Shifting Sibilants: Spectral Sensitivity of the Lombard Effect in L1 English Speakers

Andrew Lubanszky

Prof. Chandan Narayan

Major Research Paper

Graduate Program in Linguistics

York University

Toronto, Ontario

August 2025

Abstract

This paper examines the spectral sensitivity of the Lombard Effect in L1 English speakers with respect to sibilant consonants. Though previous literature has examined Lombard Speech produced under various noise conditions, no study to date has examined a 'second-order' Lombard effect by eliciting speech under speaker-specific noise designed to approximate a speaker's typical Lombard Speech—that is, speech already produced under broadband noise conditions. Doing so elucidates possible different manifestations of the Lombard Effect and how such a noise condition might elicit speech which is differentiable from broadband Lombard Speech.

Results from spectral analysis indicate a weak-to-moderate sensitivity to the spectral characteristics of noise in the majority of speakers in this study, as demonstrated by adjustments to speech characteristics in both the temporal and frequency domains. Notable changes from broadband to speaker-specific noise include raising of peak frequency, increased segment duration, and decreased skewness. In conjunction, these changes do indeed differentiate speech between noise conditions in those speakers who first demonstrated acoustic changes from quiet to broadband noise conditions. Degree and manner of change in the frequency domain are highly speaker-specific, however, and occasionally segment-specific.

These findings may suggest that acoustic augmentations which differentiate the speech signal from tailored noise may be accompanied by articulatory changes meant to increase segment stability. Further work is needed to verify articulatory changes as well as the statistical significance of spectral trends for other measures across different noise conditions.

Acknowledgements

My dearest thanks to those who have contributed their time and effort to help in completing this project. In particular, many thanks to Professor Chandan Narayan for his assistance with both the technical and the theoretical aspects of this project, without whom this work would not have been possible— supervising a project such as this can be just as tedious as actually writing it, and I thank him for his patience and commitment to the crafting of this paper. Thanks also to Dr. Thomas Kettig, who not only laid the foundations for my understanding and interest in acoustic phonetics but for his assistance in selecting and refining the topic for this major research paper.

I would also like to extend my gratitude and congratulations to fellow members of the Linguistics Graduate Program at York University, especially Makaila and Jacob, who have been invaluable not only to the writing of this paper but have been a source of camaraderie, encouragement, and immense joy throughout the year.

I must also extend gratitude to my family for all of their support over the years, without whom I surely could not have gotten this far. Your support has made all the difference in the world. I must also, of course, thank my beautiful wife for her support, encouragement, and love throughout our years together. Not only have you made this project possible, you have made me happier than I could ever wish.

Lastly, to the coffee farmers of the world— thank you for your service. Everyone knows academia is built standing on the shoulders of giants, but some of us must also stand on coffee beans.

Table of Contents

1 Introduction & Background

1.1 Speech in Noise

It is well documented that speech produced in noisy conditions varies both qualitatively and quantitatively compared to speech produced in quiet conditions. This phenomenon, the Lombard Effect (LE), was documented by the eponymous otolaryngologist Etienne Lombard in 1911 (Lombard, 1911) and has since been studied in numerous languages including English, French, Dutch, Hindi, Arabic, Korean, and others (Junqua, 1996). It has also been found as a reflex in other species besides humans, including songbirds, monkeys, some species of frog, among others (Hotchkin & Parks, 2013; Kunc et al., 2022). Though there are nuances to the types and degree of acoustic changes employed by speakers of different languages, the general parameters of the LE remain the same cross linguistically; primary changes include increasing speaking amplitude and segment duration, increasing pitch (F0), and raising of first (F1) and second (F2) formants (Lane & Tranel, 1971; Summers et al., 1988; Junqua, 1993, 1996; Lu & Cooke, 2009; Stowe & Golob, 2013; Reilly, 2020). More subtle, secondary changes have also been observed under different experimental conditions, though the specifics of the LE remain highly speaker-dependent and sensitive to environmental conditions (Garnier & Henrich, 2014). It is generally understood, however, that the LE exists as an automatic and unconscious vocal reflex in the presence of noise which aims to increase intelligibility of speech. This study is concerned with the spectral sensitivity of the LE in relation to sibilant consonants /s ʃ z/ produced by L1 English speakers under different noise conditions. In particular, it builds on previous studies aiming to determine ways in which the LE may be anticipatory of acoustic masking with respect to individual consonants.

1.2 Sibilants

Sibilant consonants are a fairly common cross-linguistic segment class defined by continuant turbulent airflow, resulting in sustained high-frequency noise (Kokkelmans, 2021). English has six sibilant consonants, four of which are fricatives and two or which are affricates (American Speech-Hearing-Language Association). The sibilants are further divided into voiced and voiceless classes as shown in the diagram below, where voiceless segments are on the left and voiced counterparts are on the right.

	Alveolar	Postalveolar
Fricative	s z	ʃ ʒ
Affricate		t͡ʃ d͡ʒ

Table 1. English sibilant consonants

The term sibilant comes from the Latin *sibilans*, meaning 'hissing.' Though there are many sibilant consonants with diverse articulations including dental, alveolar, postalveolar, alveopalatal, and palatoalveolar, they are all coronal consonants which divert an airstream through a narrow opening between the tongue and roof of the mouth (Kokkelmans, 2021). The tongue body may be grooved, palatalized, retroflex or in some combination. The airstream then strikes the bottom teeth to create turbulence and resonantes in the oral cavity between the teeth and tongue. That the airstream strikes an obstacle, the teeth, at a right angle to the flow of pulmonic air is a defining articulatory characteristic of sibilants more generally (Shadle, 1985, 1991). The English /s/, for example, is generally articulated with an apical constriction at the alveolar ridge. This results in a high-pitched hissing sound that is perceptually very salient and perceived as high intensity (Stevens, 2000). A postalveolar /ʃ/ involves a more retracted or posterior tongue position than the alveolar counterpart which both increases the space in front of

the constriction and creates a sublingual cavity. The increased size of the prelingual space and sublingual cavities both contribute to a perceptually lower pitch sibilant compared to alveolar productions due to greater resonance resulting from increased volume in the oral cavity; increasingly retracted articulations result in lower and lower average frequencies and therefore perceptually lower pitched consonants (Stevens 1971: 1185). The voiced alveolar fricative /z/ is articulatorily similar to the voiceless counterpart, albeit with the added voicing parameter. Oral articulation is the same while the vocal folds open and close in a regular periodic manner to create voicing, though voiced frication more generally is relatively rare cross linguistically. This may be due to the aerodynamic constraint of maintaining high velocity air flow across two points of constriction the glottis and the oral constriction. The voiced postalveolar /ʒ/ is not only the least common sibilant in English but also the least common phoneme in English as a whole; phonological distribution is very restricted, making it difficult to select target words of consistent form if this segment is included in the analysis (Hayden, 1950). For example, it has no dedicated grapheme and appears frequently as a coda consonant within borrowed words such as *massage* [məsɑʒ], *espionage* [ɛspijənɑʒ], etc. or as the product of [j]-coalescence in words such as *pleasure* [plɛʒɚ] *measure* [mɛʒɚ] and *treasure* [tɹɛʒɚ]/[t͡ʃɹɛʒɚ]. Sibilant affricates are slightly more complex in that they are first articulated with a complete closure in the oral cavity before constriction weakens and releases airflow in a typical sibilant manner. Sibilant affricates are sometimes considered sibilants with a delayed release (as in Chomsky & Halle, 1968). Though they are also eligible for spectral analysis in the same manner as continuant sibilants /s ʃ z/, they are not a concern of this paper due to articulatory, acoustic, and phonological differences from their fricative counterparts.

1.3 Visualizing Sounds

Sibilants are of particular interest to this study as they are essentially, in an acoustic sense, productions of sustained static noise. This means they are particularly well suited for a spectral analysis as they can be differentiated (within the sibilant class as well as across other natural classes) based on the distribution of acoustic energy to various concurrent frequencies during production. The following oscillograms, or *waveforms*, in Figures 1 and 2 represent acoustic energy as a function of intensity (*y-axis)* over time (*x-axis)*. Important to note in the oscillograms of voiceless segments is the lack of periodicity in the waveform, which corresponds to the aforementioned turbulence in the airstream created by striking the obstacle; this is characteristic of fricatives as a class more broadly (Kokkelmans, 2021). The spectrograms directly underneath each oscillogram provide a visual representation of the aforementioned energy distribution, where time and frequency are shown on the x- and y-axis, and darkness represents the relative intensity of sound at that particular frequency.

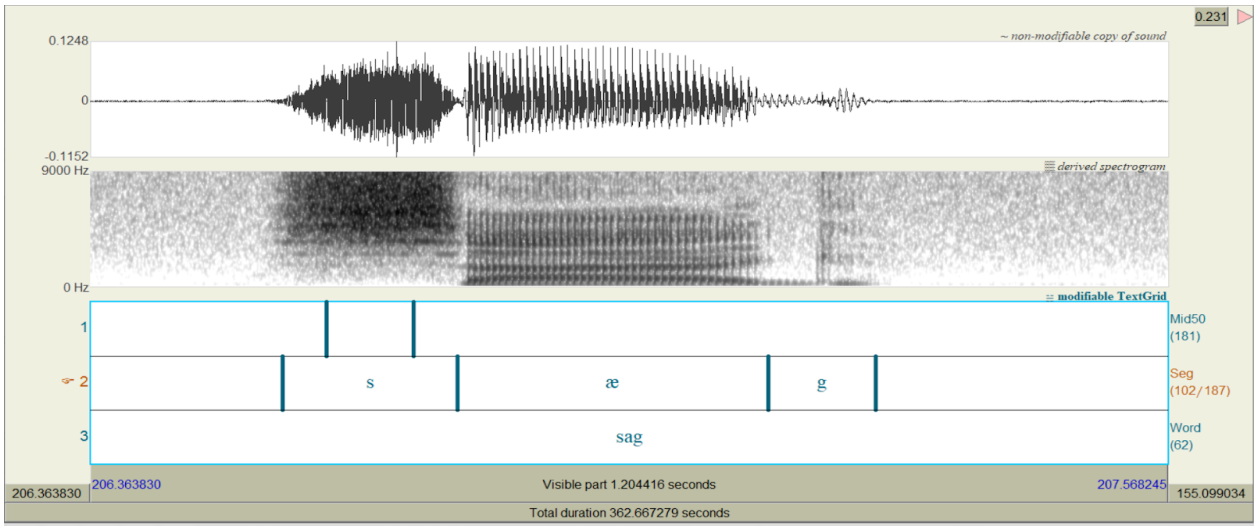


Figure 1. Oscillogram and spectrogram of one participant's pronunciation of /s/ in 'sag' while listening to white noise (C2) as shown in Praat.

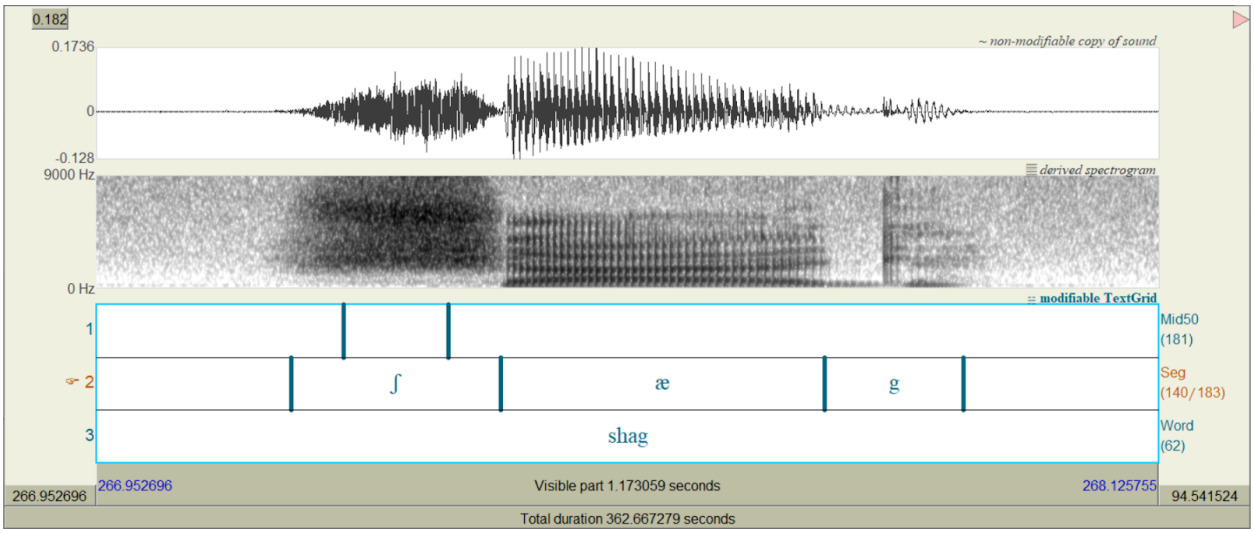


Figure 2. Oscillogram and spectrogram of one participant's pronunciation of /ʃ/ in 'shag' while listening to white noise as shown in Praat. Two dark horizontal bands can be seen in the spectrogram below the postalveolar sibilant.

Visualizing the sound with a spectrum eliminates the time dimension entirely and displays the frequency distribution at a single point in time or as an average across an entire segment. In the following diagrams, frequency (in Hz) is represented as a function of intensity (in dB), where the highest peaks represent the frequencies at which most acoustic energy is concentrated. These peaks can be understood as the dark bands seen in the previous spectrograms (Fig. 2), where a top-down view of the spectrum yields the spectrogram visualization shown previously (in this perspective, 'closeness' to the viewer corresponds with greater darkness). The spectrums also show that the majority of the high amplitude energy in the voiceless alveolar /s/ is concentrated to a single peak and placed higher in the frequency spectrum compared to the postalveolar /ʃ/. The postalveolar sibilant has energy which is more dispersed, yielding a more 'flat' distribution and reflecting the prelingual space discussed previously (see §1.4 for discussions on spectral shape). These visual representations are invaluable to understanding frequency distributions during sibilant productions and are crucial for analysis of spectral characteristics.

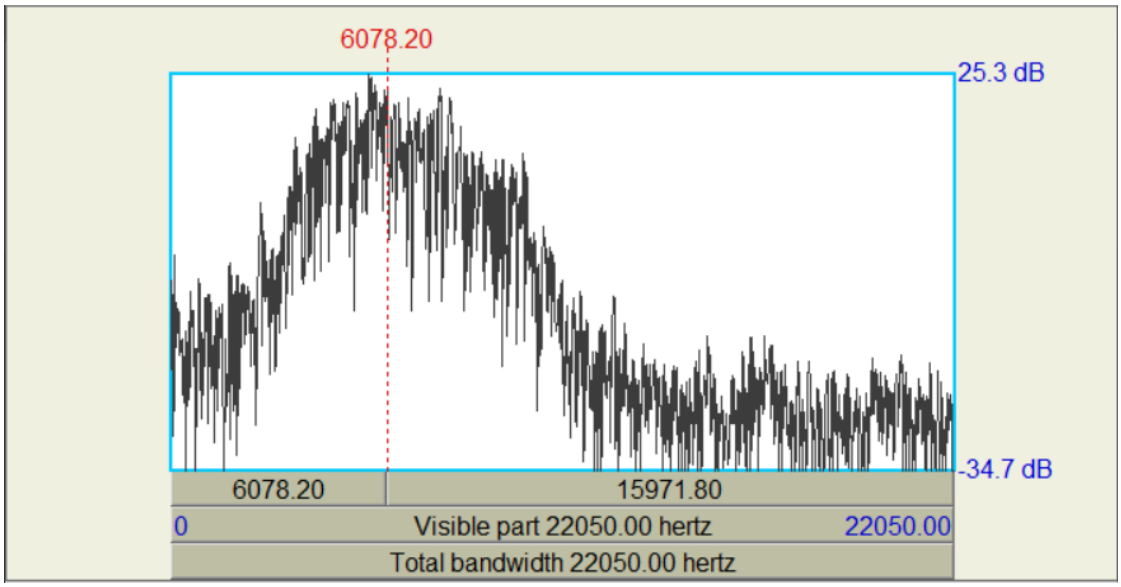


Figure 3. Spectrum derived from the middle 50% of the voiceless alveolar sibilant /s/ shown in Fig. 1.

