



LIFESPANS & STYLES

Undergraduate Working Papers on Intraspeaker Variation

Volume 3, Issue 2

Article 5

2017

The Effects of Forensically Relevant Face Coverings on the Acoustic Properties of Fricatives

Julie Saigusa
jssaigusa@gmail.com

ISSN: 2057-1720

doi: 10.2218/lv3i2.2017.1866

This paper is available at: <http://journals.ed.ac.uk/lifespansstyles>

Hosted by The University of Edinburgh Journal Hosting Service: <http://journals.ed.ac.uk/>

The Effects of Forensically Relevant Face Coverings on the Acoustic Properties of Fricatives

Julie Saigusa

This forensically motivated study investigates the effects of a motorcycle helmet, balaclava, and plastic mask on the acoustics of three English non-sibilant fricatives, /f/, /θ/, and /v/, in two individuals. It examines variation within the individual as an effect of the physical environment. Two speakers recorded a list of minimal pairs in each of the three guises and with no face covering. The results showed that facewear significantly affected fricative intensity and the four spectral moments: centre of gravity, standard deviation, skewness, and kurtosis. The acoustic changes caused by facewear have implications for judging the reliability of earwitnesses' content recall and voice identification as well as forensic speech scientists' examination of content and speaker identity in disputed recordings.

1 Introduction and Background

1.1 Motivation and Research Question

Criminal acts lend themselves to disguise, especially in a world where video and photo surveillance is increasingly pervasive. A range of disguises are exemplified in popular culture in films like *V for Vendetta*, which features the now-iconic Guy Fawkes mask, and *Point Break*, featuring robbers dressed in rubber masks of US presidents. Real-life CCTV footage circulated by news media often shows a suspect in a ski mask or other face-concealing disguise (e.g., Martinez and Lerten 2016). The evidence against such individuals must necessarily come from earwitness testimony of what was heard (i.e., content) and who was heard (i.e., speaker identification). Forensic phoneticians are also often called upon to judge the identity of a speaker or a crucial word in a bad quality recording (Fraser 2014). This study is a forensically based attempt to determine how several relevant face coverings affect the acoustic properties of certain fricatives and to explore the implications for forensic speech science. It expands the kind of phenomena examined in studies of intraspeaker variation to the passive and active effects of immediate physical obstruction on speech.

Intuitively, it can be predicted that a material obstructing sound propagation will affect how speakers produce speech, how speakers sound to listeners—either in speech content or voice quality, and how far the speech wave can physically propagate. Even the angle of the speaker in relation to the listener may affect perception: high frequency sounds do not travel as well in non-forward directions as lower frequency sounds (Thomas 2002). This study addresses the first aspect: are the acoustic properties of certain fricatives produced through facewear different from those produced without facewear? The study examines two fricative pairs: one with a voicing contrast (a contrast previously unstudied in this context), /f/ and /v/, and one with a place contrast, /f/ and /θ/.

The area of forensic facewear research is relatively new. Llamas et al. (2008) and Fecher (2014) pioneered this subfield by examining how different types of face coverings affect the acoustic signal, and in turn, how listeners perceive that signal. Since there has been little work on this topic to date, it is important not only to expand on previous findings by bringing in new factors and manipulations, but also to replicate what has been found so far. It is predicted that fricative acoustics will be affected by facewear and manifest in differing spectral properties and intensities across the conditions.

1.2 Forensic Speech Science and Facewear Research

Earwitness testimony is common in court cases. One legal meta-analysis (Laub et al. 2013) analysed 226 US court cases that focused on earwitness testimony. Earwitnesses may be asked to report what they heard a suspect say or to identify a suspect by their voice, such as in a voice lineup (Broeders 2013). Though the unreliability of eyewitnesses has been exhaustively researched (Wells and Olson 2003), less has been done on aspects of auditory evidence, though this is by no means an unstudied area. A number of studies have focused on the reliability of content recall (e.g., Ling and Coombe 2005), while others have investigated earwitness ability to identify a voice (e.g., Yarmey 2007). Briefly, Fecher (2014), which is discussed in more detail below, examined the ability of listeners to discriminate between voices when stimuli were recorded through facewear. The study found that compared to near-ceiling performance in the control condition, facewear degraded listeners' abilities to tell whether a pair of consonant-vowel syllables was spoken by the same or different people. In addition, covert recordings made by law enforcement or tapes of calls to emergency services are often of degraded quality

and require careful acoustic analysis to determine their contents (Fraser et al. 2011). In these cases, knowledge of the type of face covering an individual is wearing, if any, could be of crucial importance to forensic practitioners.

Masks and other types of facial concealment are prevalent in crime. In addition to the iconic balaclava, criminals may use handkerchiefs, rubber masks, tights, party masks, or even motorcycle helmets to disguise themselves, to name only a few types of facial concealment (Fecher 2014). The current study uses a balaclava and helmet, as in Fecher (2014), and delves into the wide range of other face coverings with a plastic mask.

Llamas et al. (2008) is the first forensic study on facewear. Their first experiment focused listener perception of speech produced through a balaclava, surgical mask, niqab (full-face veil), and no cover (control). Confusions between fricatives, especially place distinctions, were among the most common errors, inspiring the present investigation into the acoustic consequences of facewear on fricatives in particular.

The second experiment investigated the acoustic transmission loss properties of different materials on a non-speech signal produced by a loudspeaker. The aim was to observe how the materials themselves attenuate sound independent of how a speaker may change production while wearing facewear. The range of fabrics in this study was greater than in the perception experiment, and included the niqab (polyester), balaclava (acrylic yarn), surgical mask (paper), handkerchief (cotton), scarf (wool/acrylic blend), stockings (nylon), and loudspeaker cover fabric (a supposedly “acoustically transparent” woven fibre). The materials were placed between a loudspeaker that played a sequence of pulses and a microphone that picked up the pulses. Unexpectedly, only the surgical mask showed any significant transmission loss. The other fabrics, even thicker ones, showed little or no difference to the control. As Llamas et al. (2008) point out, this is not an accurate representation of how natural speech is produced or transmitted, but it gives a preliminary insight into characteristics of a range of fabrics and how they may affect speech. Their study did not examine any type of plastic face covering, which the current one does.

