

Spectral properties of fricatives: a forensic approach

Natalie Fecher

Department of Language and Linguistic Science, University of York, UK

<https://doi.org/10.36505/ExLing-2011/04/0017/000186>

Abstract

This paper reports on the acoustic-phonetic 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. Speech samples were extracted from an audio-visual ‘face cover’ corpus that was collected for the purpose of investigating multimodal speech and speaker recognition in a forensic context. Findings are discussed with regard to constraints in speech production and acoustic damping effects caused by certain mask materials.

Key words: forensic phonetics, acoustics, fricatives, spectral moments

Introduction

Forensic speech scientists are particularly confronted with the inherent intra- and inter-speaker variation in speech, language and the human voice. The practitioners, however, have to cope not only with this speaker-induced variability, but also with the discrepancies between studio-quality speech material and often low-quality (authentic) forensic samples. One aspect that has to date not been studied in detail is the extent to which changes in a speaker’s visual appearance have an impact on the speech signal. Hereby we refer to various types of headgear and face-concealing garments (henceforth FCGs) typically worn for occupational, recreational and religious purposes, or for the commission of crimes such as armed robberies and assaults. All FCGs obscure parts of the talker’s articulators and will, to a varying degree, get in the way of their normal functioning and are likely to cause further modifications to the acoustic (and consequently the auditory) signal.

Material and method

Six speakers (3 M, 3 F, aged 21-36) were recorded in a professional TV studio at the University of York. All were native British English speakers with normal vision and hearing, training in phonetics, and no experience in regularly wearing any kind of face covering. Their task was to repeatedly read aloud a list of 64 phonotactically legal /C₁aC₂/ syllables embedded in the carrier phrase *He said <stimulus>*, each time wearing one of the FCGs shown in Figure 1. For the acoustic analysis, two tokens per syllable position per fricatives /s, ʃ, f, θ/ were extracted. On the basis of relevant previous studies (e.g. Jongman et al., 2000) five parameters capturing spectral properties and intensity of the frication noise were chosen for analysis. A repeated-measures ANOVA was applied to investigate the effects of the

dependent factors *intensity*, *peak frequency*, and the first four statistical moments of the averaged FFT power spectrum, i.e. *centre of gravity*, *variance*, *skewness* and *kurtosis*, on the independent factors *place of articulation* (POA), *syllable position* (SYL) and *disguise condition* (FCG). Measures were taken from spectra computed over non-filtered/non-pre-emphasised speech (48kHz/16-bit) in *Praat 5.1.44* (see also Fecher, 2011).



Figure 1: Control condition and face coverings. Selection criteria for the FCGs were forensic relevance, mask material and parts of the face covered.

Results

The ANOVA revealed a high level of significance ($p < .001$) for the main effects of POA ($F(3,12)=63.53$), SYL ($F(1,4)=69.48$) and FCG ($F(7,28)=12.68$), and also for the interactions between POA and SYL ($F(3,12)=29.00$) and POA and FCG ($F(21,84)=3.33$), on the *intensity* of all fricatives.

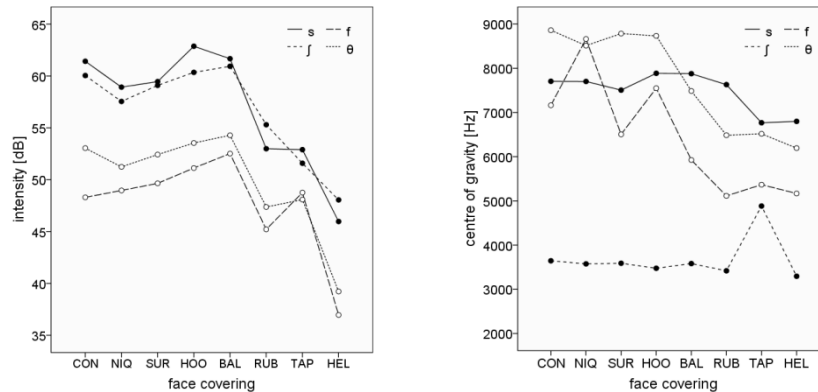


Figure 2 a) Intensity (*left*) and b) centre of gravity (*right*) for all fricatives and FCGs, averaged across speakers and syllable positions.

As expected, certain mask materials (HEL/RUB/TAP) absorb the sound energy more than others (see Figure 2a). However, intensity measures for several face coverings (BAL/HOO/SUR/NIQ) show even higher values than the control. This may have been caused by some speakers compensating for the (perceived) decreased loudness of their speech by speaking with greater vocal effort and thus ‘overriding’ the acoustic damping effects.

