*, Pittsburgh, 2154-57.

Schwartz, J.L., Berthommier, F., Savariaux, C. 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* **93** (2), B69-78.

Sluijter, A. M. C., van Heuven, V. J., Pacilly, J. J. A. 1997. Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Am*. **101** (1), 503-513.

Watt, D., Llamas, C., Harrison, P. 2010. Differences in perceived sound quality between speech recordings filtered using transmission loss spectra of selected fabrics. *Talk at IAFPA 2010*, Trier.

Zhang, C., Tan, T. 2008. Voice disguise and automatic speaker recognition. *Forensic Science International* **175** (2-3), 118-122.