Llamas et al. (2008) laid the foundations for Fecher (2014), which inspired the current study. Fecher’s work is a comprehensive PhD thesis which examined the acoustic properties of voiceless stops and fricatives produced through facewear, listener perception of those consonants in both auditory only and audiovisual conditions, and listener discrimination of facially covered speakers. Fecher recorded five males and five females to produce the Audio-Visual Face Cover corpus (AVFC). The speakers were recorded wearing a motorcycle helmet, balaclavas with and without mouth holes, a strip of tape across the mouth, a niqab, a surgical mask, a scarf across the nose and mouth with a hoodie over the head, and a full-head rubber mask. The material recorded consisted of a nonsense consonant-vowel-consonant syllable (to prevent top-down word processing) in the carrier sentence “He said X”. A range of consonants were used, including voiceless stops and fricatives in both onset and coda position, except where prohibited by English phonotactic rules, preceded and followed by the vowel /a:/.

The voiceless fricatives produced for the AVFC were analysed for a number of spectral measures including spectral peak, centre of gravity, standard deviation, skewness, kurtosis, and intensity (see Section 3 for a discussion of these terms). Non-sibilant fricatives (the focus of the current study) showed a significant effect of facewear on intensity and spectral moments. Most fricatives were lower in intensity than the control. With few exceptions, centre of gravity was lower than the control for /f/ and /θ/. /f/ showed more positive skewness in the balaclava rubber mask, tape, and helmet conditions. Kurtosis increased in the balaclava and helmet conditions for /θ/, and was affected overall for /f/. Additionally, syllable position was significant. Based on these results, facewear can be expected to affect intensity and spectral measures in the current study. In an auditory-only forced-choice identification task, non-sibilant fricative confusions were again among the most common error type. /θ/ was commonly misidentified as /f/ across all conditions, but again the opposite (/f/ for /θ/) occurred less than 10% of the time. /v/, the other phoneme used in the present study, was usually identified correctly.

The present study aims to extend Fecher’s (2014) findings by replicating some previously used facewear conditions and adding a previously untested one. Llamas et al. (2008) tested an array of voiced sounds, including /v/, but Fecher (2014) only examined voiceless sounds acoustically. This study narrows the phonetic focus of Llamas et al. (2008) but expands on the findings of Fecher (2014) by testing two minimally contrastive phoneme pairs, /f-θ/ and /f-v/.

Finally, a number of non-forensic studies have examined the effects of various head and face coverings on communication in medical, military, and other industrial contexts. These results confirm the negative effect of facewear on intelligibility, though so far this area has lacked detailed investigation into acoustics or patterns of phonetic confusion. Surgical masks, respirators, and Air Force headgear have all have been found to reduce intelligibility (Radonovich Jr. et al. 2009, Sommer 1976, Wittum et al. 2013).

1.3 Speech Production and Acoustics

The first research question is concerned with the acoustic properties of fricatives produced through face coverings. A purely acoustic analysis cannot determine what changes come from natural inhibition of the sound

wave by the material and what changes the speaker is actually making to their articulation in response to facewear. However, previous research on how speakers compensate for inhibited articulation provides a framework for what characteristics can be expected in this study.

Speakers perturbed by jaw weights or bite blocks have been found to adapt quickly to inhibited movement in one articulator by increasing use of another one (Lindblom et al. 1979), producing much the same acoustic output. Stevens' (1989) influential quantal theory proposed that languages exploit articulatory regions within which changes to articulation do not result in major changes to acoustic output. In the context of this study, these findings suggest that even with slightly restricted articulatory movement, changes in production will not be the major source of acoustic difference. Some studies, however, have reported that consonants require a much more precise articulatory configuration than vowels, finding incomplete adaptation (Flege et al. 1988, McFarland and Baum 1995). McFarland and Baum (1995), in particular, found that bite-block perturbed sibilants and stops differed acoustically from controls immediately after presenting perturbation and after a 15-minute accommodation period. They also observed a degree of individual variability in adaptation ability. The current study is not focused on articulatory perturbation, but findings in that area can untangle what acoustic changes may result from the restriction of jaw movements by crash helmets or other coverings.

Two of the face coverings chosen for this study cover the speakers' ears as well as their mouths and noses (balaclava and helmet). Earwear affects the natural feedback loop that speakers use to monitor themselves. When wearing a garment that covers the ears, speakers perceive their own voices differently, and subsequent adjustments have been the focus of many studies (e.g., Garnier et al. 2010, Tufts and Frank 2003, Martin et al. 1976). Garnier et al. (2010) investigated the methodological implications of using headphones to transmit noise on manifestations of Lombard speech (i.e., speech in noise). The study found that the acoustics of speech produced by headphone-wearers differed significantly from noise transmitted by loudspeaker. Notably, Tufts and Frank (2003) examined speech produced when participants wore different types of earplugs in quiet and in noise. They found that, overall, speakers wearing earplugs produced lower intensity speech than those with uncovered ears as well as less higher frequency sound energy. Indeed, Fecher (2014) also observed a migration of sound energy from higher to lower frequencies in facewear conditions, which was attributed to the attenuating characteristics of the material itself. Llamas et al. (2008) reported fabrics attenuating energy above 10kHz. In the context of the current study, the earwear effect supports the prediction that facewear will affect speech acoustics.

1.4 Social Connotations of Phonetic Variation

One of the features investigated here, the realisation of /θ/ as [f], has been examined as "TH-fronting" in sociolinguistics and is a linguistic stereotype in parts of the English-speaking world. It is primarily found in working-class and urban areas of the south of England, though it has also recently been observed in the northeast and in some parts of urban Scotland (Clark and Trousdale 2009, Levon and Fox 2014). TH-fronting is important in the present context in that its realisation could potentially be acoustically altered enough by a facial covering for listeners to misperceive it. It has been found that listeners' perceptions are affected by the dialect they believe they are hearing, even given identical stimuli (Niedzielski 1999, Thomas 2002). Perceiving /θ/ as [f] may well influence listeners' perceptions of the speaker's characteristics and their ability to recall voice qualities (Thomas 2002). In addition to non-facewear confusion studies which found voiceless non-sibilants (/f/ and /θ/) the most easily confusable, Fecher (2014) and Llamas et al. (2008) found facewear to significantly affect error rates for these consonants. In the present study, the extension of this effect to the perception of real words (cf. nonsense syllables) would support a TH-fronting illusion hypothesis.