The main effects of POA ($F(3,12)=43.92$), SYL ($F(1,4)=125.93$) and FCG ($F(7,28)=8.81$) on *centre of gravity* were also found to be significant at $p<.001$. As can be seen in Figure 2b, certain masks lower the centre of gravity for /f, θ/, but not for /s, ʃ/. These FCGs are likely to absorb energy particularly in higher frequency bands, and the non-sibilants may be more prone to this damping due to their greater spectral diffuseness and overall lower energy (Shadle&Mair, 1996). The same explanation holds for the *variance* (standard deviation) of non-sibilants, which is generally higher for the non-sibilants than for the sibilants (POA: ($F(3,12)=49.21$, $p<.001$).

Regarding the *peak frequency* there is a significant main effect at $p<.001$ for POA ($F(3,12)=14.18$) and at $p<.01$ for SYL ($F(1,4)=30.92$), but none for FCG ($F(7,28)=1.03$, $p=.43$). *Skewness* and *kurtosis* were significant at $p<.001$ only for POA ($F(3,12)=82.22$; $F(3,12)=32.33$) (see Figure 3).

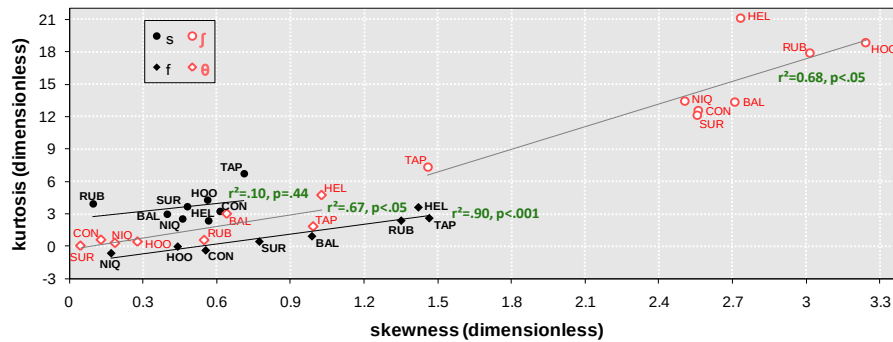


Figure 3: Skewness and kurtosis for all fricatives. The means for both spectral moments determine the data point for each face covering. For each fricative except /s/ a significant positive correlation can be observed.

Discussion and conclusion

The choice of fricatives for this study was motivated by their high perceptual confusability, their relevance as consonantal features in forensic phonetics, and an anticipated larger attenuation by certain FCGs of energy in higher frequency bands that are particularly discriminative for this phoneme class. The shifts in the spectral patterns may be caused by acoustic damping effects of certain mask materials, leading to energy being absorbed at higher frequencies (Llamas et al., 2009; Watt et al., 2010). When an FCG obstructs the talker's face, it may also interfere with speech production. Physiological and somatosensory effects, such as lip/nose contact, restricted jaw movement (Iskarous et al., 2009) and skin stretching (Fuchs et al., 2010) can lead to modified articulatory behaviour. Simultaneously, each subject may reveal individual compensation strategies, e.g., by increasing the overall intensity (loudness). The speakers may produce speech with greater pulmonary/glottal

effort, with the effect that the spectrum is not amplified uniformly, but higher frequencies are elevated more (Sluijter et al., 1997). This effect may be reinforced by some masks covering the speakers' ears, in this way impairing auditory self-monitoring. The outcome of the present study will be of particular value in combination with the results of upcoming perception tests in which participants will be asked to identify the stimuli under visually and acoustically degraded conditions. This may hinder the successful mapping between facial cues and auditory percepts, reducing intelligibility. Findings will be beneficial for forensic work (e.g., speaker comparison, transcription tasks, earwitness testimony) and research on audio-visual speech processing.

Acknowledgement

This work is supported by the FP7 Marie Curie ITN "Bayesian Biometrics for Forensics" (<http://bbfor2.net>). Many thanks to Dom Watt for his support.

References

- Fecher, N. 2011. Speaking under cover: The effect of face-concealing garments on spectral properties of fricatives. Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, August 2011 (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/M.: Peter Lang, 247-264.
- Iskarous, K., Shadle, C., Proctor, M. 2008. Evidence for the dynamic nature of fricative production: American English /s/. Proceedings of the 8th International Seminar on Speech Production, Strasbourg, 405-408.
- 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.
- Shadle, C., Mair, S.J. 1996. Quantifying spectral characteristics of fricatives. Proceedings of Interspeech, Philadelphia, 1521-24.
- 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 the IAFPA Conference 2010, Trier.