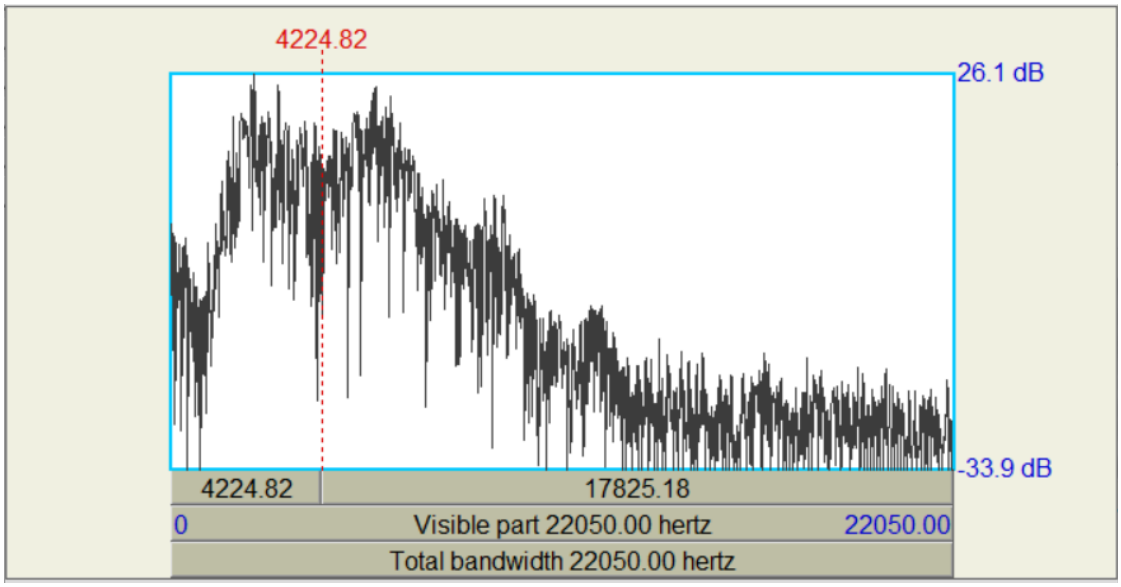


Figure 4. Spectrum derived from the middle 50% of the postalveolar sibilant /ʃ/ shown in Fig. 2. The two highest peaks on either side of the red line correspond to the two dark bands in the Fig. 2 spectrogram.

1.4 Spectral Moments

Sibilants can be quantitatively described and differentiated based on their acoustic characteristics. There are four such measurements which have been used to differentiate both the place and manner of articulation of different sibilants, namely the first four *spectral moments*: Centre of Gravity (CoG), standard deviation (SD), skewness, and kurtosis (Forrest et al., 1988; Jongman et al., 2000). They may also be referred to as *Moment 1* or *M1, Moment 2* or *M2,* etc. Though there is currently work on improving their efficacy and discriminatory power, spectral moments are well established measurements in the field of acoustic phonetics and are considered reliable means of quantifying spectral characteristics of fricative consonants more broadly. CoG is the first and most prominent, describing the main frequency (in Hz) around which most of the acoustic energy is centred and therefore is an overall average frequency of the segment. CoG is calculated "by weighing the frequencies in the spectrum by their power densities," (Boersma and Hamann, 2008: 229). For the voiceless alveolar /s/, Adult males tend to have a CoG in the range of 5000-6000 Hz and females/children tend to have a CoG exceeding 6500 Hz (Jongman et al., 2000; Fox & Nissen, 2005; Stuart-Smith, 2020), though various coronal articulations yield typical CoG values anywhere between 2000-9000 Hz (Boersma & Hamann, 2008). CoG is marked on the spectrums in Figures 3 and 4 with a vertical dotted red line.

Unfortunately, CoG alone is not sufficient to categorize and differentiate sibilants within and across consonant classes; spectral shape is also important as sibilant consonants tend to have very distinct peaks in the spectrum as seen in Figures 3 & 4. These peaks represent particular frequencies at which energy is most concentrated and can be further quantified with measures such as skewness (3rd spectral moment). Skewness describes the degree to which the frequency

distribution ‘leans’ to one side or the other, as opposed to a distribution of 0 which corresponds to a perfect bell curve/normal distribution.

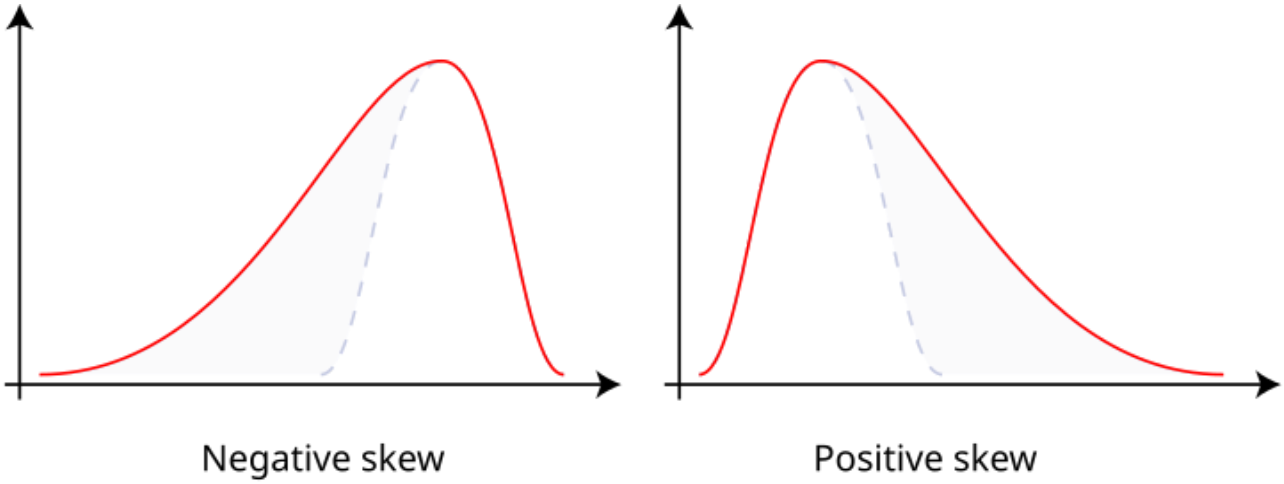


Figure 5. Negative and positive skewness in a normal distribution. Picture taken from Wikimedia Commons (Wikimedia: Rodolfo Hermans, 2008).

Sibilants, with higher frequencies having more concentrated energy and greater intensity than concurrent lower frequencies, often show negative skewness, though this is not a universal; previous studies have found voiceless alveolar sibilants consistently show a negative skewness while voiceless postalveolars consistently show a positive skewness (Haley et al., 2010). It remains that skewness is a useful quantification within individual speakers but should not be taken as an objective indicator of place of articulation independent of other measures. Figures 3 and 4 shown spectrums with positive skewness.

Standard deviation, or 2nd spectral moment, is calculated by finding the square root of the variance, which itself simply describes how spread out or *dispersed* values are from the average mean. A high standard deviation indicates values are widely dispersed around the mean ie. not tightly clustered about the centre. Sibilants tend to show low standard deviation as the acoustic energy, as mentioned, is highly concentrated at specific frequencies. Kurtosis, as the 4th spectral moment, describes the ‘peakedness’ of the distribution, where negative values

correspond with a very flat spectrum and positive kurtosis corresponds to distinct, high peaks (Forrest et al., 1988; Jongman et al., 2000).

The final measure of importance to this experiment is peak frequency, which is the particular frequency at which energy is most concentrated. CoG, on the other hand, is an average measure which is derived from the weighted energy distribution (relative intensities) of frequencies across the entire spectrum. Peak frequency and CoG are usually similar but not exactly the same, with CoG providing a more holistic representation of the segment while peak frequency is more instantaneous. In Figure 3, peak frequency is 5586 Hz, slightly below the marked CoG at 6078 Hz. In Figure 4, peak frequency is 2340 Hz, below the marked CoG at 4224 Hz.These five measures do well to differentiate different sibilants both within and across speakers and will be used as the primary means of assessing sibilant pronunciations obtained in this experiment.

2 Literature Review

2.1 Automaticity

A number of previous studies have pointed to the automaticity of the LE. One study by Stowe and Golob (2013) used a picture naming task and found the LE was manifest under broadband noise (0.2-20 kHz) which contained frequencies typical of human speech; expectant changes to pitch, intensity, and raising of formant frequencies were all found. When the same broadband noise had speech-like frequencies in the range of 0.05 - 4 kHz removed (notched noise), the acoustic changes typical of LS were not found (Stowe & Golob, 2013). The researchers also tested the effect of bandpass noise containing exactly these frequencies and found a lesser effect than was manifest under a broadband noise condition, though there is little discussion of why this

is the case (Stowe & Golob, 2013). This study speaks to an important characteristic of the LE as a vocal reflex which is sensitive to the type of environmental noise under which speech is produced and the presence of speech-typical frequencies. This sensitivity has been shown by Garnier, Henrich & Dubois (2010) who found the greatest elicitation of Lombard Speech (LS) while participants were exposed to multi-talker babble, also known as cocktail-party noise, compared to types of broadband and bandpass static noise, again suggesting the LE is specifically elicited when the ambient acoustics contain speech-typical frequencies. That the effect is lesser or not manifest at all when speech frequencies are not present in environmental noise suggests that the LE is somehow anticipatory of acoustic masking and seeks to augment certain acoustic and articulatory characteristics in order to further differentiate speech from environmental noise. When no such masking will occur and suprasegmental augmentations are unnecessary, as in the notched noise condition, the LE is not manifest.

Interestingly, Hay et al. (2017) found that elicitation of LS may also be a conditioned behaviour that is influenced by environmental priming. In their study, participants were observed using LS while seated in a car and exposed to typical car noise. LS was also exhibited when participants were seated in a parked car that was not running or turned on at all. Hay et al. ultimately take these findings to suggest that location-specific memories (ie. repeated exposure to ambient car/driving noise while seated in a vehicle) can trigger location-specific speech behaviours. In conjunction, these studies suggest that the LE is indeed automatic and dependent on the type of ambient noise, whether or not such noise is even present at the time of speech production.

Other studies pertaining to sound immersion have found a greater elicitation of the LE when participants heard noise through headphones as compared to ambiently through speakers.

A study by Garnier, Henrich and Dubois (2010) found noise in headphones, communicative burden, and multi-talker babble to all directly increase elicitation of the LE compared to broadband noise played ambiently through speakers (Garnier et al., 2010). They also found that, “adding a self-monitoring feedback into headphones reduced this effect but did not completely compensate for it,” (Garnier et al., 2010). Though the authors do not speculate, it may suggest that suprasegmental augmentation strives to produce speech which matches the perceived acoustic characteristics of speech-in-quiet from the perspective of the speaker themselves. Based on previous literature, it seems an unstated assumption that the LE is not goal oriented (at least not to one which is universal across speakers). Rather, acoustic and articulatory changes simply result from the speaker avoiding acoustic masking in any way which is feasible. Whether this avoidance involves, say, a consistent and quantifiable augmentation above the environmental noise (ie. raising F0 10% above the perceived main frequency of noise) or a more concrete end point is an unresolved issue. It may hold that LS is an attempt to make speech-in-noise sound to the speaker as their own speech-in-quiet, or it may hold that the LE manifests speech thought to be most perceptible to the listener, whether or not such speech matches the general acoustic parameters of speech in quiet. This requires further examination to verify but does present an interesting step forward in verifying the ultimate “end goal” of the LE as it is manifest under these different conditions, both experimentally and naturalistically.

2.2 Intelligibility

That the LE is an automatic mechanism designed to improve speech intelligibility is strongly supported by findings that content words (ie. agents, objects, locations) show the greatest degree of change when utterances are produced in noise. Patel and Schell (2008) examined productions in noise and found that low to moderate levels of noise affected uniform

changes to speech across an entire utterance. As noise increases, however, The LE becomes most apparent on content words, such as nouns and verbs, which show continued increases to duration and raising of F0 (Patel & Schell, 2008). With greater noise, intensity remains rather uniformly distributed across the entire utterance though segment duration and F0 remain robust primary cues for marking content words. Functional items such as prepositions show the least increases with regards to these measures.

In further support of these findings, previous studies have found that the presence of a communicative burden on speakers increased elicitation of the LE compared to vacuous speech without the presence of a speech partner. This has been shown in experiments requiring interlocutors to participate in communication-based tasks. This was tested by Garnier and Henrich in 2014 and further reinforces the intelligibility-based account with findings that LS is perceived as more intelligible than regular speech (Dreher & O'Neill, 1957; Lu & Cooke, 2008; Maniwa et al., 2008, 2009) This was found when the speech was presented to listeners in isolation but also when presented with the accompanying environmental noise in which it was produced (Garnier & Henrich, 2014). In conjunction, findings such as these strongly indicate that the LE is primarily a communicative mechanism which seeks to improve transmission of semantic information and augment communication as a whole.

2.3 Spectral Sensitivity

Though elicitation of LS is variable both within and across speakers, it does remain that the effect is at least somewhat sensitive to the spectral characteristics of noise. A 2020 study by Reilly examined the spectral sensitivity of first order LS with acoustic filters which shaped static noise into vowel-shaped and sibilant-shaped noise specific to individual speakers. These filters were based on participants' speech in quiet conditions. It was found that sibilant productions in

both noise conditions involved expected increases to CoG, sibilant intensity, and duration, though the largest acoustic changes to sibilant consonants were obtained under the sibilant-noise condition (Reilly, 2020). Reilly found that, “no sibilant changes were specific to the vowel noise condition,” and concluded that, “the sensitivity of sibilant sounds to sibilant noise indicates that speech responses to noise were affected by the frequency content of background noise,” (2020). Spectral variance and skewness were not found to systematically vary in the sibilant noise condition, though kurtosis increased in both the vowel and the sibilant noise conditions (Reilly, 2020). Overall, the experiment indicates that sibilant productions are indeed sensitive to the frequency content of environmental noise and show the greatest acoustic changes when there is spectral similarity between the produced segment and background noise. This correlation can be taken to indicate an attempt at decreased acoustic masking based on spectral overlap of speech and noise.

In contrast, a study by Lu & Cooke (2009) found that the LE can place speech in an area of the acoustic spectrum which actually increases acoustic masking by ambient noise. Their experimental conditions involved participants listening to either high-pass or low-pass filtered noise while reading and producing target words. In both noise conditions, the LE raised intensity, F0, F1, and CoG compared to quiet conditions; this had the effect of decreased masking in the low-pass noise condition but ultimately placed productions into the crowded portion of the spectrum during the high-pass condition and increased acoustic masking (Lu & Cooke, 2009). However, it should be noted that the experimental conditions present a possible confound which may make these results unrepresentative of the nature of the Lombard mechanism more generally. Given that the LE has been only ever observed as a systematic raising of pitch and formants in noisy conditions, it is unsurprising that participants employed these changes in the

high-pass condition. These changes obviously then resulted in greater acoustic masking due to the frequency overlaps between speech and high-pass noise. To take these findings as indicative of a non-anticipatory nature of the LE would perhaps be unfair, as the alternative option of lowering pitch and formants was, as best can be understood, never a possible manifestation of the Lombard mechanism in the first place. Productions which do lower pitch and formants compared to speech in quiet conditions, should they be found, should likely not be considered as part of the LE but rather considered either an intentional and salient decision on behalf of the speaker or attributed to other perceptual-acoustic speech mechanisms.

2.4 Acoustic strategies

Work by Garnier and Henrich in 2014 outlines three possible strategies by which speakers may differentiate their speech from the noise signal, namely *boosting, bypass,* and *modulation* strategies. Boosting involves global increases to vocal intensity, especially in those portions of the acoustic spectrum which is most crowded. This corresponds to global frequency and intensity increases, particularly for those frequencies where noise is maximally energetic. Bypass strategies involve redistribution of acoustic energy to regions where background noise is least energetic. Bypass strategies may be restricted only to those frequency bands of greatest phonetic importance, such as F0, first and second formants, etc. Lastly, modulation involves increasing the temporal modulation of intensity and of F0 (Garnier & Henrich, 2014). In a study designed to evaluate the potential application of these specific strategies, Garnier and Henrich found different patterns for vowels and consonants consistent with previous studies that vowels show the greatest change in noisy conditions (Summers et al., 1988; Hansen, 1988; Lu & Cooke, 2009; among others), whereby duration and intensity increases are greatest for vowels; consonants were actually seen to decrease in duration while vowels increased, possibly in an attempt to

further differentiate the two segment classes. Garnier and Henrich found, in line with Lu and Cooke (2009), that high frequency noise led to a raising of CoG and vowel formants. This is in contrast with other studies such as Mokbel's (1992) finding of specific boosting only in those regions of maximum acoustic energy. Garnier and Henrich instead found female speakers to show greatest boosting in the range of 4-5 kHz and males in the range of 2-4 kHz, with females showing greater CoG increases than males both under broadband and cocktail party noise. Their final conclusions state that speakers appear to adopt increased vocal effort in order to maintain a positive signal-to-noise ratio, where increased vocal effort also corresponds to a flatter spectrum consistent with shouted speech, whether or not background noise is present (Rostolland, 1982; Schulman, 1989; Liénard and Di Benedetto, 1999; Traumüller and Erikkson, 2000). Garnier and Henrich did find specific frequency bypass strategies in use in the cocktail noise condition, however, such as adjusting F0 "so that [speakers'] most frequent $f0$ or $2f0$ value was specifically adjusted in the frequency region where the CKTL noise presented a local minimum in energy" (Garnier & Henrich, 2014). A similar effect was found for F1 frequencies though this was restricted to high vowels [i] and [u]. These findings suggest that a broad understanding of boosting and bypass strategies is likely not fine-grained enough to capture the frequency-specific implementations of these strategies by speakers under different noise conditions, where certain important frequency bands (formants) may show boosting while others (F0) could be considered to show bypass. This is taken to indicate two levels of adaptation in LS, where the primary level is indeed global increases to vocal effort leading to global increases in intensity, flatter spectra, and raised characteristic frequencies (F0 and formants). The secondary level can be understood as the frequency-specific bypassing observed by speakers in cocktail noise. Ultimately, Garnier and Henrich's work paints a picture of the LE that is layered and variable across speakers,

dependent on noise type, sex, and communicative burden, as in previous work (Garnier and Henrich, 2010, 2014).