1.5 Intensity and Spectral Moments

The acoustic parameters examined in this study follow Fecher (2014). They are intensity, centre of gravity, standard deviation, skewness, and kurtosis. Intensity is the measure of the loudness of a sound, here measured in decibels. Centre of gravity is the first of four measures referred to as spectral moments (Hardcastle et al. 2010). It measures the frequency in Hertz at which the sound energy is primarily concentrated (Watt 2013). The second spectral moment, standard deviation, is a measure in Hertz of how spread out the energy is around the centre of gravity (Watt 2013). A higher standard deviation indicates that energy is more spread across the spectrum, while a lower value indicates that energy is more concentrated around the centre of gravity. The third spectral moment, skewness, is a measure of spectral tilt, or energy asymmetry around the centre of gravity. Positive values indicate more energy in lower frequencies, and vice versa (Watt 2013). Kurtosis, the fourth spectral moment, is a measure of how concentrated spectral energy is in a peak relative to the energy distribution of the spectrum (Hardcastle et al. 2010), i.e., how the shape of a spectrum differs from the normal distribution. A positive value indicates a leptokurtic, or more peaked, distribution. Kurtosis and standard deviation are often correlated, since they are concerned with the energy distribution (Fecher 2014).

1.6 Summary

The current study aims to determine how a face covering affects the acoustic characteristics of three non-sibilants: /f/, /θ/, and /v/. Findings from pioneering forensic facewear studies (Fecher 2014, Llamas et al. 2008) indicate that facewear will have a significant effect on the intensity and spectral properties of fricatives, especially by moving sound energy to lower frequencies. Theories of speech production and studies on articulatory perturbation found that variable motor input can and often does lead to a minimally varied acoustic output (Stevens 1989). However, evidence from “earwear studies” such as Tufts and Frank (2003) indicates that intensity and energy concentration frequency can both decrease when a speaker’s ears are covered. This suggests that any spectral effects of facewear in the current study are likely to be due to attenuating characteristics of the face covering itself and to the speakers’ impaired ability to monitor the sound of their own voice rather than atypical articulatory configurations. The current study builds on previous work by examining a voiced fricative and expanding the range of facewear used.

2 Methods

2.1 Materials

The read materials consisted of 42 monosyllabic English words (see appendix for list) with one of the target fricatives (/f/, /θ/, or /v/) in onset or coda position. The words were minimal or near-minimal pairs of either a /f/-/θ/ (place) or /f/-/v/ (voicing) contrast. Word frequency was balanced within place and voicing contrast pairs (/f-θ/: $t(14.62) = -0.081$, $p = 0.93$, /f-v/: $t(15.08) = -0.1583$, $p = 0.88$) (Davies 2004). Vowel height and backness were not perfectly balanced due to the limited number of possible minimal pairs (see list in appendix). Each contrast had a majority of high front vowels. The words were presented on individual index cards and randomised between each repetition.

The face coverings chosen for this experiment were a full-head motorcycle helmet, a balaclava (no mouth hole), and a plastic party mask (no mouth hole) (see Figure 1). The selection of these face coverings was motivated by Fecher’s (2014) use of a helmet and balaclava and mention of plastic Halloween-type masks as prevalent in criminal and social situations, though that study did not include them.



Figure 1: The three face coverings used in the experiment (left: a full-head motorcycle helmet; middle: a hard plastic party mask with no mouth hole; right: a knit balaclava with no mouth hole).

2.2 Participants

The participants consisted of two native Standard Southern British English speakers: one 20-year-old female and one 24-year-old male, neither of whom reported any speech or hearing disorders. In addition, neither speaker normally exhibited TH-fronting. The male speaker reported previous experience wearing motorcycle helmets.

2.3 Audio Recording

Participants were recorded individually in a sound-treated booth and recordings were sampled at 44.1 kHz. The microphone was placed approximately 30 cm from the participant's mouth. They were told to try to stay at that position but adherence to this was not monitored. They were given a stack of 42 index cards and instructed to read each word, making sure not to flip to the next card until they had finished speaking. No special instructions on how to speak were given. Each participant recorded the list twice for every face covering as well as for no face covering, but not consecutively. After each repetition, the list was recorded in a different face covering; this order was varied between the subjects. This resulted in eight repetitions per participant. The word list was randomised by shuffling 8 times per participant after each repetition. In total, 672 tokens were recorded.

2.4 Processing

Each recording was hand-segmented for word and phoneme in Praat (Boersma and Weenink 2009). The fricatives were identified visually on the spectrogram and the waveform by energy onset and offset as well as by auditory examination through Sennheiser HD202 headphones (see Figure 2). Each word and fricative was extracted from the larger .wav file.

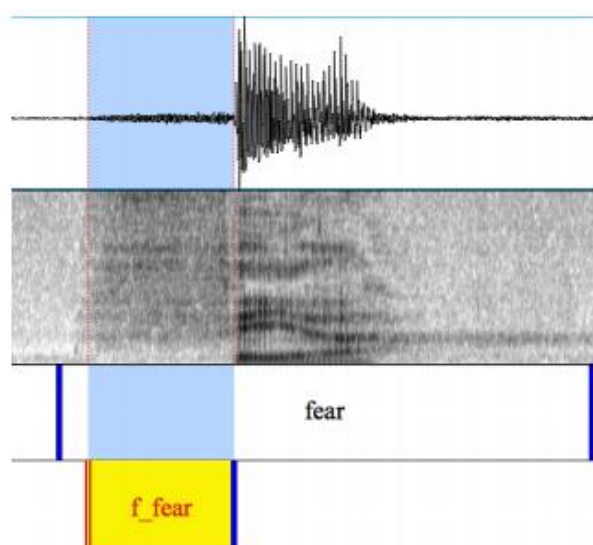


Figure 2: Spectrogram of a *fear* token showing segmentation.

3 Results

The individuals' data were analysed separately and are referred to as the “female speaker” and the “male speaker” to distinguish between them, but this does not imply that gender was a causal influence on any of the parameters investigated. Subsets of the data were examined visually to confirm normality. In total, 672 tokens were collected, of which 18 tokens were excluded because of noise in the recording, leaving 654 for analysis. Data were analysed in R (R Core Team 2015). All two-way ANOVAs tested cover condition levels (balaclava, control, helmet, and mask) and syllable position (onset and coda). Syllable position was included as a predictor to ensure that the main effects of facewear were not falsely reported, since position is known to have phonetic effects on segments (Fecher 2014).

3.1 Intensity

Intensity measurements were extracted from Praat with a script (Elvira-Garcia 2014) which extracted mean intensity across the fricative. The mean intensity of each phoneme in each cover condition for the two speakers is shown in Figure 3, pooled across onset and coda positions. For the female speaker, a two-way ANOVA examining cover condition and syllable position reported a significant effect of cover condition for each phoneme: /f/ ($F(3,162) = 25.76$, $p < 0.001$), /θ/ ($F(3,61) = 10.34$, $p < 0.001$), and /v/ ($F(3,86) = 14.46$, $p < 0.001$), and also for the male speaker: /f/ ($F(3,153) = 8.23$, $p < 0.001$), /θ/ ($F(3,63) = 3.99$, $p < 0.02$), and /v/ ($F(3,83) = 6.54$, $p < 0.01$).

There was a significant main effect of syllable position for /θ/ ($F(1,63) = 9.24, p < 0.01$), and a significant interaction between cover and position for /θ/ again ($F(3,63) = 6.72, p < 0.01$) for the male speaker.

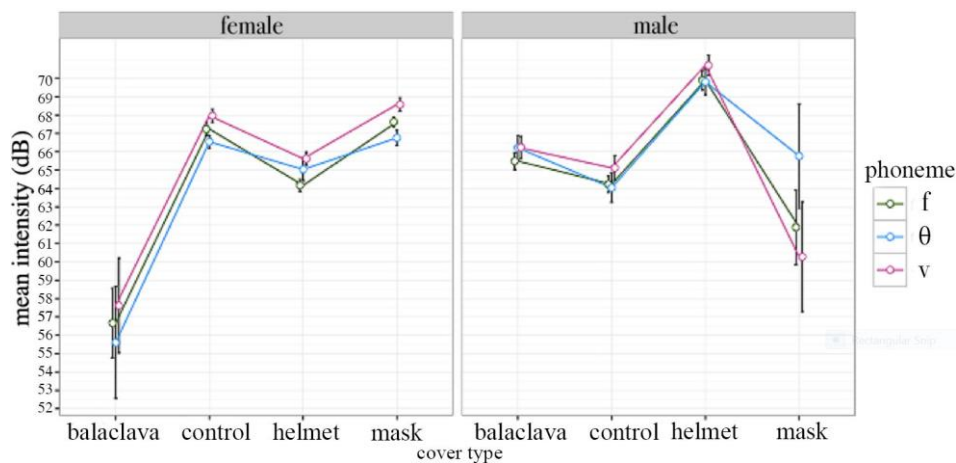


Figure 3: Mean intensity (dB) of each phoneme for the two speakers. Error bars show the standard error of the mean.

For the female speaker, T-tests conducted using a Bonferroni-corrected significance level of $p = 0.017$ reported a significant difference in intensity between the control and the balaclava ($T(87.61) = 7.73, p < 0.001$), the control and the helmet ($T(159.38) = 7.77, p < 0.001$), but not the control and mask ($T(166.56) = 1.40, p = 0.16$). For the male speaker, the helmet differed significantly from the control ($T(154.12) = 11.60, p < 0.001$). The main effect of cover condition is consistent with findings in Fecher (2014) and support the hypothesis that facewear affects fricative intensity.

3.2 Centre of Gravity

All spectral moments were measured over the centre 80% of the fricative via script in Praat (Lennes 2002). Figure 4 displays the mean centre of gravity for each phoneme across cover conditions for each speaker individually. Syllable positions were pooled together. Overall, mean centre of gravity was lower in the covered conditions than in the control for /f/ and /θ/, while /v/ exhibited different behaviour. For the female speaker, two-way ANOVAs again reported a main effect of cover condition for all three phonemes: /f/ ($F(3,162) = 42.61, p < 0.001$), /θ/ ($F(3,61) = 18.08, p < 0.001$), and /v/ ($F(3,86) = 0.08, p < 0.001$), and also for the male speaker: /f/ ($F(3,153) = 11.91, p < 0.001$), /θ/ ($F(3,63) = 2.90, p < 0.05$), and /v/ ($F(3,83) = 8.54, p < 0.001$).

The significant effect of cover condition on the three phonemes is consistent with Fecher's (2014) observation that centre of gravity for non-sibilants was lower in facewear conditions than in the control and supports the current study's hypothesis that facewear affects spectral properties of fricatives.

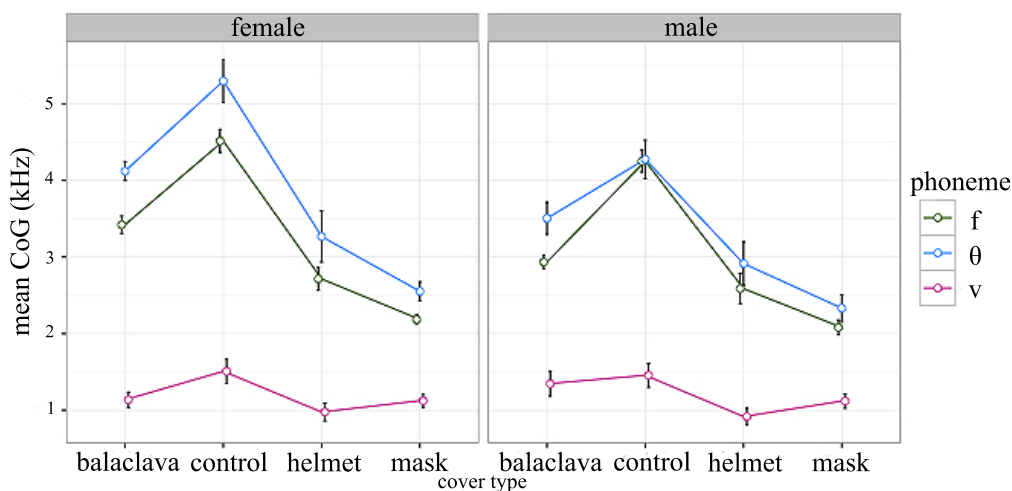


Figure 4: Mean centre of gravity (kHz) of each phoneme for the two speakers. Error bars show the standard error of the mean.

3.3 Standard Deviation

Figure 5 shows the mean standard deviation in Hz for each phoneme across cover types and speaker. Two-way ANOVAs for the female speaker showed a significant main effect of cover condition on standard deviation for all three phonemes: /f/ ($F(3,162) = 43.25, p < 0.001$), /θ/ ($F(3,61) = 18.09, p < 0.001$), and /v/ ($F(3,86) = 21.61, p < 0.001$). For the male speaker, there were significant main effects of cover condition for /f/ ($F(3,153) = 8.53, p < 0.001$) and /v/ ($F(3,83) = 5.94, p < 0.002$), but not for /θ/ ($F(3,63) = 2.45, p = 0.07$). However, there was a significant interaction of cover condition and position for /θ/ ($F(3,63) = 4.75, p < 0.01$).