2.5 Current study

Despite numerous previous studies tailoring ambient noise to unique acoustic characteristics of individuals' speech, no study to date has investigated a 'second-order' LE by tailoring noise to match acoustic characteristics of individuals' Lombard pronunciations, ie. those already elicited under broadband/white noise. The point of this design is to assess a possible anticipatory nature of the LE which, as mentioned, has been shown to primarily function as a method of increasing speech intelligibility. Such design arises from a simple question; if the LE changes acoustic parameters of speech to increase intelligibility, what if ambient noise matches the acoustic characteristics of speech that has already been augmented? What sort of changes do speakers employ then? To employ the same acoustic changes/augmentations as manifest under broadband noise is to then majorly or totally overlap with the new noise condition (C3). Direct overlap leads to major/complete acoustic masking and is likely to directly impact the intelligibility of speech. This is of course counter to the presumed function of the LE as an unconscious and automatic vocal response in the presence of detractive environmental noise which increases speech intelligibility. If the same augmentations are manifest under this second-order condition (C3), it is evidence that the LE is not anticipatory of acoustic masking based on specific spectral characteristics of noise and the effect is not, for lack of a better word, multifaceted in this way. If it is anticipatory, however, speakers are expected to employ one of the aforementioned strategies to avoid increased masking based on the spectral characteristics of the new noise.

3 Methodology

3.1 Participants

A total of 10 L1 English speakers, 5 female and 5 male, participated in the study. Though L1 English was a criteria for eligibility, there was self-reported multilingualism by most participants (Appendix A for age and language distribution). All participants were in the age range of 18-30 years old. An age of 40 was chosen as the maximum age of participation so as to avoid any decreased sensitivity to higher-range frequencies due to age-related hearing decline which might affect perception of noise and therefore affect specific acoustic augmentations/changes in each condition (Lin, 2024). Participants were recruited through various means such as emails distributed to York University undergraduate students by third-parties or through recruitment posters on campus. A $10 inducement was offered for participation in the study.

3.2 Targets

Target words were minimal triplets of the form #SV(C)(C) which contrasted one of three English sibilant consonants /s z ʃ/ in word-initial, prevocalic position. This created minimal sets such as [sip, ship, zip]. Post-sibilant environment was controlled to one of five vowels from the English inventory /i ɪ u ɜ æ/. Participants were also presented with distractor items of the form #$C_{[+continuant]}$VC(C) with the same five vowel nuclei. (Appendix C.1 for target words).

Target words were not presented in carrier sentences so as to avoid any inadvertent focus prosody on the target item. This was crucial considering that English focus prosody typically involves systematic acoustic changes such as increased pitch, intensity and duration on the focused item (Kim & Arnhold, 2024, for example). Given the overlap with changes seen in

typical LS, focus prosody would have affected the acoustic characteristics of the sibilant consonants independently of the actual LE and confounded experimental results.

Unique word lists were created for each participant based on a 30x30 Latin square design. Those orders used in the study were selected if they began with a distractor item; this was to ensure that the very first productions in each recording session were not target items (sibilant-initial words) as participants could modify their speaking volume or speech characteristics as they adjust to the experimental conditions. (Appendix C.2 for sample randomized word lists). Lists presented in C2 were in reverse order from those in C1. The same list given in C2 was presented in C3.

3.3 Experimental Procedure

For all experimental conditions, participants were seated in a sound-proof recording booth on Keele Campus while individual target words were presented on a laptop screen approximately 12 inches in front of participants. Audio was captured using a Sony F-730 microphone placed approximately 4 inches in front of the speaker and was recorded in Audacity(R) v3.5.1 audio software with a sampling rate of 44.1kHz (Audacity Team, 2023). In all conditions, participants were instructed to speak in a normal and comfortable manner, with each word pronounced three times before moving to the next item. In order to impose a communicative burden, participants were instructed that the experimenter would be listening and writing each word as it was read aloud. This was an important condition as previous studies, in line with the intelligibility-based account arising elsewhere, found that word lists and lack of a communicative burden decreases elicitation of LS compared to interlocutory contexts where effective communication is a premium, even under the same noise condition (Garnier & Henrich, 2014).

Condition 1 (C1) involved productions without any ambient noise. Condition 2 (C2) involved productions using the same randomized word list as in C1, though in reverse order, while white noise was played continuously to participants through Audio-Technica ATH-M50x headphones. Broadband noise (random Gaussian) was generated using Audacity audio software with a sampling rate of 44.1 kHz. Before beginning recording, participants were asked to listen to the white noise and set it to their maximum comfortable volume— this determined the volume that was used during C2. All participants stopped at or near the same volume, which was then approximated and standardized to 60dB for all trials. This same volume was then used in Condition 3 (C3). Productions in C2 amounted to typical LS and were analyzed in order to create the acoustic filters for C3 (see §3.4 for creation of acoustic filters and §3.5 for acoustic analysis procedure).

Approximately two weeks after the initial recording session, participants returned to record in C3 in which they produced words in the same manner as C1 and C2. In this condition, participants listened to white noise passed through acoustic filters designed to match their previous Lombard sibilant productions from C2. Filtered noise was coordinated so as to match the target items as they were presented in the randomized lists; /s/-initial words were produced while speakers listened to their own unique /s/-filtered noise, /z/-initial words were produced under that speaker's unique /z/-filtered noise, and likewise for /ʃ/. Distractor items were produced while the speaker heard either white, pink, or blue noise that was randomly assigned to the distractor items. This was to reduce predictability of the noise conditions during productions of target sibilant items. For both C2 and C3 conditions, the same target sibilant never appeared more than twice in succession, meaning the same noise conditions in C3 also never appeared more than twice in succession.

3.4 Acoustic Filters

To create the filtered noise, speaker-specific sibilant productions were grouped by consonant and the middle 50% of each sibilant consonant was concatenated to create three separate audio files. Using the middle 50% of each consonant for analysis and filter creation reduced transition effects such as slow onset of acoustic energy at the initiation of sibilant production or coarticulation effects of the following vowel. For each file, a spectral envelope was created using Praat's built-in LPC generation function. LPC generation used Praat's standard settings (window length of 0.025s, time step of 0.005s, pre-emphasis frequency of 50 Hz). A prediction order of 18 was used as this provided a balanced approximation of the recorded sibilant productions without overfitting and capturing noise, where small amounts of ambient noise during speech are incorporated into the LPC and affect the overall spectral shape.

The generated LPC was converted to a spectrum and then back to sound with a sampling rate of 44.1 kHz. The sound was then convolved with the white noise samples used in C2 productions to create audio files that were 10 seconds in length. This created three unique filters for each speaker for a total of 30 different, sibilant-like static noises.

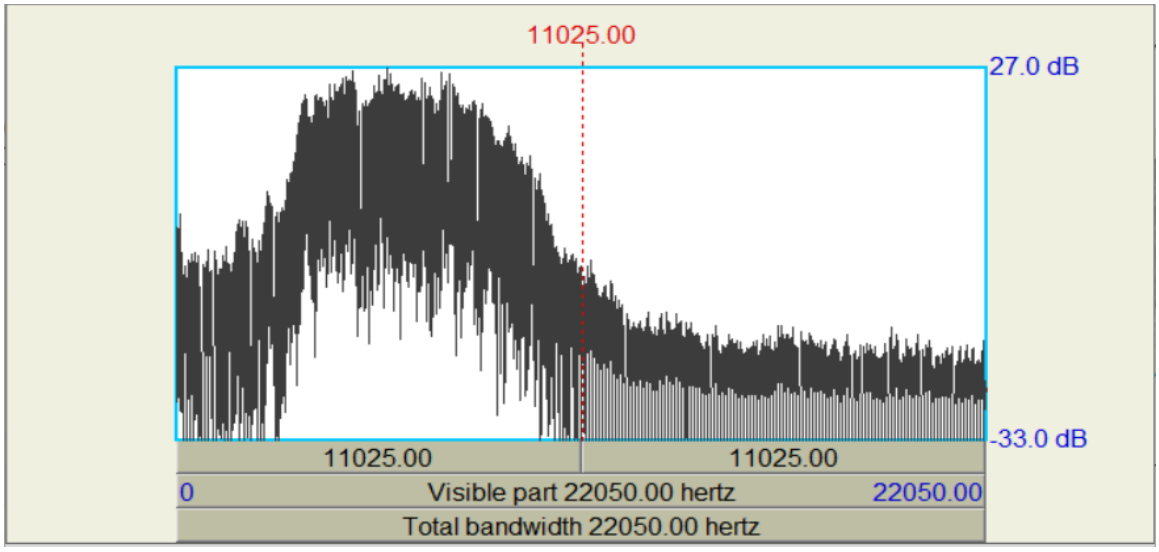


Figure 6. Spectrum of one speaker's concatenated /s/ productions.

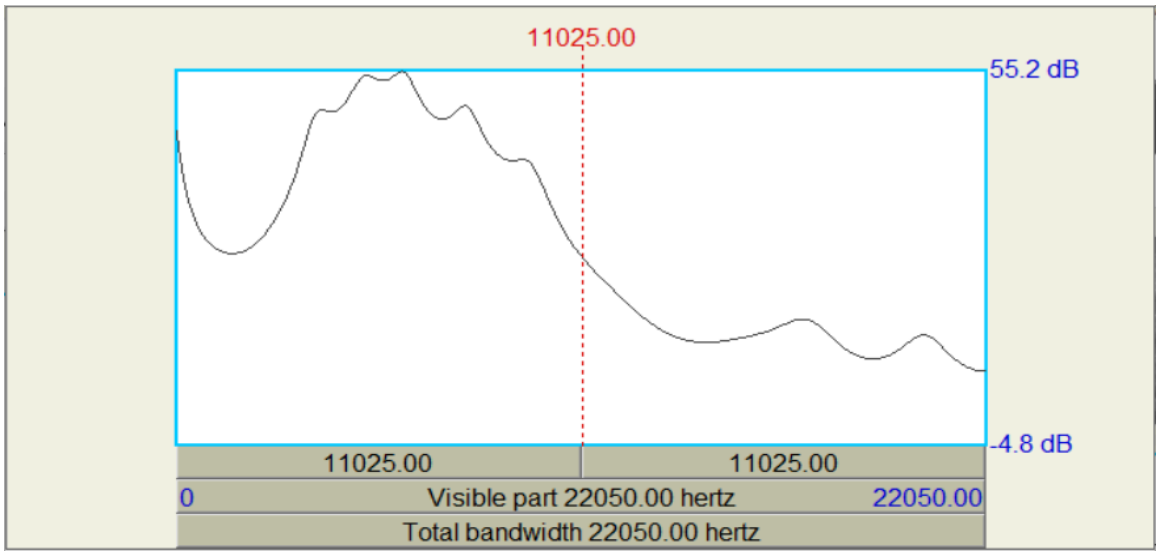


Figure 7. Spectrum derived from LPC of the spectrum shown in Fig. 6.

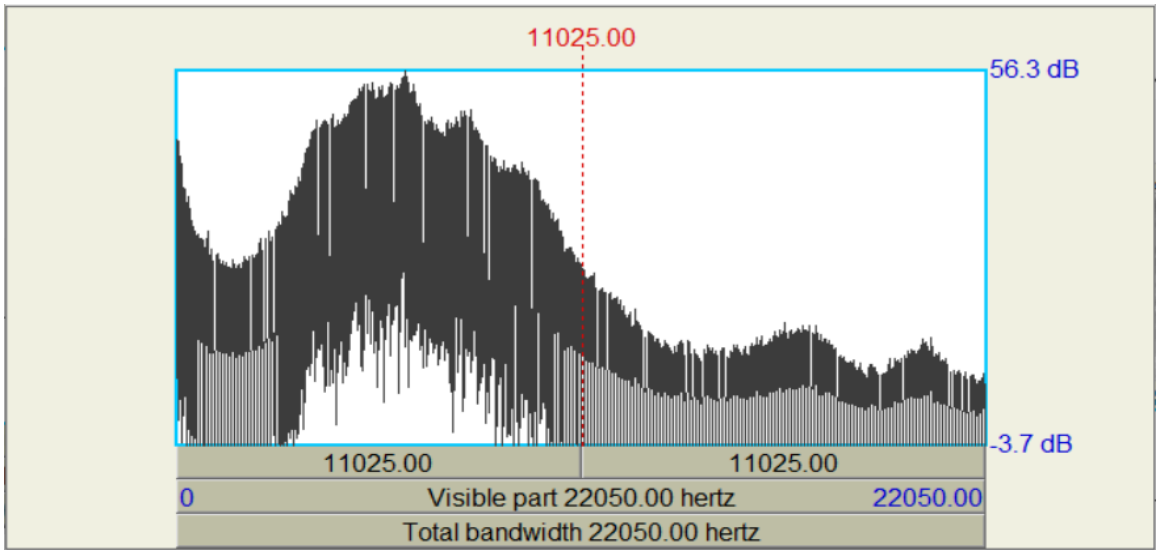


Figure 8. Spectrum of sibilant-shaped noise for one speaker, created by convolving the LPC spectrum from Fig. 7 with white noise

The synthesized noises were then manually volume-corrected by the experimenter using Audacity software to match the perceived volume of the white noise audio. This was to counter biases in human hearing in which perceived loudness does not vary linearly with frequency. The ISO 226 is a standardized equal-loudness contour, revised in 2023, which shows greatest hearing sensitivity to frequencies in the range of 2-5 kHz ie. in the range of typical speech (International

Organization for Standardization [ISO], 2023). As the generated sibilant-shaped noises contained mostly high frequencies above the 2-5 kHz range, the intensity of generated noise had to be increased for all participants in order to match the perceived volume of the white noise used in C2.

To ensure validity of the synthesized audio and faithfulness to participant's average sibilant pronunciations, certain productions from C2 were discarded by the experimenter and not used in the filter-creation process. Discarding was either due to clear mispronunciation of target items, such as the voiceless alveolar fricative /s/ being pronounced as a postalveolar [ʃ], or hesitation by the speaker leading to drastically reduced amplitude or unstable articulation of target items. Of the cumulative 450 tokens produced in the white noise condition across speakers, three were discarded.

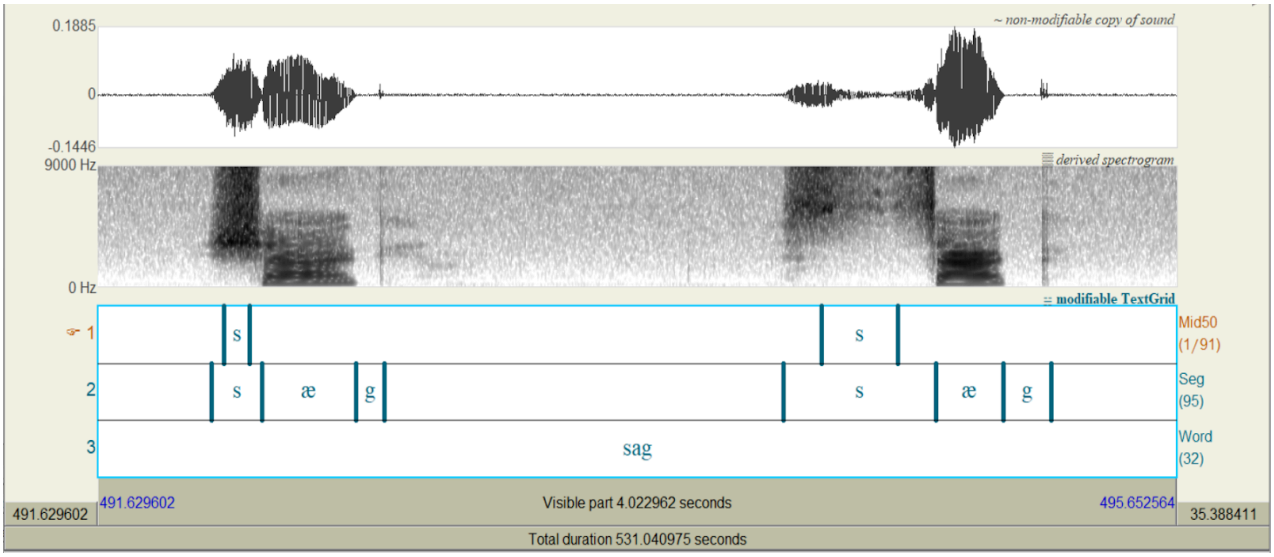


Figure 9. Two discarded pronunciations of 'sag' by one speaker in C2. Sag2.1 was pronounced as a postalveolar /ʃ/ based on the lower CoG consistent with that speaker's pronunciation of the postalveolar sibilant in other target words, as well as the experimenter's perception of the pronunciation. The next token, Sag2.2, shows clear inconsistencies in intensity and frequency across the sibilant segment as evident in both the oscillogram and spectrogram. Perceptually, it seems the participant noted their mispronunciation in Sag2.1 upon beginning Sag2.2 and faltered/hesitated, leading to

unstable pronunciation and correction. Sag2.3 showed a more typical alveolar /s/ pronunciation and was included in the LPC generation process.

3.5 Acoustic Analysis Procedure

To analyze the spectral characteristics of participants' speech, recordings from C2 were loaded in Praat acoustic software v3.5.1 (Boersma & Weenink, 1999) and a textgrid was created to demarcate the beginning and end of each sibilant consonant. Sibilant onset was defined by the initiation of high-frequency acoustic energy as evident in the spectrogram and initiation of aperiodic turbulent airflow as evident in the waveform. Sibilant offset was defined as the initiation of the following vowel, marked by offset/diminishing of high frequency energy in the spectrogram, a loss of aperiodicity and initiation of periodicity in the oscillogram. Using these demarcations, the middle 50% of each sibilant consonant was derived based on the segment's total duration. This was to ensure the transitions in and out of the sibilant consonant were not included in the acoustic analysis. The first four spectral moments, segment duration, and peak frequency were calculated for these middle 50% windows using Praat's in-built measurement mechanics and automated with a Praat script (Appendix E).

An important consideration in the spectral analysis of sibilant consonants is the restriction of the analysis functions above a particular frequency floor. When deriving CoG and peak frequency, for example, the voicing parameter of voiced segments such as /z/ can confound calculations and result in anomalous measurements such as a CoG in the 200-500 Hz range. Based on previous frequency ranges of typical sibilants (Boersma & Hamann, 2008) a floor of 2 kHz was chosen for spectral analysis.

3.6 Statistical Analysis Procedure

Data was analyzed using linear mixed effects models in R with each of the six acoustic measures serving as dependent variables (DV) while segment and condition served as independent variables (IV). Models used the formula DV ~ seg * cond + (1 | sub_rank) in R's lme4 package where sub_rank is a random intercept accounting for repeated measures for individual participants (Bates et al., 2014, 2015). A Type III 2-way ANOVA was used to assess statistical significance with Satterthwaite approximation for degrees of freedom (lmerTest package; Kuznetsova et al., 2017). Post-hoc pairwise comparisons used the Tukey method to adjust for multiple comparisons within each family of three estimates (Tukey 1949) and the Kenward-Roger method was used to calculate degrees of freedom (Kenward & Roger, 1997).

4 Results

Duration showed statistically significant changes across experimental conditions, with the broadband noise condition eliciting greatest duration for all segments. Peak frequency shows a small yet significant change for all segments with the greatest increase under C3. Centre of Gravity and SD did not show statistically significant changes across experimental conditions, suggesting the first and second spectral moments are not particularly mutable under different noise conditions. Skewness showed very significant changes across conditions, with C3 eliciting drastically lower skewness for all segments compared to C1. Kurtosis also showed significant changes across conditions, though it is highly variable with each of the three sibilants showing very different patterns across conditions. Overall, data suggests a spectral sensitivity effect is restricted to peak frequency, skewness, and kurtosis. Important to consider however is the limited sample size, which reduces the authority of significance thresholds. Though pairwise

comparisons may not indicate statistical significance, there are still notable trends in the data which may be more pronounced with larger participant pools, such as the increase in CoG across experimental conditions. Figures 10-16 show Estimated Marginal Means (EMM) for each of the six measures.

4.1 Duration

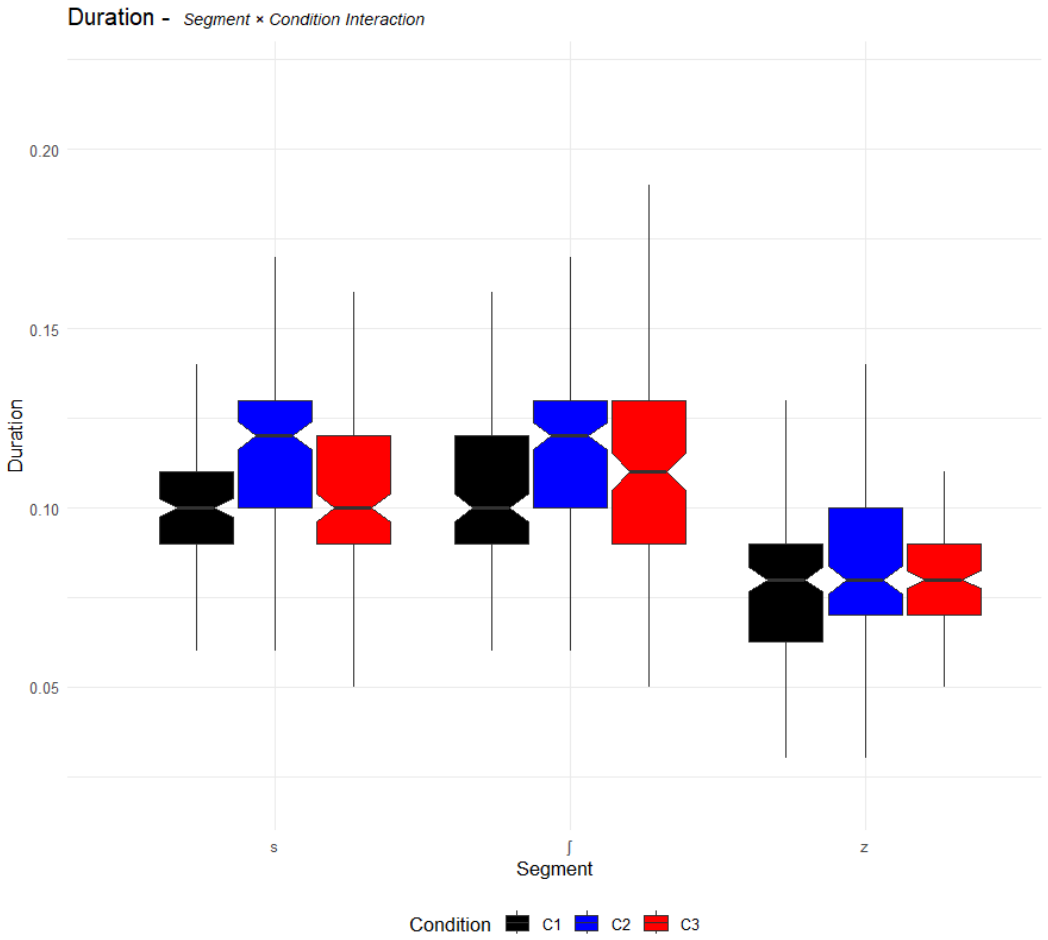


Figure 10. Estimated marginal means (EMM) of duration for target sibilants.

ANOVA indicates a significant effect of condition on duration ($F(2,1327) = 33.168$, $p<0.01$), with C2 showing the greatest duration for all three segments and C1 showing the least. An

interaction effect was also found ($F(4,1327) = 3.1211$, $p<0.05$). Post-hoc tests show that segment /s/ had a significant increase in duration from C1 to C2 ($p<0.0001$) but not in C3. Segment /ʃ/ showed a significant increase from C1 to C2 ($p<0.0001$) as well as from C1 to C3 ($p<0.001$). Segment /z/ did not show a significant change in C2 or C3, suggesting little to no change in duration. The significant increases in C2 for /s/ and /ʃ/ may be due to the experimental procedure in which C1 and C2 were completed consecutively while C3 was completed in isolation. The stark difference between quiet and noisy environments in a single sitting may have led to a dramatic increase in duration from C1 to C2 compared to the lesser increase from C1 to C3. It should be noted that C3 did still show greater duration than the quiet condition, C1, consistent with general findings that ambient noise leads to greater segment durations compared to quiet conditions. Given that such increases to duration are uncontested in the literature, duration was only a supplementary measurement in this experiment, though it remains useful in verifying the presence of temporal effects elicited under different noise conditions.

4.2 Peak Frequency

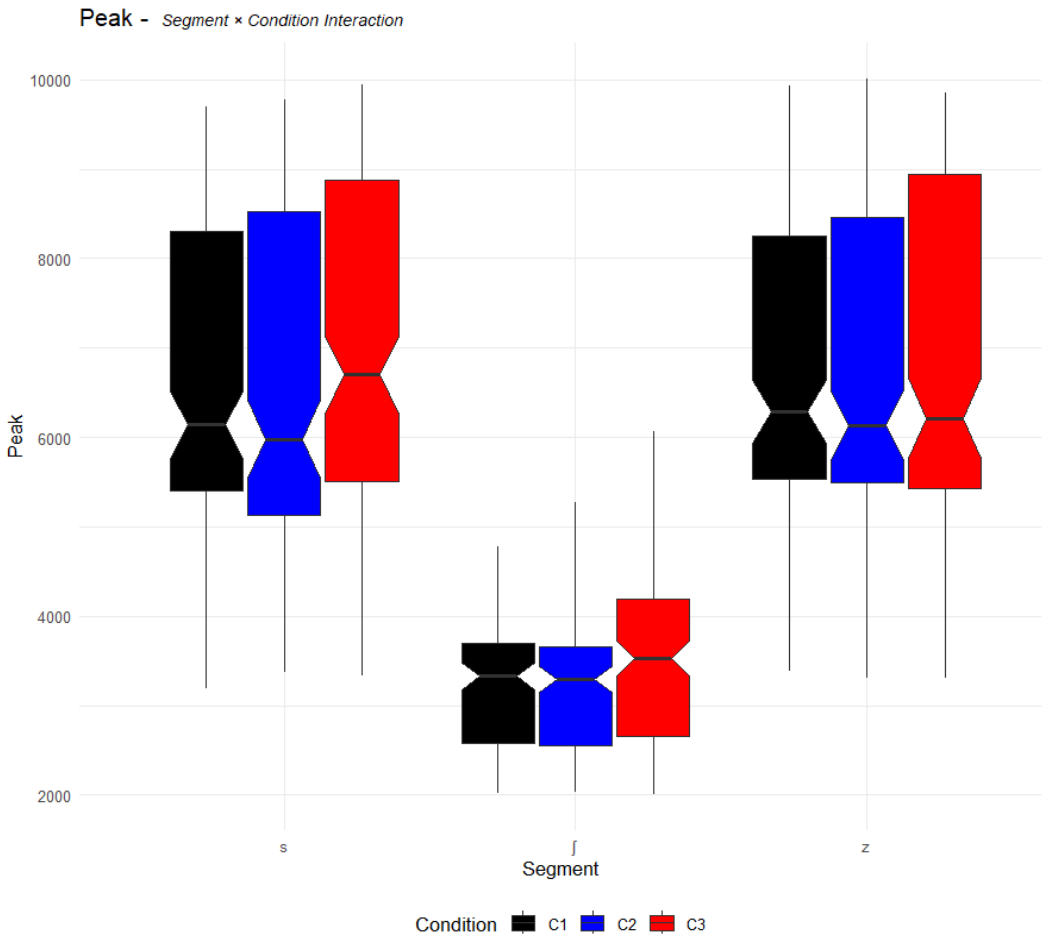


Figure 11. EMM of peak frequency for target sibilants.

Both alveolar sibilants show roughly equal peak frequencies in all conditions which follows from the fact that oral configurations are the same. The voiceless postalveolar has much lower peak frequencies in all conditions which is consistent with a weaker, more posterior oral constriction creating larger prelingual and sublingual cavities. Peak frequency shows a positive trend across conditions for all segments, though with varying degrees of statistical significance. ANOVA indicates that condition had a significant effect on peak frequency ($F(2,1327) = 6.4447$, $p = <0.01$), but no interaction effect was found. Segment /s/ showed a significantly higher peak from

C2 to C3 ($p<0.05$) but not from C1 to C2. Segment /ʃ/ had a significant increase between C1 and C3 ($p<0.05$), though not between C1 and C2 or C2 and C3. The increase in duration from C1 to C3 for segment /z/ approaches statistical significance but does not quite meet the threshold ($p=0.5096$).

4.3 Spectral Moments

4.3.1 Centre of Gravity

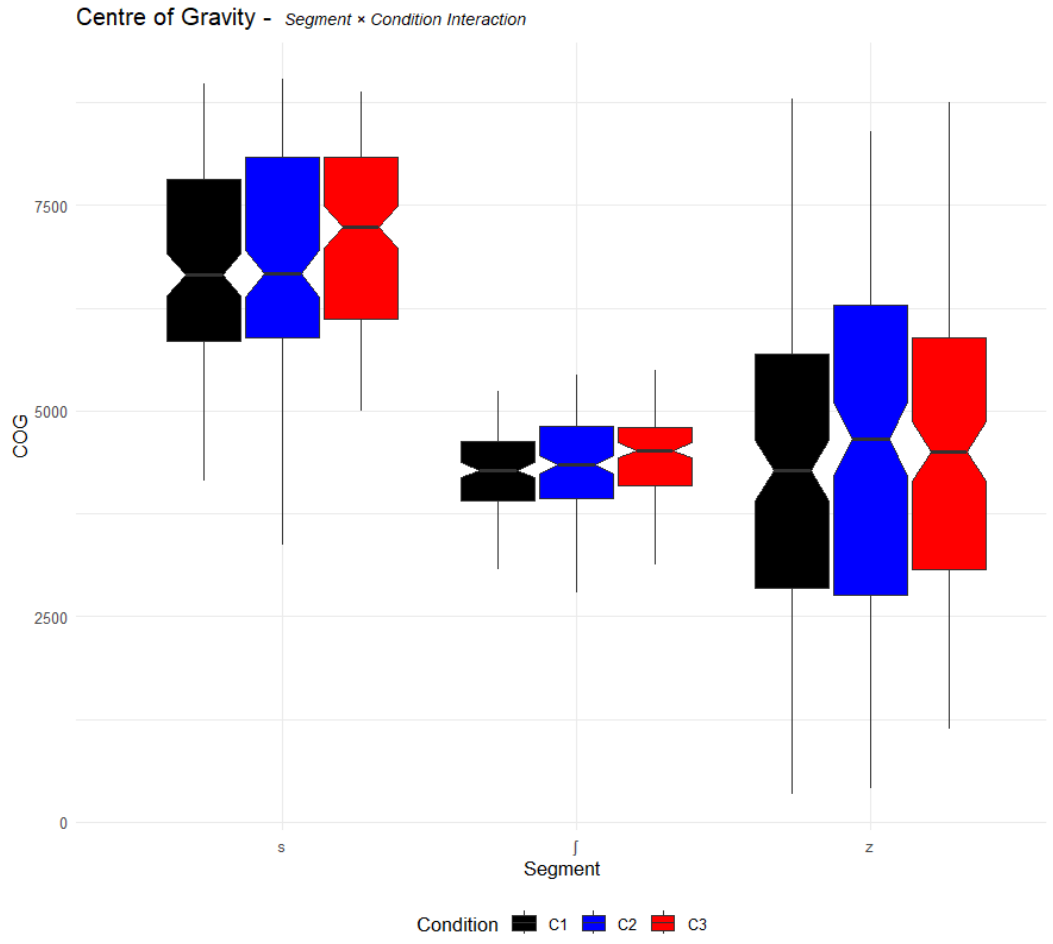


Figure 12. EMM of centre of gravity for target sibilants.

Average CoG for segment /s/ and /ʃ/ are consistent with typical acoustic characteristics whereby the alveolar sibilant is higher than the postalveolar. Segment /z/ has an average CoG near that of /ʃ/ though with the greatest data range of the three sibilants. This is likely the result of the voicing parameter which is variable independently of the high-frequency frication and thus contributes to the broad data range. Statistically, a non-significant raising of CoG was found across experimental conditions, though not uniformly across all segments. The only statistically significant increase was found for segment /z/ between C1 and C2, showing an increase in CoG by 326 Hz ($p<0.05$). C3 did not show a significant increase compared to quiet or broadband conditions for any segment. ANOVA results indicate a significant main effect of condition ($F(2, 1327) = 4.98$, $p<0.01$) though this is not reflected in any pairwise comparisons. No interaction effect was found between segment and condition.

4.3.2 Standard Deviation

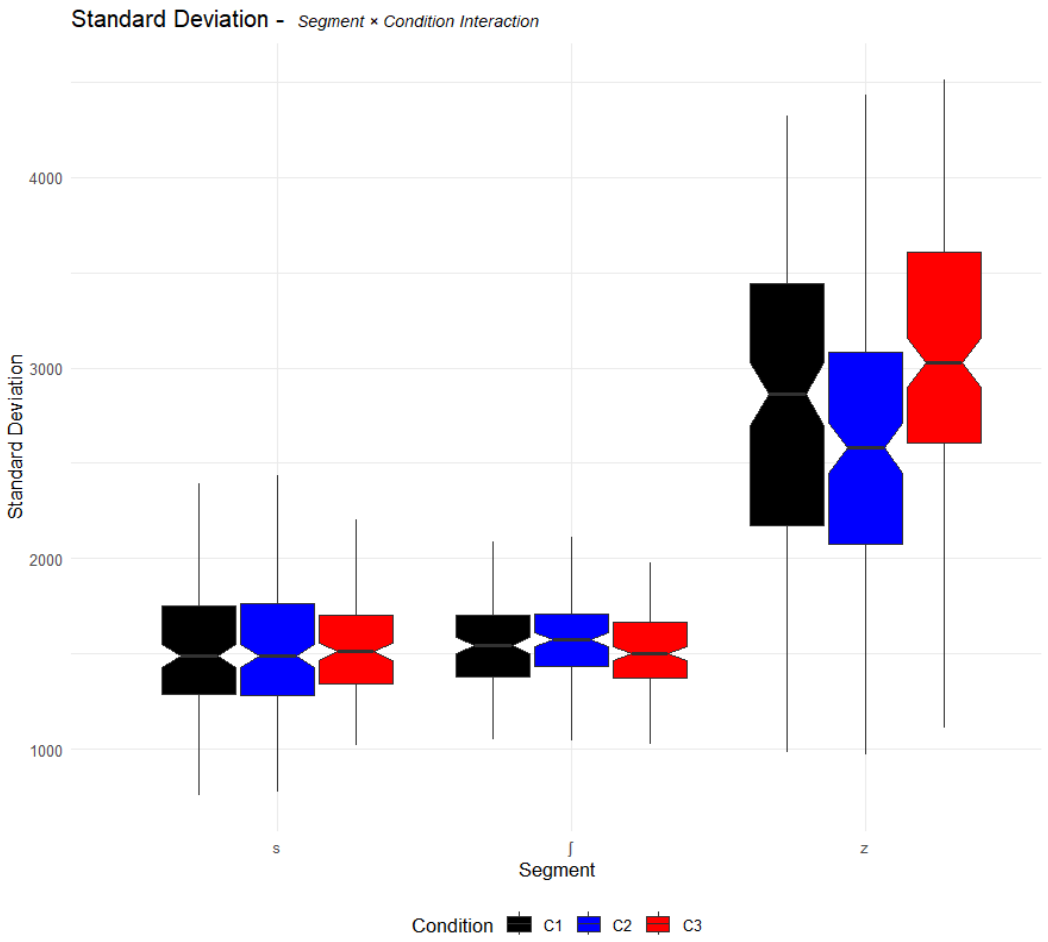


Figure 13. EMM of standard deviation for target sibilants.