The significant effects of facewear found for /f/ differ from Fecher's (2014) finding that facewear did not significantly affect the standard deviation of /f/. Fecher also found a significant effect of facewear on /θ/ that was not observed in this male speaker. As seen in Figure 5, the helmet appeared to increase the spread of energy around the mean for /f/ and /θ/, and decrease it for /v/.

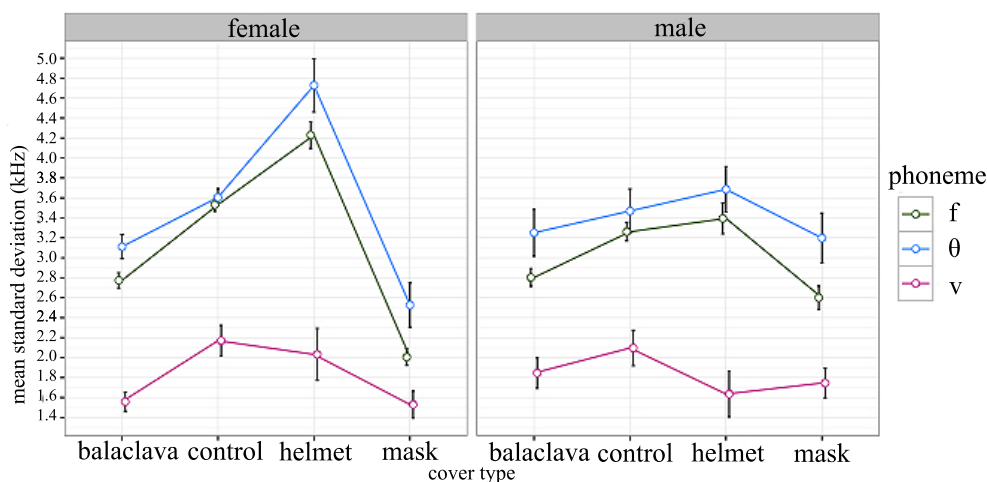


Figure 5: Mean standard deviation (kHz) of each phoneme for the two speakers. Error bars show the standard error of the mean.

3.4 Skewness

Figure 6 shows the mean skewness of each phoneme across conditions and speakers. For the female speaker, there was a significant main effect of cover condition for all three phonemes: /f/ ($F(3,162) = 100.32, p < 0.001$), /θ/ ($F(3,61) = 78.05, p < 0.001$), and /v/ ($F(3,86) = 73.52, p < 0.001$); the same applied in the case of the male speaker: /f/ ($F(3,153) = 15.50, p < 0.001$), /θ/ ($F(3,63) = 6.40, p < 0.001$), and /v/ ($F(3,83) = 7.11, p < 0.001$). For the female speaker, a significant effect of position was also reported for /v/ ($F(3,86) = 9.77, p < 0.01$). However, a significant interaction was also found between cover and position for /v/ ($F(3,86) = 7.20, p < 0.01$).

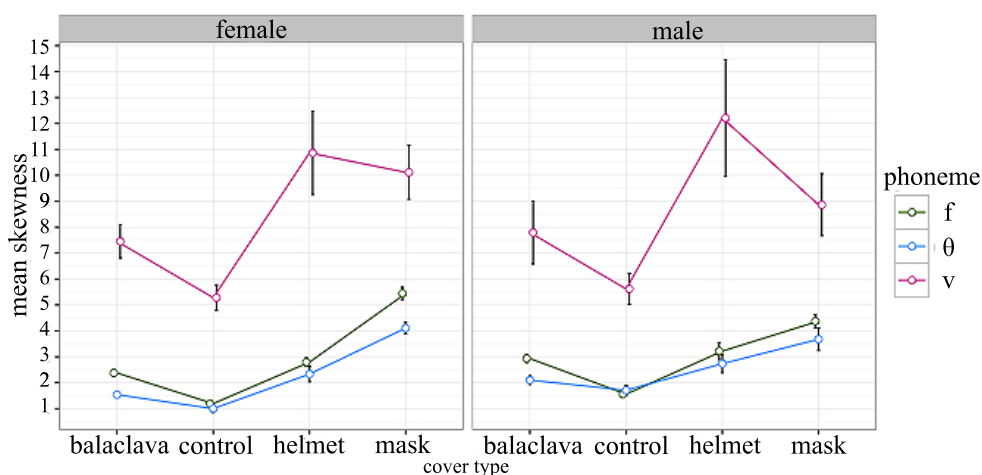


Figure 6: Mean skewness of each phoneme for the two speakers. Error bars show the standard deviation of the mean.

The significant effect of cover condition on skewness is consistent with Fecher (2014), who also reported a lower skewness in the control (more energy in higher frequencies) than in the facewear conditions. This

supports the hypothesis that facewear affects fricative acoustics, in this case, energy distribution, and the suggestion by Fecher (2014) that intervening materials shift energy downwards.

3.5 Kurtosis

Figure 7 shows mean kurtosis values for the three phonemes across conditions and speakers. Two-way ANOVAs again showed a significant effect of cover condition for all phonemes for the female speaker: /f/ ($F(3,162) = 55.19, p < 0.001$), /θ/ ($F(3,61) = 46.18, p < 0.001$), and /v/ ($F(3,86) = 43.28, p < 0.001$), and for the male: /f/ ($F(3,153) = 16.80, p < 0.001$), /θ/ ($F(3,63) = 6.27, p < 0.001$), and /v/ ($F(3,83) = 7.31, p < 0.001$). For the female speaker, there was also a significant effect of position for /v/ ($F(1,86) = 6.08, p < 0.02$) and an interaction between position and facewear ($F(3,86) = 7.49, p < 0.001$).

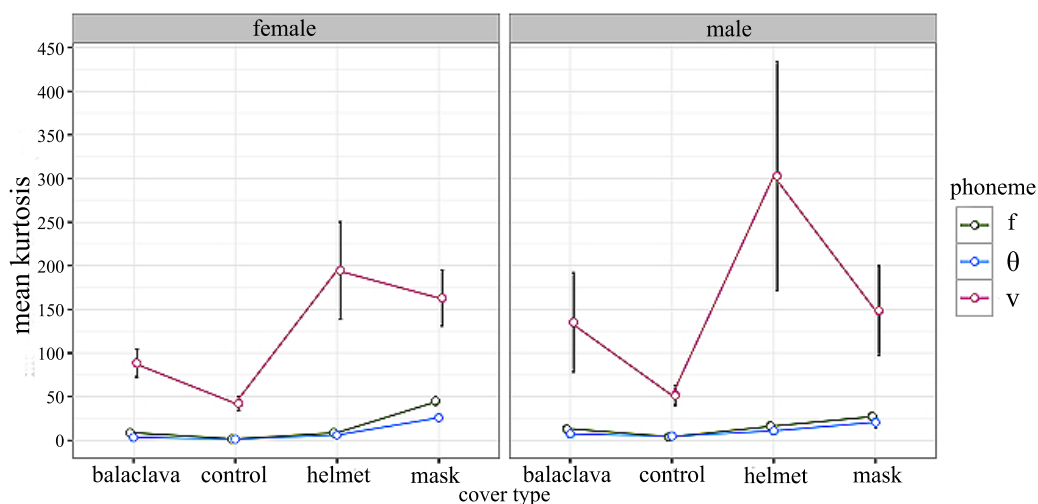


Figure 7: Mean kurtosis of each phoneme for the two speakers. Error bars show the standard deviation of the mean.

The higher mean kurtosis in the balaclava and helmet conditions relative to the control were also observed by Fecher (2014). The higher spectrum “peakedness” (Fecher 2014) is supported by the more concentrated energy for the balaclava and mask in Figure 5, as standard deviation and kurtosis are often correlated (Fecher 2014).

4 Discussion

4.1 Acoustic Properties of Fricatives Produced through Face Coverings

The aim of this experiment was to determine if and how face coverings affected the acoustic properties of /f/, /θ/, and /v/. It was hypothesised, based on findings in Fecher (2014), that the acoustic properties of the fricatives recorded would differ significantly from those in the control. The evidence reported in Section 3 supports the hypothesis that covering the face changes the acoustic output. Intensity and all four spectral moments measured were significantly affected by facewear.

4.1.1 Intensity

Intensity is a measure of sound loudness. The interpretation of these results requires synthesis of several possible sources of difference. As outlined in the introduction, because the analysis did not involve articulatory measurements, it is not always possible to distinguish between any changes in speaker behaviour and physical absorption qualities of the face covering. Furthermore, as the two will always co-occur in facewear speech, separating them does not further the practical purpose of forensic research.

For the female, helmet and balaclava speech was lower than the control, while the mask did not significantly differ. This finding is supported by Tufts and Frank’s (2003) observation that intensity was lower for speech produced in quiet when the speaker’s ears were covered. Both the helmet and the balaclava covered the ears, while the mask did not. For the male speaker, the intensity of helmet fricatives was higher than the control. A possible explanation stems from his reported previous experience of wearing motorcycle helmets. It has previously been found that speakers seated in a car even in the absence of noise produce speech similar to

that produced in car-noise (Podlubny et al. 2016). In short, their previous knowledge of the effects of their physical environment caused them to compensate even when the communication conditions were not adverse. Similarly, the male speaker may have known how his speech would be affected and how he would need to compensate to be heard, shifting his speech style (i.e., speaking louder) in response to the physical environment (i.e., the helmet).

4.1.2 Spectral Moments

The results indicate that face coverings affect the spectral properties of some fricatives. All three phonemes showed differences in the four spectral moments across the different types of cover.

These findings are well supported by previous studies that observed differences in sound energy concentration in the presence of intervening materials. Fecher (2014) reported that at least in the helmet condition, sound energy migrated into the lower frequency bands compared to the control, resulting in a lower centre of gravity and a more positively skewed spectrum (i.e., more energy in lower frequencies compared to higher ones). Observing the results in Figures 4 and 6, it can be seen that this, indeed, was the case with the two speakers. For every face covering, centre of gravity measures were lower and skewness was higher than in the control. /v/, which was not examined in Fecher (2014), showed especially large differences in skewness and kurtosis between control and experimental conditions compared to the voiceless fricatives. However, though /v/ also showed a lower centre of gravity in facewear conditions than in the control, the difference was not as large as for the voiceless sounds.

The lowering of spectral energy concentration was also found by McFarland and Baum (1995) and Flege et al. (1988) in studies of articulatory perturbation. However, those studies used a bite-block perturbation that increased jaw widening and therefore the dimensions of the front cavity as well. Here, jaw movement was most likely to be restricted rather than increased, especially in the helmet condition. This lends support to the evidence that properties of the covering material account for migration of sound energy to lower frequencies rather than a large articulatory change on the part of the speaker. However, Tufts and Frank's (2003) earwear study found a lowering of spectral energy frequency, as well as a decrease in intensity. This suggests that the hearing obstruction caused by the balaclava and mask additionally affected centre of gravity.

The overall effect of syllable position for some phonemes was also found in Fecher (2014). This was expected given that coda consonants exhibit more articulatory reduction than onsets generally (Ohala and Kawasaki 1984). The acoustic analysis indicates that a material covering the face can significantly affect spectral properties that experts who are consulted in court cases use to determine individual speaker likelihood or disputed content. This is discussed further in Section 4.3.

4.2 Social Perception and a TH-fronting Illusion

/f/ and /θ/ have been found to be highly confusable (Fecher 2014, Llamas et al. 2008, Miller and Nicely 1955). In Fecher (2014), /θ/ was more often identified as /f/ than the opposite (/f/ for /θ/). It has been claimed that /θ/ to /f/ is a perceptually motivated sound change currently occurring in English (Blevins 2004).

The movement of energy to lower frequencies found in this study may be a reason why covered /f/ and /θ/ are easily confused: higher frequency noise information (especially above 10kHz) has been found to play a role in identifying these sounds (Tabain 1998, Tabain and Watson 1996), especially for /f/. This means that in a situation where a crime is occurring, /θ/ may be heard as /f/. As mentioned before, the phenomenon of TH-fronting, where the phonemic /θ/ in words like *thin* is realised as [f] (making *thin* and *fin* homophones), is well known. It has been the focus of numerous studies investigating regional and class dialects (Levon and Fox 2014, Schlee and Ramsammy 2013) and sound change (Blevins 2004), and is a growing feature in many southern English dialects as well as in parts of the northeast and Scotland (Bennett 2012, Clark and Trousdale 2009). TH-fronting is often characterised as a feature of working-class speech, or more specifically, “chavs”. A “chav” is a social stereotype of the poor working class, often vilified by the media and other online users. For example, a highly rated definition of “chav” on Urban Dictionary is “amoral”, and the term is described as applicable to “every culture with a nasty, thieving element” (chavspotting 2004).