Standard deviation was found to be quite variable across conditions, with the voiced sibilant showing a very broad data range. Again, the voicing parameter of /z/ contributes to more variable measurements, with greater SD corresponding to much greater diffusion of acoustic energy about the mean and /z/ showing highly significant changes across all conditions; C2 elicited a lowering of SD by 150 Hz compared to quiet conditions ($p=0.03$) and C3 showed a raising of SD by 238 Hz ($p<0.001$) compared to quiet. Voiceless segments are more consistent, however, with non-significant changes across conditions. ANOVA indicates a significant effect

of condition (F(2,1327) = 6.73, p<0.01) as well as interaction effects (F(4,1327) = 7.78, p<0.001), though post-hoc tests show this is restricted to the voiced alveolar sibilant only.

4.3.3 Skewness

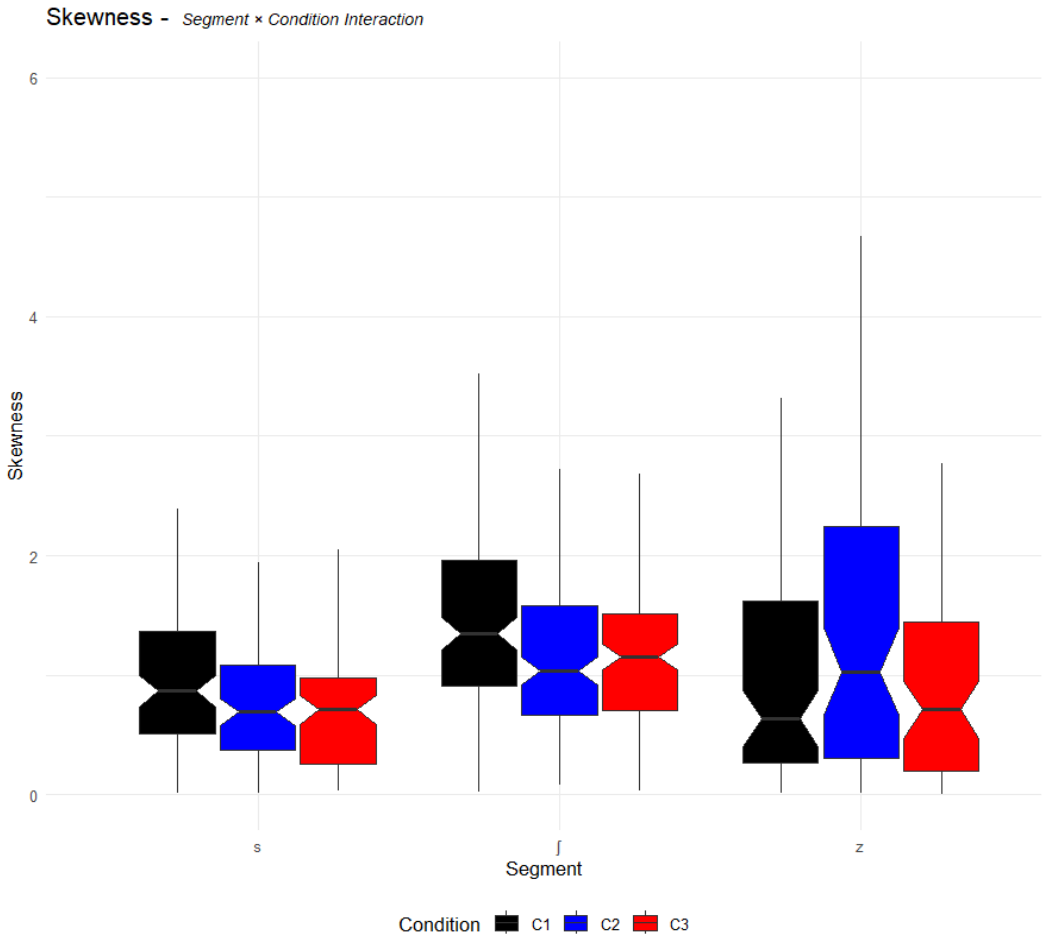


Figure 14. EMM of skewness for target sibilants.

ANOVA revealed that experimental condition had a highly significant effect on skewness for all segments (F(2,1327) = 18.6977, p<0.001) with C3 eliciting significantly decreased skewness compared to quiet conditions. No interaction effect was found (F(4, 1327) = 0.7159, p>0.5). Post-hoc tests revealed segment /s/ only showed significant decrease from C1 to C3 (p<0.005) with C2 not significantly different from C1 (p=0.21). Segment /ʃ/ showed significant decreases

across all conditions, with C2 lowering skewness by 0.2953 ($p<0.05$) compared to C1 and C3 lowering by 0.34 ($p<0.05$) compared to C1. C2 and C3 were not significantly different from each other ($p>0.5$). Segment /z/ showed a significant decrease in skewness from C1 to C3 ($p=<0.0001$) and from C2 to C3 ($p<0.05$).

4.3.4 Kurtosis

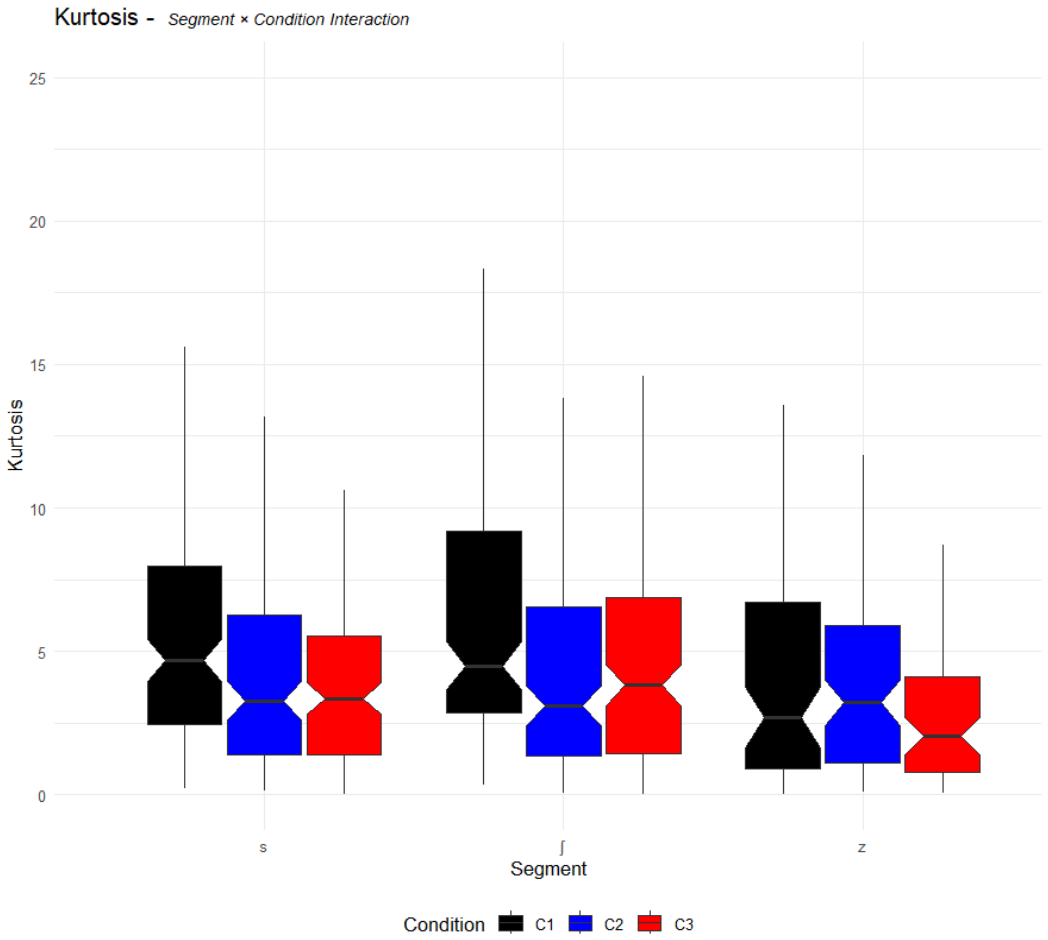


Figure 16. EMM of kurtosis for target sibilants.

Kurtosis was highly variable for all segments, each of which displaying different trends across conditions. ANOVA indicates that condition has a significant effect on kurtosis ($F(2,1327) = 13.5403$, $p<0.0001$) and found an interaction effect between segment and condition ($F(4,1327) =$

4.8134, $p<0.001$). Pairwise comparisons show segment /s/ had a significantly lower kurtosis in C2 ($p<0.05$) and C3 ($p<0.05$) compared to C1. Segment /ʃ/ had a significant decrease from C1 to C2 ($p<0.005$) as well as from C1 to C3 ($p<0.05$) with C2 and C3 not significantly different from each other. Segment /z/ showed a significant decrease from C1 to C3 ($p<0.001$) and from C2 to C3 ($p<0.0001$) with C1 and C2 not significantly different from each other. Based on these comparisons, it may be concluded there is a weakly negative trend across conditions for all segments, though with high variability.

4.4 Participant Spectra

LPC was used to generate spectra for each participants' average pronunciations in each condition, again derived from the middle 50% of each sibilant (as in the noise generation process). Figures 17-26 show spectra of average sibilant pronunciations from each speaker in all three conditions, as well as spectra of generated noise used in C3 for each participant. These visualizations help show the different frequency distributions in each condition, with C1 and C2 sibilants often showing similar acoustic envelopes. That they are often very similar in shape is consistent with Garnier and Henrich's boosting strategy in which global increases to vocal intensity lead to global increases to spectral measures and not any major redistributions of acoustic energy. Theoretically, this follows from the nature of white noise as containing all frequencies in the audible range (20 Hz - 20 kHz) whereby global increases are a simple and reliable means of overcoming acoustic masking; bypass strategies would theoretically be inadequate as there is no portion of the acoustic spectrum which is not crowded. Most important in the spectra are the difference in shape between noise spectra and C3 productions, representative of speakers' attempts to overcome and avoid masking based on the speaker-specific noise. In this condition, notable redistributions of frequency can be seen

including the significant raising of peak frequency. This is represented in the spectra overlays by a right-shifted peak frequency compared to spectra of LPC-generated noise and C1/C2 productions. Spectral overlays for each participant are presented in Figures 17-27 where black, blue and red lines represent productions in C1, C2, and C3, respectively. Speaker-specific noise is represented by a dashed green line which often appears with an extremely similar shape to C2 spectra. As noise was generated by convolving C2 productions with white noise designed to be louder than participants' speech, the resulting noise spectra are often higher in amplitude than sibilant spectra. Peak frequency for speaker-specific noise and for C3 productions are marked by a dotted vertical line. Frequency values for each peak are given under each plot. Not all participants showed changes to peak frequency or major redistributions of acoustic energy, however, with some speakers showing nearly identical spectra for sibilant productions across conditions; Participant 4 is a good example of such consistency, shown in Fig. 20.

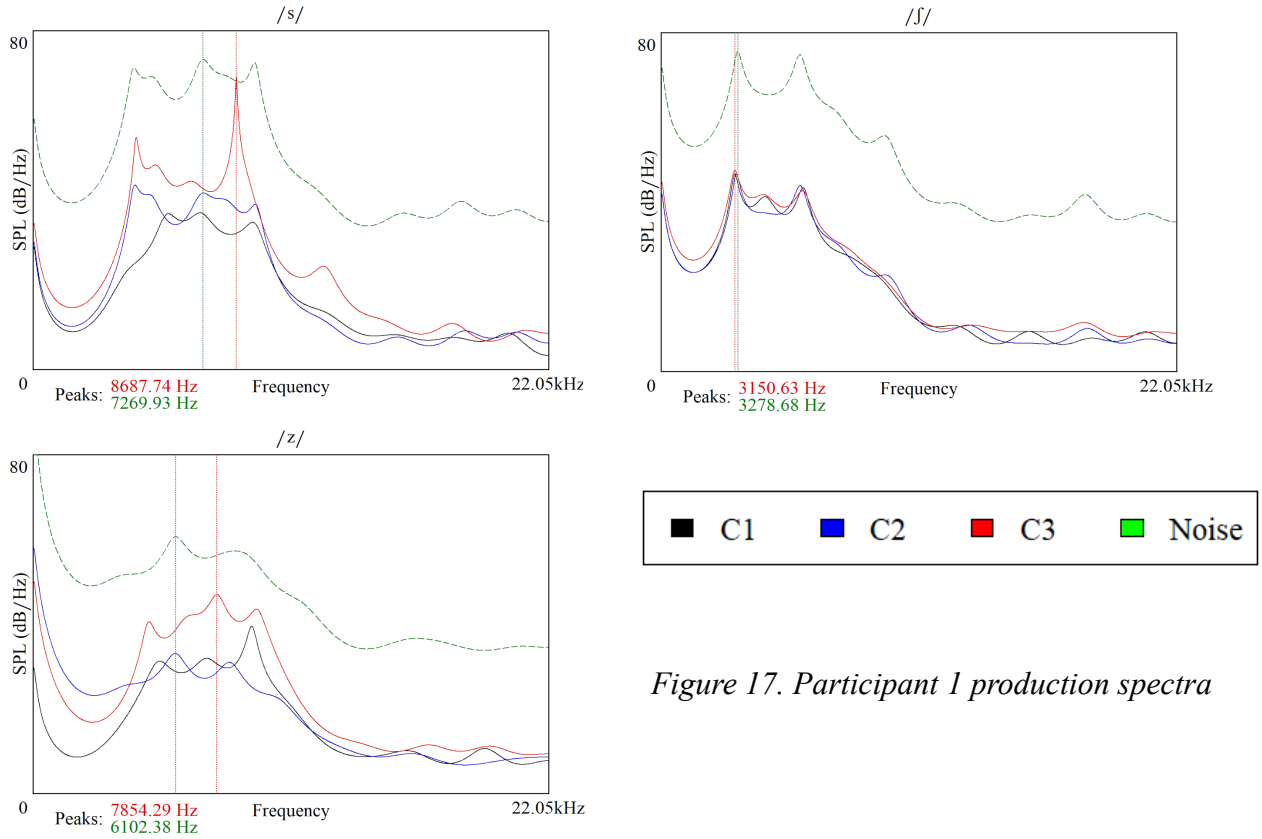


Figure 17. Participant 1 production spectra

Participant 1 shows very different spectral shapes for segments /s z/ compared to noise spectra, including a raising of peak frequency. It is interesting that the C3 peak frequencies roughly sit between two local maximums in the nearby range. For example, noise spectra for segment /s/ shows two high peaks on either side of the C3 peak— the noise peak at 7269 Hz and secondary peak near 9000Hz. Segment /ʃ/ does not show any significant shifting of peak frequency or redistribution of energy across all conditions, suggesting this speaker's postalvolar sibilant is not particularly mutable under different noise conditions.