Combined with the prevalent cultural stereotype of petty criminality in “chavs”, the increased ambiguity and potential misperception of /θ/ may lead a witness to false impressions of the perpetrator's dialect and social class. Though the effects of facewear on other salient dialectal characteristics such as t-glottaling and h-dropping (Bennett 2012) were not investigated, a TH-fronting illusion could be an important factor in determining the reliability of eyewitness testimony. In addition to the sound-attenuating characteristics of any material itself, the lack of visual cues for place of articulation that necessarily occur when the face is obscured make these fricatives difficult to distinguish (Fecher 2014). Niedzielski (1999) found that participants' phonemic perception shifted depending on what dialect they believed they were listening to. This effect could easily be exacerbated by /θ/ misperception.

4.3 Forensic Speech Science: Masks in Crime Revisited

One of the main purposes of the study was to examine what the results of facewear experiments mean forensically, for example, in cases involving earwitness situations or disputed recordings. Firstly, fricatives are usually examined by forensic speech scientists in casework (Gold and French 2011). In a case where determining if a recorded voice belongs to a suspect or profiling a voice's characteristics is of the utmost importance, accurate measurement is crucial. It is well known that intraspeaker variation makes reliance on any one acoustic measurement unviable—indeed, conversation context has been found to affect individuals' fricative measurements (Saigusa 2016). This study has shown that face coverings can also alter those same spectral properties. Fecher's (2014) speaker discrimination experiment additionally found that facewear impairs a listener's ability to tell if syllables are spoken by the same or different people. Consequently, knowledge of how facewear affects speech is necessary to make an accurate analysis and present the best possible judgment. If an analyst does not know that the speaker's face was covered, their speaker comparison is likely to suffer.

Incriminating audio recordings present the possibility for misinterpretation with serious consequences. Recordings can be made by a hidden microphone (e.g., in a house or car) to capture incriminating evidence in a police investigation. Similarly, recordings of private telephone calls or calls made to emergency services are often provided as evidence in criminal cases (Fraser et al. 2011). These recordings are often of degraded quality due to the equipment (e.g., a small microphone), the background noise (such as crowd or engine noise), and in the case of emergency calls, an extremely distressed speaker. A famous example of how these factors can affect perception is taken from a murder case in 1994 (Fraser et al. 2011). In this instance, a man made a call to emergency services, where he allegedly confessed to murder by saying "I shot the prick". However, other listeners heard him accuse his father, "He shot them all". Fraser's study showed that listeners' interpretations of the call were not independent of context but varied significantly depending on the information they were given.

Uncovered speech in these recordings is clearly already acoustically ambiguous—often forensic phoneticians are called in to make an expert judgment on the content of the recording. If suspects were wearing any type of face covering at the time, the reliability of these recordings will be even more impaired. The results of this study showed that face coverings dampen noise in high frequency bands; if it is known that a speaker in a recording was wearing a face covering, its effects on the acoustic signal can be interpolated and taken into account when offering a judgment on content or identity. In fricatives, the migration of a peak to a different frequency due to a face covering may give a misleading acoustic impression. Knowing that a face covering was involved can help forensic phoneticians to make the most accurate analyses.

5 Conclusion

This study explored how facewear affected both the acoustic output and the perception of /f/, /θ/, and /v/ within two particular individuals. As hypothesised, facewear significantly affected fricative intensity and four spectral measures for both speakers. In particular, sound energy migrated from higher to lower frequencies in facewear conditions compared to the control, and in general, intensity was lower than in the controls. In addition to the passive attenuating effects of the facewear, there was some evidence that speakers shifted production to compensate for the obstruction. These results suggest that facewear, a common feature in criminal cases, can change the acoustic properties of speech and disrupt a witness's ability to hear certain words accurately. The acoustic confusability of /f/ and /θ/ and the potential to mistake facewear effects for TH-fronting could lead listeners to make false assumptions about speaker dialect and characteristics. Future research on a TH-fronting illusion, or even other salient dialectal features, would give better insight into how facewear interacts with dialect perception. The potential real-word consequences of the acoustic effects found in this study demonstrate the need for further research that will not only broaden scientific understanding of how the physical environment can affect speech style and acoustics but also enable experts in criminal cases to provide accurate testimony.

References

- Bennett, Joe. 2012. 'And what comes out may be a kind of screeching': The stylisation of chavspeak in contemporary Britain. *Journal of Sociolinguistics* 16(1):5–27. <http://dx.doi.org/10.1111/j.1467-9841.2011.00521.x>
- Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Boersma, Paul, and David Weenink. 2009. Praat: Doing phonetics by computer [Computer programme, Version 6.0.09]. Accessed 1 January 2016, <http://www.praat.org>