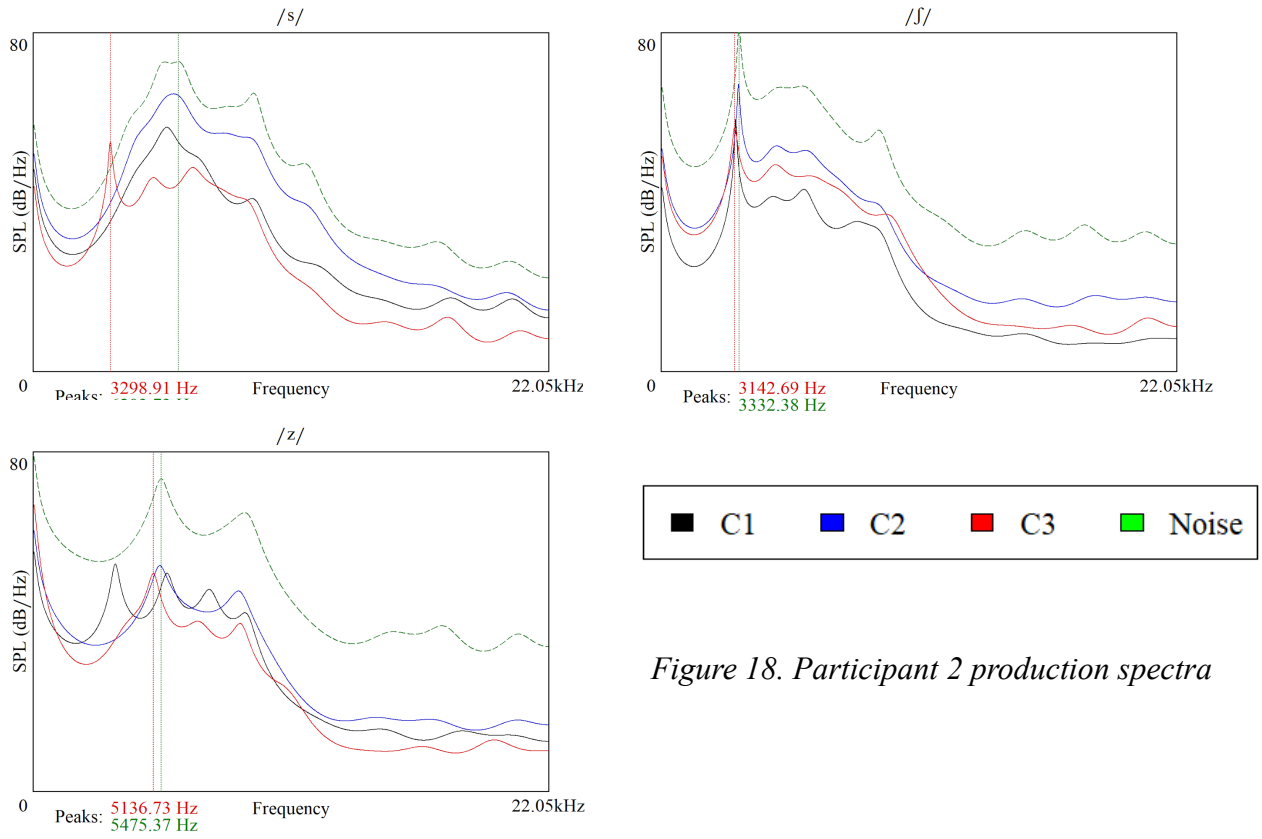


Figure 18. Participant 2 production spectra

This speaker shows clear energy redistributions for both alveolar sibilants, where C3 peak frequency is lowered compared to noise spectra. For segment /s/, productions in C1, C2, and noise spectra all show extremely similar spectral shapes while C3 has a notably lower peak

frequency, suggesting this speaker did employ intensity boosting strategies in C2 but employed frequency bypass strategies in C3. Segment /z/ shows a right shifted C2 spectra compared to C1, though spectral shape is overall very similar. This corresponds to a frequency shift but not global increases to intensity. Segment /ʃ/ shows nearly identical envelopes for all conditions which closely match the noise spectra.

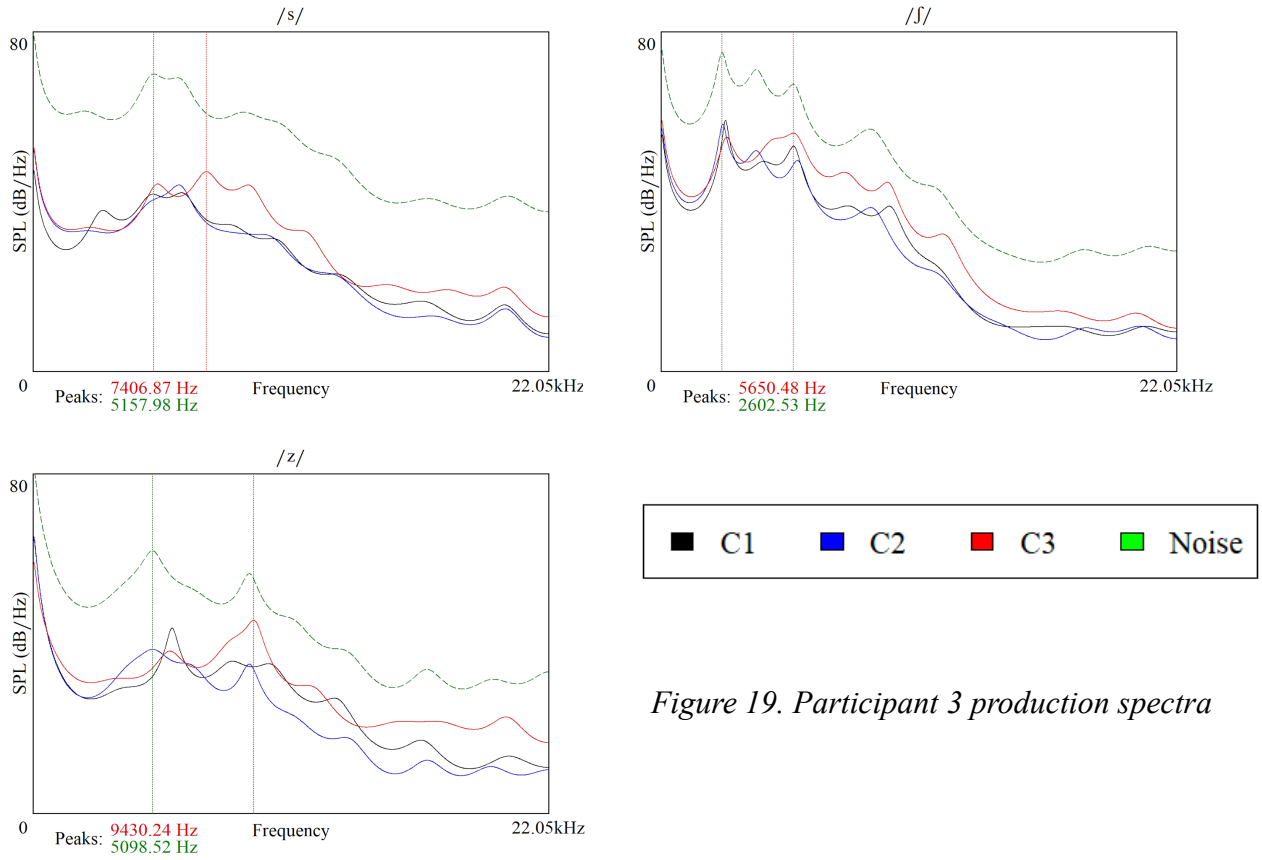


Figure 19. Participant 3 production spectra

Participant 3 displays an increase to peak frequency for all segments in C3 compared to C2, as well as increased amplitude for frequencies above the peak. This was an interesting consistency to find as Participant 3 has a notable lisp, reflected both by listener perceptions of speech as well as spectral measurements; this subject consistently produced /s/ with higher SD and lower kurtosis compared to other speakers, representing a greater diffusion of acoustic energy about the mean and less peaked spectra in C1 conditions. This is consistent with other findings on the

spectral characteristics of sibilants produced by speakers with a lisp, where spectral moments have been used in clinical settings to non-invasively assess such speech impediments (Lowell, 2011). The consistent raising of intensities and peaks in the high frequency range suggests a raising of CoG for all segments in C3.

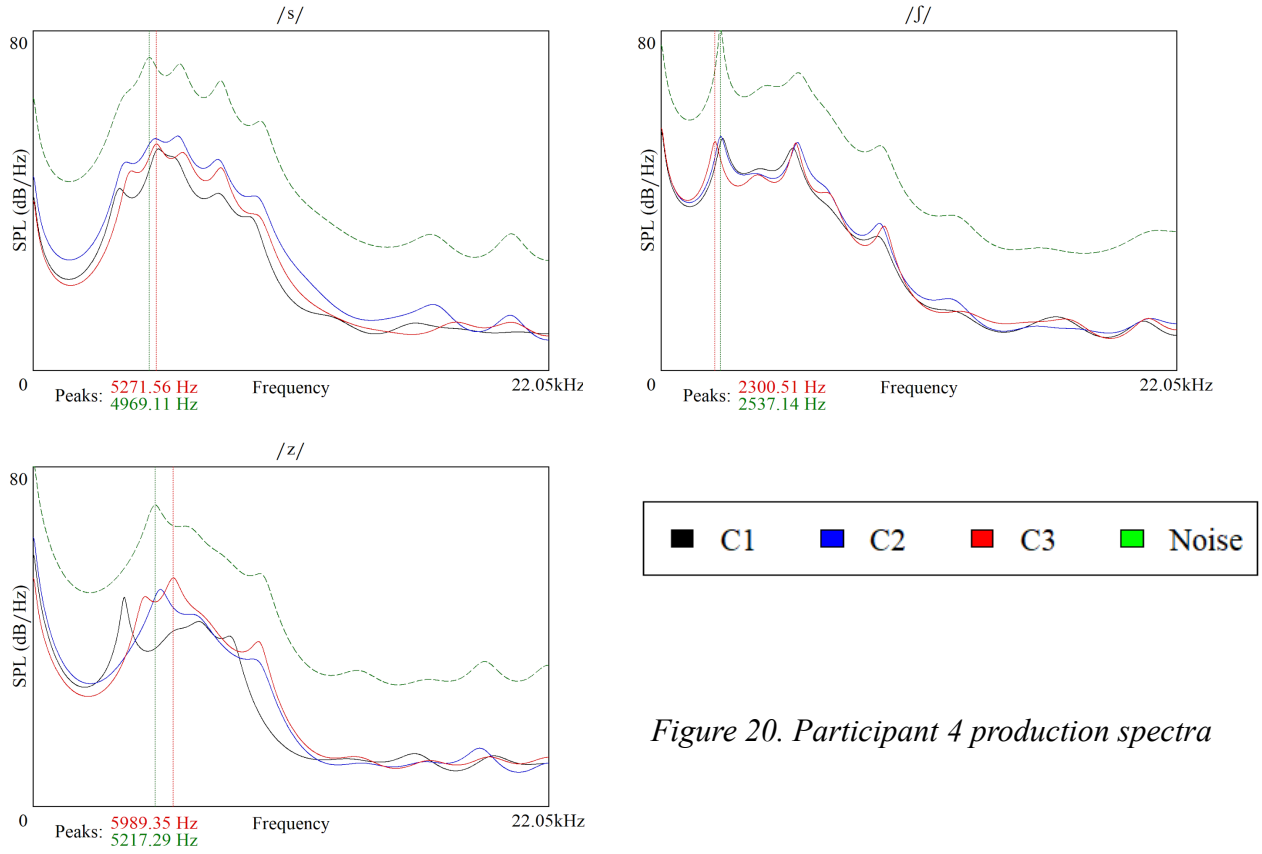


Figure 20. Participant 4 production spectra

This speaker shows the most consistent productions across conditions, especially with respect to voiceless sibilants— spectral shapes are nearly identical across all three experimental conditions and closely resemble noise spectra; peak frequencies have remained almost entirely unchanged, with the greatest increase for segment /z/.The voiced alveolar does show some variability between quiet and noise conditions, however, where C2 and C3 productions show a higher peak frequency and greater amplitudes in the typical sibilant range of 2-9 kHz compared to C1

frequencies. C3 frequencies and intensities are slightly increased compared to C2 but overall maintain a very similar spectral shape.

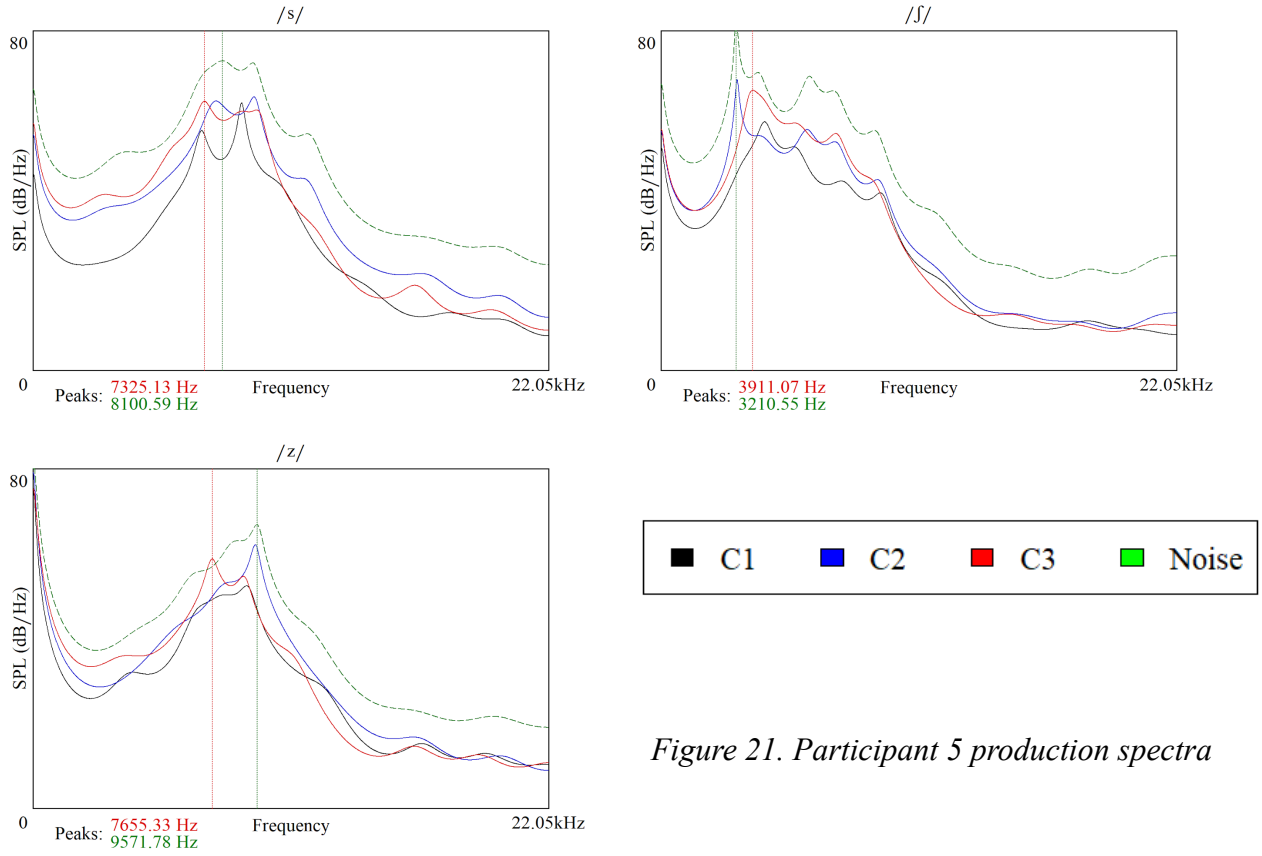


Figure 21. Participant 5 production spectra

This speaker shows interesting redistributions of energy across conditions for all segments, where only the postalveolar /ʃ/ showed a raised C3 peak frequency compared to noise. In the typical sibilant frequency range, C3 frequencies above the peak also show greater intensity compared to C2 productions. Both alveolar sibilants show a decreased peak frequency in C3 compared to other conditions and noise. For segment /s/, the C2 and C3 spectrum are also notably less peaked than C1, corresponding to slightly more diffuse energy about the mean. Segment /z/ shows the inverse, where C2 and C3 spectra are more peaked than C1.

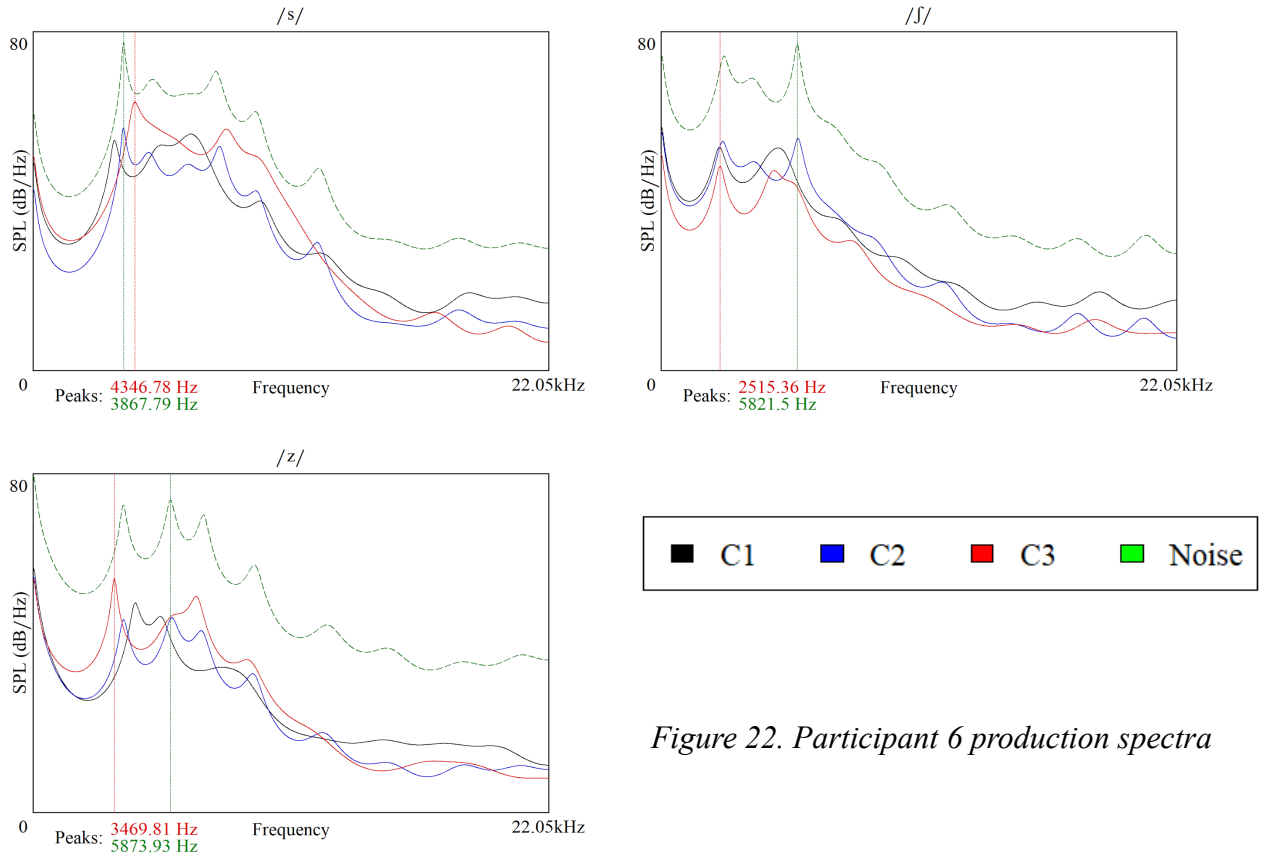


Figure 22. Participant 6 production spectra

Intensity and frequency both vary across noise conditions for all segments, with alveolars showing the greatest degree of change. Segment /ʃ/ shows similar frequency distributions in all conditions but shows a notably decreased intensity in C3 compared to C1 and C2. For all segments, all four spectra show at least two distinct peaks in the typical sibilant range of 2-9 kHz. For segments /ʃ/ and /z/, however, there is alternation between which of the two peaks corresponds to the peak frequency in the C2/noise and C3 spectra. For example, C2 and noise spectra for /ʃ/ show a peak frequency near 5800 Hz and a secondary peak near 2500 Hz. In C3, 2500 Hz is now the peak frequency and 5800 Hz has a lesser amplitude. Similar alternation occurs for /z/, where a secondary peak in C2/noise becomes the peak frequency in C3.