- Broeders, A. P. A. 2013. Earwitness identification: Common ground, disputed territory and uncharted areas. *International Journal of Speech, Language and the Law* 3(1):3–13. <http://dx.doi.org/10.1558/ijssl.v3i1.3>
- chavspotting. 2004. chav [Urban Dictionary entry]. Accessed 29 March 2016, <http://www.urbandictionary.com/define.php?term=chav>
- Clark, Lynn, and Graeme Trousdale. 2009. Exploring the role of token frequency in phonological change: Evidence from th-fronting in east-central Scotland. *English Language and Linguistics* 13(1):33–55. <http://dx.doi.org/10.1017/S1360674308002852>
- Davies, Mark. 2004. Byu-bnc [Corpus based on the British National Corpus from Oxford University Press]. Accessed, 27 March 2016, <http://corpus.byu.edu/bnc/>
- Elvira-Garcia, Wendy. 2014. Zero crossings and spectral moments [Praat script v.1.3.]. Accessed, 26 February 2016, <http://stel.uab.edu/labfon/sites/default/files/zero-crossing-and-spectral-moments13.praat>
- Fecher, Natalie. 2014. Effects of forensically-relevant facial concealment on acoustic and perceptual properties of consonants. Doctoral dissertation, University of York.
- Flege, James Emil, Samuel G. Fletcher, and Abdullah Homiedan. 1988. Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data. *The Journal of the Acoustical Society of America* 83(1):212–228. <http://dx.doi.org/10.1121/1.396424>
- Fraser, Helen. 2014. Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science. *Language and Law/Linguagem e Direito* 1(2):5–21.
- Fraser, Helen, Bruce Stevenson, and Tony Marks. 2011. Interpretation of a crisis call: The persistence of primed perception of speech. *The International Journal of Speech, Language and the Law* 18 (2):261–292. <http://dx.doi.org/10.1558/ijssl.v18i2.261>
- Garnier, Maeva, Nathalie Henrich, and Daniele Dubois. 2010. Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research* 53(3):588–608. [https://doi.org/10.1044/1092-4388\(2009\)08-0138](https://doi.org/10.1044/1092-4388(2009)08-0138)
- Gold, Erica, and Peter French. 2011. International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law* 18(2):293–307. <http://dx.doi.org/10.1558/ijssl.v18i2.293>
- Hardcastle, William J., John Laver, and Fiona E. Gibbon. 2010. *The Handbook of Phonetic Sciences*. Chichester: John Wiley & Sons.
- Laub, Cindy E., Lindsey E. Wylie, and Brian H. Bornstein. 2013. Can the courts tell an ear from an eye: Legal approaches to voice identification evidence. *Law and Psychology Review* 37:119–158.
- Lennes, Mieta. 2002. Save labeled intervals to wav sound files [Praat script]. Accessed 29 March 2016, http://www.helsinki.fi/~lennes/praat-scripts/public/save_labeled_intervals_to_wav_sound_files.praat
- Levon, Erez, and Sue Fox. 2014. Social salience and the sociolinguistic monitor: A case study of ing and th-fronting in Britain. *Journal of English Linguistics* 42(3):185–217. <http://dx.doi.org/10.1177/0075424214531487>
- Lindblom, Björn, James Lubker, and Thomas Gay. 1979. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7(2):147–161. <http://dx.doi.org/10.1121/1.2016039>
- Ling, Jonathan, and Allison Coombe. 2005. Age effects in earwitness recall of a novel conversation 1. *Perceptual and Motor Skills* 100(3):774–776. <http://dx.doi.org/10.2466/pms.100.3.774-776>
- Llomas, Carmen, Philip Harrison, Damien Donnelly, and Dominic Watt. 2008. Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers in Linguistics (Series 2)* 9:80–104.
- Martin, A., Howell, K., and Lower, M. 1976. *Hearing Protection and Communication in Noise (Vol. 2)*. London: Academic Press London.
- Martinez, Lauren, and Barney Lerten. 2016. Ski-masked gunman holds up NE Bend bank, gets away. Accessed 29 March 2016, <http://www.ktvz.com/news/skimasked-gunman-robs-northeast-bend-bank-gets-away/38254268>
- McFarland, David H., and Shari R. Baum. 1995. Incomplete compensation to articulatory perturbation. *The Journal of the Acoustical Society of America* 97(3):1865–1873. <http://dx.doi.org/10.1121/1.412060>
- Miller, George A., and Patricia E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America* 27(2):338–352. <http://dx.doi.org/10.1121/1.1907526>
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1):62–85. <http://dx.doi.org/10.1177/0261927x99018001005>
- Ohala, John J., and Haruko Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology* 1:113–127. <http://dx.doi.org/10.1017/s0952675700000312>
- Podlubny, Ryan G., Jen Hay, and Megan McAuliffe. 2016. Car-talk: How physical environment influences speech production and perception. Full abstract of paper presented at LabPhon15. http://www.labphon.org/labphon15/long_abstracts/LabPhon15_Revised_abstract_82.pdf
- R Core Team. 2015. R: A language and environment for statistical computing [Computer software manual]. Accessed 15 March 2016, <https://www.R-project.org/>
- Radonovich Jr., Lewis J., Robert Yanke, Jing Cheng, and Bradley Bender. 2009. Diminished speech intelligibility associated with certain types of respirators worn by healthcare workers. *Journal of Occupational and Environmental Hygiene* 7(1):63–70. <http://dx.doi.org/10.1080/15459620903404803>
- Saigusa, Julie. 2016. Jane Lynch and /s/: The effect of addressee sexuality on fricative realization. *Lifespans and Styles* 2(1):10–16. <http://dx.doi.org/10.2218/lis.v2i1.2016.1426>
- Schleef, Erik, and Michael Ramsammy. 2013. Labiodental fronting of /θ/ in London and Edinburgh: A cross-dialectal study. *English Language and Linguistics* 17(01):25–54. <http://dx.doi.org/10.1017/s1360674312000317>
- Sommer, Henry C. 1976. Speech communication capability and hearing protection of USAF inflight headgear devices (Technical Report, DTIC). <http://www.dtic.mil/docs/citations/ADA029007>

- Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17:3–45.
- Tabain, Marija. 1998. Non-sibilant fricatives in English: Spectral information above 10 kHz. *Phonetica* 55(3):107–130. <http://dx.doi.org/10.1159/000028427>
- Tabain, Marija, and Catherine Watson. 1996. Classification of fricatives. In *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*, ed. P. McCormack and A. Russell, 623–628. Canberra: Australian Speech Science and Technology Association.
- Thomas, Erik R. 2002. Sociophonetic applications of speech perception experiments. *American Speech* 77(2):115–147. <https://doi.org/10.1215/00031283-77-2-115>
- Tufts, Jennifer B., and Tom Frank. 2003. Speech production in noise with and without hearing protection. *The Journal of the Acoustical Society of America* 114(2):1069–1080. <http://dx.doi.org/10.1121/1.1592165>
- Watt, Dominic 2013. Research methods in speech acoustics. In *The Bloomsbury Companion to Phonetics*, ed. M. J. Jones and R.A. Knight, 79–97. London: Bloomsbury.
- Wells, Gary L., and Elizabeth A. Olson. 2003. Eyewitness testimony. *Annual Review of Psychology* 54(1):277–295.
- Wittum, Kelsi J., Lawrence Feth, and Evelyn Hoglund. 2013. The effects of surgical masks on speech perception in noise. In *Proceedings of Meetings on Acoustics* 19, 060125. <http://dx.doi.org/10.1121/1.4800719>
- Yarney, Daniel A. 2007. Earwitness descriptions and speaker identification. *International Journal of Speech, Language and the Law* 8(1):113–122. <http://dx.doi.org/10.1558/ijssl.v8i1.113>

jssaigusa@gmail.com

Appendix

Wordlist

deaf death
fie thigh
fin thin
first thirst
fought thought
free three
frill thrill
oaf oath
reef wreath
fan van
fast vast
fault vault
fear veer
fend vend
Fife five
file vile
grief grieve
safe save
strife strive
waif waive
wife's wives