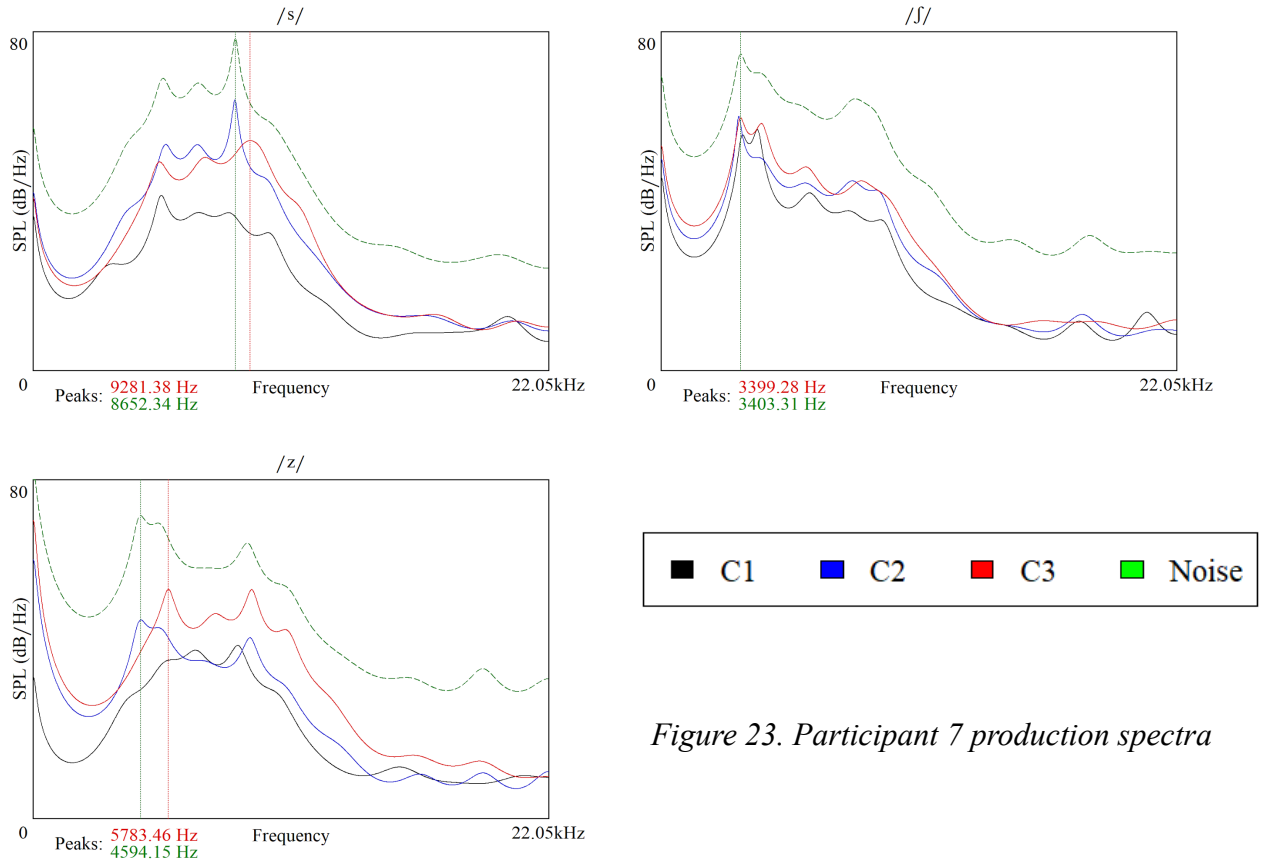


Figure 23. Participant 7 production spectra

Speaker 7 shows some redistribution of energy and shifting of peak frequency, especially with respect to alveolar segments. Both /s/ and /z/ show increased peak frequencies compared to noise, with /z/ showing higher intensity in C3 for all frequencies above the peak. This spectrum is also notably more peaked compared to the C1 spectrum, while /s/ shows a less peaked distribution in C3 compared to C1 and C2. Segment /ʃ/ shows very similar spectra in all conditions and an insignificant raising of peak frequency in C3 compared to noise. Again, the postalveolar seems less mutable under noise compared to alveolars but does show higher amplitudes compared to C2 and C1 productions. This corresponds to global intensity increases without major alteration to frequency distribution, with C3 noise eliciting greater increases to amplitude compared to the white noise condition.

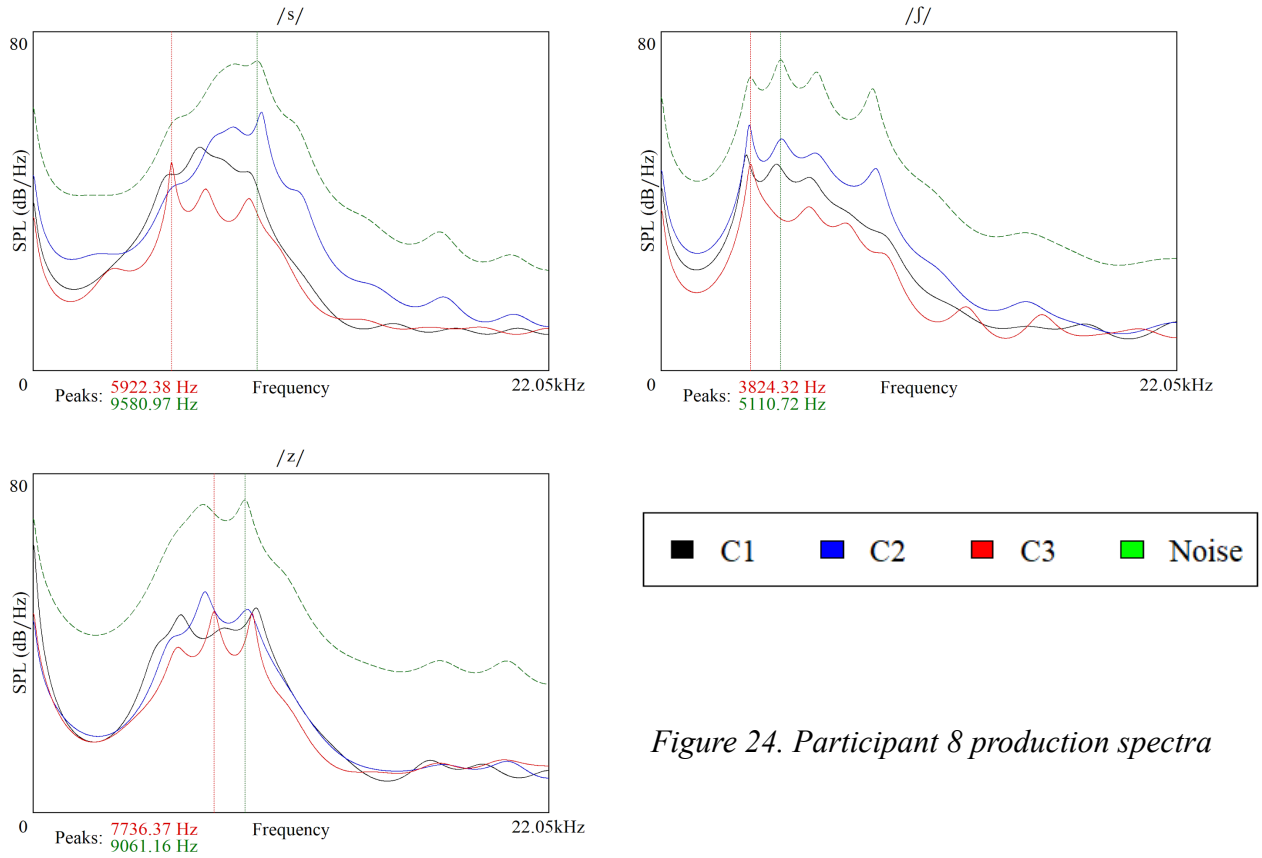


Figure 24. Participant 8 production spectra

Segment /z/ shows the least degree of variability across conditions while the two voiceless sibilants show notable changes to frequency distribution and to intensity. The voiced sibilant shows a decrease in peak frequency in C3 compared to noise though it is still higher than in C2. Spectral shape is not drastically different, however, with two distinct peaks in the sibilant frequency range for C2, noise, and C3 spectra. Amplitudes are slightly reduced in this range compared to C1 and C2. Segment /s/ shows the most variability, with a decreased C3 peak frequency compared to other spectra and showing decreased amplitudes across most of the frequency range. There are also three distinct peaks in the C3 spectrum where other spectra show one. Segment /ʃ/ shows very similar peak frequencies in all three conditions which is incongruent with the generated noise, whose peak frequency corresponds to the secondary peak of the production spectra. Amplitudes for C3 spectra of all segments are consistently lower than that of

C2 and C1; this may be the result of some procedural error given the nature of the LE as in literature as well as elsewhere in this study as generally raising amplitudes.

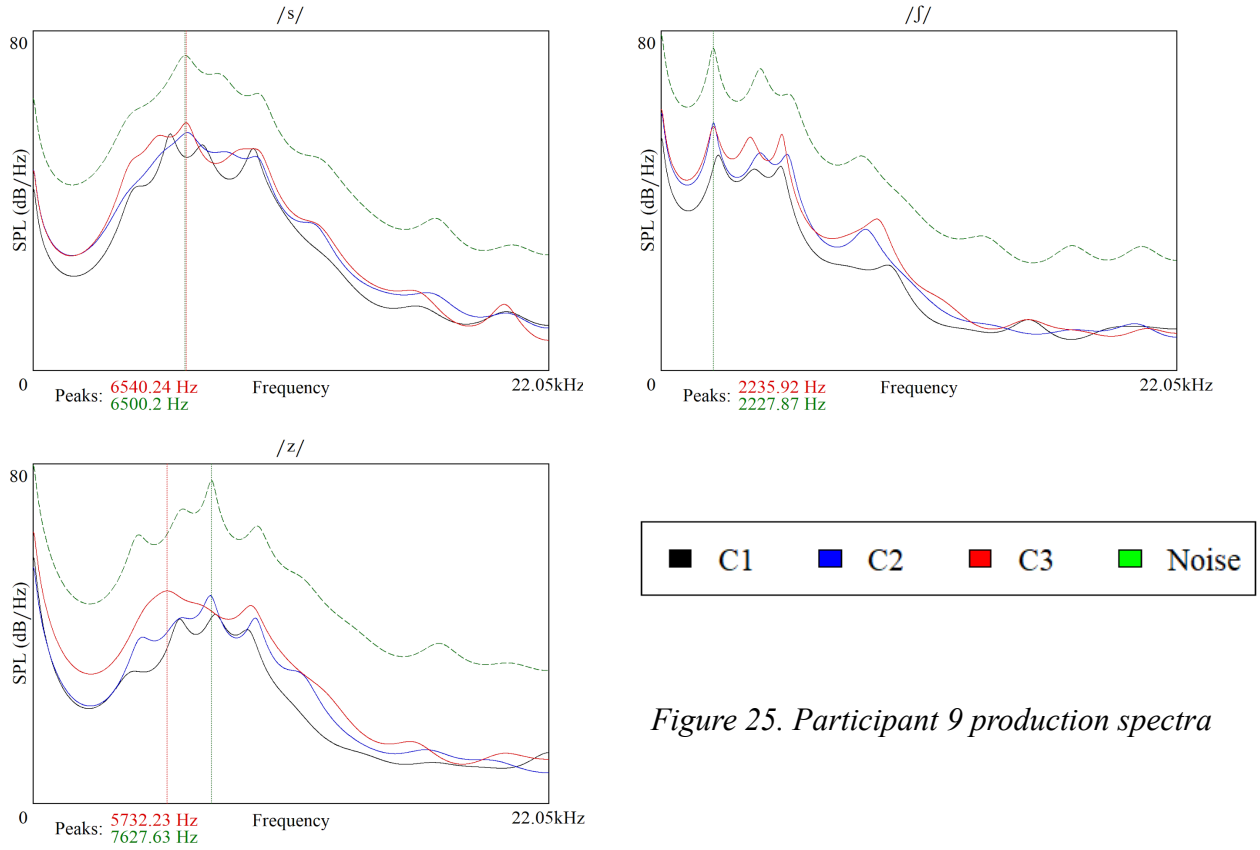


Figure 25. Participant 9 production spectra

This speaker had rather consistent energy distributions across experimental conditions, with voiceless segments especially showing minor changes to frequency and/or amplitude. Segment /s/ had a peak frequency near 6500 Hz in C1 and C3 which in turn matches the peak frequency of speaker-specific noise. Frequencies above this peak have extremely similar amplitudes as in C2 as well as similar frequency peaks. Segment /ʃ/ is similar in that peak frequency did not change across all conditions and also matches the peak of generated noise. Amplitudes in C3 are slightly increased compared to C1 and C2, especially where the spectrum has sharper peaks in the 2-7 kHz range. Segment /z/ shows the greatest variation across experimental conditions, with a lower peak frequency compared to noise. This is the only segment to show noteworthy changes to peak

frequency. Amplitudes across the entire frequency range are greater than C1 and C2 spectra albeit with minor exceptions in the very high frequency range of 15-22.05 kHz.

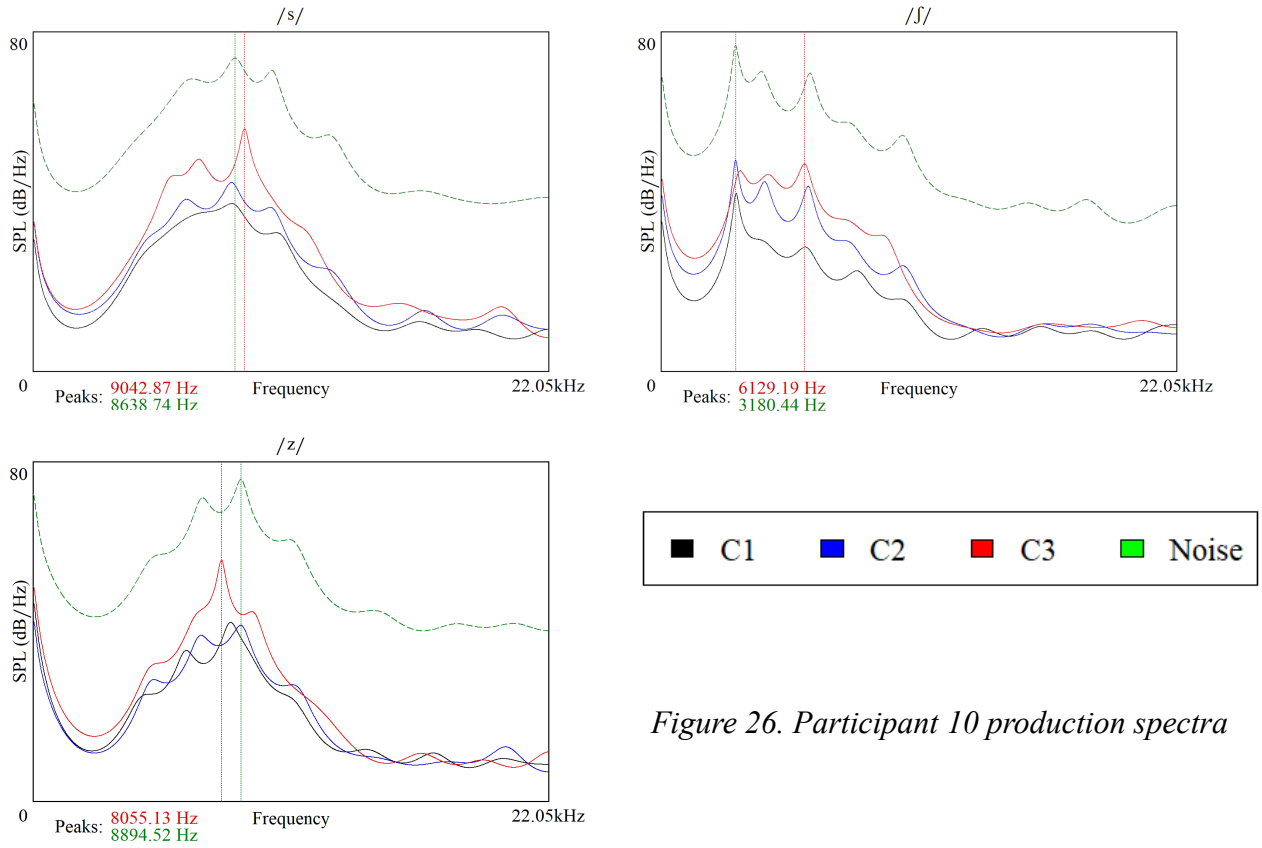


Figure 26. Participant 10 production spectra

All segments show changes to peak frequency and intensity across conditions. Segment /s/ shows a distinctly sharp peak frequency which is slightly higher than that of the noise spectrum. The peak is also notably greater in amplitude compared to C2 and C1. The C3 spectrum also shows greater amplitudes across the entire spectrum, though with some minor exceptions in the upper half of the frequency range (10 kHz+). The postalveolar segment shows an alternation between peak frequencies in C3 compared to noise; primary and secondary peak noise frequencies near 3200 Hz and 6200 Hz, respectively, have alternated in C3 such that they are now secondary and primary, respectively. Amplitudes across the lower half of the frequency range are globally increased compared to C2, which in turn are higher than C1. The voiced alveolar /z/ shows

similar changes to the voiceless counterpart /s/, where the peak C3 frequency is slightly shifted but much sharper and greater in amplitude compared to noise. Here, the peak is slightly lowered but sits equidistant between two local maximums in C2/noise. Amplitude is greater across the lower half of the frequency range compared to other conditions.

5 Discussion

5.1 Spectral changes

Data indicates statistically significant changes to peak frequency, skewness, and kurtosis under speaker-specific noise compared to broadband conditions, though the limited sample size should incite skepticism regarding the statistical insignificance of overall trends for other measures. For example, CoG raising did not meet the statistical significance threshold according to ANOVA and pairwise comparisons, though a clear upward trend can be seen in Fig. 12. Given the logical connection between peak frequency, which showed a statistically significant upward trend across conditions, and the CoG measure, an average of frequency distributions as a function of intensity, it is suspected that a greater sample size would reveal a statistically significant upward trend in CoG which corresponds to the upward trend in peak frequency. Articulatorily, a raising of peak frequency and CoG is consistent with a more anterior tongue position which forms a smaller prelingual cavity during sibilant productions compared to productions in quiet and broadband conditions. The raising of peak frequency while CoG remains stagnant logically indicates a redistribution of acoustic energy toward lower frequencies in the acoustic spectrum, though this is difficult to discern with current methods of analysis. Given the nature of the LE as generally raising prominent frequencies (such as increases to F1 and F2), it is more likely that energy is not redistributed lower in the spectrum but that the lack

of significance of the CoG trend is a result of limited data. This is reinforced by findings of decreased skewness in C3 compared to C1— while segments are generally positively skewed in quiet conditions, a near-zero skewness in C3 indicates an upward redistribution of energy leading to more symmetrical spectral shapes. This should also contribute to a higher CoG consistent with boosting strategies and contrast with the downward redistribution of acoustic energy implied by a stagnant CoG. Examination of segment spectra also confirms this notion of upward redistribution, whereby C3 productions often show a right-shifted highest peak and higher SPL for frequencies above the peak. In conjunction, it can thus be indirectly concluded that frequencies are indeed slightly upwardly redistributed consistent with boosting strategies, though the degree to which CoG actually increases across conditions cannot be adequately discerned from these data.

Because of the limited data size, statistical power (or lack thereof) is another reason why speaker-specific envelopes are important to analyze. Though statistical analysis indicates a weakly positive correlation for some of the six acoustic measures, visual analysis reveals the ways in which different speakers employ different strategies to overcome ambient noise in C3. Some participants show global increases to frequency and intensity while others might be seen as employing bypass strategies. Such distinctions are not necessarily clear or mutually exclusive, however, as some speakers' changes to frequency and intensity occur quasi-independently. For example, C3 intensities in the lower half of the frequency range may be increased compared to noise/C2 though frequency distribution remains largely unchanged. Simultaneously, intensities at frequency peaks and troughs in the upper half of the frequency range may appear unchanged while energy is distributed among higher frequencies overall (ie. a rightward-shifted C3 spectrum with similar peaks and troughs). Further, other speakers seem to not majorly alter

energy distributions across any noise conditions and show very consistent spectra for all segments, as in Fig. 20. The high degree of variability in speakers' strategies thereby weakens the overall strength of trends which might be better understood with a more fine-grained categorization of speakers. Here, Garnier and Henrich's multilayered understanding and distinctions between boosting and bypass might again become useful— some speakers such as participants 7 and 10 show clear global increases to amplitude and frequency across conditions as indicated by the spectral overlays in Fig. 23 and 26. Others, such as speakers 5 and 6, show C3 peaks in Fig. 21 and 22 where other conditions show troughs, indicating very particular redistributions of energy in order to bypass those noise frequencies where amplitude is greatest. Others still show changes which cannot be easily categorized, such as participant 9 (Fig. 25), while those such as participant 4 show little to no changes across conditions, as in Fig. 20. Results are therefore consistent with Garnier and Henrich's (2014) findings that global increases to vocal effort translate to increased amplitudes and higher frequencies while subtle, secondary strategies are simultaneously engaged, especially in the lower half of frequency range.

Duration, as a supplementary measure in this experiment, is helpful in verifying the temporal effects of noise on sibilant production but, as mentioned, is not a highly contested characteristic of the LE. In conjunction with other significant trends, especially decreased skewness, the increases to duration across noise conditions may be understood to indicate increasingly prolonged and stable sibilant articulation, whereby more typical and gradual onsets of turbulence generally lead to asymmetric distributions as discussed in §1.4. That skewness is reduced, in conjunction with duration increases, may also indicate a more instantaneous segment onset. Articulatorily, this follows from a stronger initial constriction which ensures earlier production of high frequency energy resulting from a narrower turbulent air channel. Compared

to the positively skewed (on average) spectra produced in quiet conditions, a weaker initial constriction which gradually strengthens across the sibilant duration (to some peak before weakening into segment offset) would explain why spectra are skewed toward lower frequencies in the 0-22.05 kHz frequency window. A more instantaneous onset might help explain, along with the upward frequency distributions, why lower frequencies are less prominent by the time the midpoint of segment production is reached. Articulatory stability is taken as the primary novel finding of this experiment in conjunction with confirmation of Garnier and Henrich's multilayered strategy-based approach, where LS is adjusted in both frequency and temporal domains compared to speech in quiet. This experiment also confirmed the high degree of speaker variability in adopting these strategies and showed they are indeed implemented at a segment-specific level, at least with regard to sibilant consonants.

5.2 Limitations

LPC is one particular method of synthesizing speech that has been used in audio signal and speech processing for several decades (Atal & Remde, 1982; Valin & Skoglun, 2019). One of its major advantages is that it can synthesize based on actual audio input (ie. recordings of participants) and provides an adequately detailed approximation of the spectral envelope with distinct peaks. For sibilants, and fricatives more broadly, it may fall short in capturing variation across a segment's duration (Hunt et al. 1989; Little et al. 2006). This is why the middle 50% of each target segment was used for generation as opposed to entire durations, as this provides a more stable spectral shape for analysis and replication. This study is also not particularly concerned with determining coarticulation effects of the vocalic nucleus but is focused on speaker-specific sibilants as a class, not particular phonetic/phonological environments. Other synthesis methods have their own advantages and disadvantages. For example, Butterworth

filters as used by Reilly (2020) are good for synthesizing sound which matched the general spectral shape of sibilant consonants but they are limited in their ability to replicate sharp and distinct peaks in the acoustic envelope and utilize maximally-smooth passbands which can result in temporal smearing and affecting intelligibility (Drullman et al., 1994; Elliott & Theunissen, 2009). They are thus suitable for synthesizing sibilant sounds as a general class but are not as precise in their ability to replicate the minute differences which differentiate various sibilants produced by an individual speaker, especially considering coarticulation effects with the following vowel.

Alternatively, noise could theoretically be synthesized to exactly match the spectral moments of productions in C2, though the issue of tailoring M3 and M4 to exact values while also matching M1 and M2 is an ongoing issue in the field of mathematics and audio engineering and is beyond the resources and scope of this investigation to solve or employ (see, for example, Nakamura et al., 2014; Takamichi et al., 2017; Henter et al., 2018).

The LPC method used in this experiment has potential limitations regarding the voiced sibilant /z/, as the generated envelope is influenced by the F0 voicing parameter. This parameter is inconsistent both within and across speakers; some voiced alveolar pronunciations employ clear prevoicing before producing any sibilance while other consecutive tokens fully alternate between a voiced [z] and voiceless [s]. F0 was thus not consistent across the three repetitions in any one noise condition and further differed between target words. Voicing is further complicated when prosodic effects are considered, such as the use of list intonation (see Morrill et al., 2014; Grice et al., 2020; for example). A more robust method of accounting for F0 and its effect on LPC generation should be employed in future studies involving voiced sibilants such as /z/.

Coordination of sound to items could also be refined. Coarticulation due to the following vocalic environment can influence spectral characteristics of the preceding sibilant (for example, preemptive raising of the jaw to articulate high front vowel /i/ reduces size of the oral cavity which raises frequencies in the preceding sibilant). Though onsets and offsets of sibilants were not included in the analysis, the stable middle 50% may still be influenced by these coarticulation effects. Acoustic filters could be generated more specifically so as to match the SV pair and not the S pronunciations across vowel conditions.

Volume normalization could also be refined. Due to the vastly different frequency content and distribution for each speaker's sibilants, normalization methods such as equalizing LUF and RMS energy did not create volumes that were perceptually equal. Manual adjustment and quantitative normalization should be used in conjunction in future studies.

Lastly, the participant pool should be expanded and refined to better control for multilingual confounds and age.

5.3 Directions for Future Work

Continuations of this work should verify the restriction of significant noise condition effects to measures of peak frequency, skewness, and kurtosis; it may be that trends in CoG and SD simply require a larger sample size to be recognized as statistically significant. The primary conclusion that decreased skewness corresponds to increasing stable articulations, while true in theory, could be vetted with use of ultrasonic imagery or other methods which are sensitive to changes in oral articulation across segment durations. First and foremost, however, the noise generation process should be refined and better controlled to account for nucleic environment and co-articulation effects. Additionally, communicative burden should be made more salient by introducing an

interlocutor to the immediate speaking environment and by performing cooperative communication tasks instead of reading from word lists.

6 Conclusions

Weak/moderate effects restricted to secondary spectral measures (M3/M4) suggest that when exposed to speaker-specific noise, overall spectral shape does not largely change but energy is more balanced and symmetric. In conjunction with greater duration, articulations might be becoming more stable for a longer period such that sibilant productions are more careful and persistent. This ensures that the speech signal is better established amongst the noise signal without drastically changing overall acoustic characteristics. Though intensity was not measured, global increases to vocal effort would raise CoG more than was found here despite the observed raising of peak frequency. Combined with secondary acoustic augmentation, this is taken to suggest that global boosting strategies are indeed implemented but only to the extent where overall spectral shape and key acoustic characteristics can be preserved, thus ensuring intelligibility and a positive signal-to-noise ratio are maintained. To return to the primary research question of this experiment— these changes do indeed differentiate productions under speaker-specific noise from those in the broadband condition, suggesting that the LE is multifaceted and, at the very least, displays a weak to moderate spectral sensitivity in some speakers. The manner in which this sensitivity manifests is highly speaker-dependent.

References

American Speech-Hearing-Language Association . (n.d.). *ENGLISH PHONEMIC INVENTORY 1*. https://www.asha.org/siteassets/uploadedfiles/englishphonemicinventory.pdf

Atal, B., & Remde, J. (1982, May). A new model of LPC excitation for producing natural-sounding speech at low bit rates. In *ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 7, pp. 614-617). IEEE.

Audacity Team. (2023). *Audacity* (Version 3.4.2) [Computer software]. https://www.audacityteam.org/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(2), 217–270. https://doi.org/10.1017/s0952675708001474

Chomsky, N., & Halle, M. (1968). *THE SOUND PATTERN OF ENGLISH* . https://web.mit.edu/morrishalle/pubworks/papers/1968_Chomsky_Halle_The_Sound_Pattern_of_English.pdf

Dreher, J. J., & O'Neill, J. (1957). Effects of Ambient Noise on Speaker Intelligibility for Words and Phrases. *The Journal of the Acoustical Society of America*, *29*(12), 1320–1323. https://doi.org/10.1121/1.1908780

Elliott, T. M., & Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Computational Biology*, *5*(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302

Fox, R. A., & Nissen, S. L. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives. *Journal of Speech, Language, and Hearing Research*, *48*(4), 753–765. https://doi.org/10.1044/1092-4388(2005/052)

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, *28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Henrich, N., & Dubois, D. (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)

Grice, M., German, J. S., & Warren, P. (2020). Intonation Systems Across Varieties of English. *The Oxford Handbook of Language Prosody*, 284–302. https://doi.org/10.1093/oxfordhb/9780198832232.013.18

Haley, K. L., Seelinger, E., Callahan Mandulak, K., & Zajac, D. J. (2010). Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics*, *38*(4), 548–554. https://doi.org/10.1016/j.wocn.2010.07.006

Hayden, Rebecca E. (1950). The Relative Frequency of Phonemes in General-American English. *Word*, 6:3, 217-223. https://doi.org/10.1080/00437956.1950.11659381

Henter, Gustav Eje, King, S., Merritt, T., & Degottex, G. (2018). Analysing Shortcomings of Statistical Parametric Speech Synthesis. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1807.10941

Hermans, R. (2008). *Negative and positive skew diagrams (English)* [Digital image]. Wikimedia Commons.

https://commons.wikimedia.org/wiki/File:Negative_and_positive_skew_diagrams_(English).svg

Hotchkin, C., & Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. *Biological Reviews*, *88*(4), 809–824. https://doi.org/10.1111/brv.12026

International Organization for Standardization. (2023). *ISO 226:2023 – Acoustics — Normal equal-loudness-level contours*. https://www.iso.org/standard/83117.html

Jean-Sylvain Liénard, & Maria-Gabriella Di Benedetto. (1999). Effect of vocal effort on spectral properties of vowels. *Journal of the Acoustical Society of America*, *106*(1), 411–422. https://doi.org/10.1121/1.428140

Junqua, J. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, *93*(1), 510–524. https://doi.org/10.1121/1.405631

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20*(1-2), 13–22. https://doi.org/10.1016/s0167-6393(96)00041-6

Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, *53*(3), 983. https://doi.org/10.2307/2533558

Kim, J., & Arnhold, A. (2024). Phonetic and phonological aspects of prosodic focus marking in Canadian English. *Laboratory Phonology*, *15*(1). https://doi.org/10.16995/labphon.9316

Kokkelmans, J. (2021). *The Phonetics and Phonology of Sibilants: A Synchronic and Diachronic OT Typology of Sibilant Inventories* . https://roa.rutgers.edu/content/article/files/1864_kokkelmans_1.pdf

Kunc, H. P., Morrison, K., & Schmidt, R. (2022). A meta-analysis on the evolution of the Lombard effect reveals that amplitude adjustments are a widespread vertebrate mechanism. *Proceedings of the National Academy of Sciences*, *119*(30). https://doi.org/10.1073/pnas.2117809119

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709. https://doi.org/10.1044/jshr.1404.677

Lin, F. R. (2024). Age-Related Hearing Loss. *New England Journal of Medicine˜ the œNew England Journal of Medicine*, *390*(16), 1505–1512. https://doi.org/10.1056/nejmcp2306778

Little, M. A., McSharry, P. E., Moroz, I. M., & Roberts, S. J. (2006). Testing the assumptions of linear prediction analysis in normal vowels. *The Journal of the Acoustical Society of America*, *119*(1), 549–558. https://doi.org/10.1121/1.2141266

Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales Des Maladies de l'Oreille et Du Larynx*, *37*(2), 101–119.

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2010). Spectral- and Cepstral-Based Measures During Continuous Speech: Capacity to Distinguish Dysphonia and Consistency Within a Speaker. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2010.06.007

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, *124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Lu, Y., & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America*, *126*(3), 1495–1499. https://doi.org/10.1121/1.3179668

Maniwa, K., Jongman, A., & Wade, T. (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *123*(2), 1114–1125. https://doi.org/10.1121/1.2821966

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973. https://doi.org/10.1121/1.2990715

Mokbel, C. (1992). *Reconnaissance de la parole dans le bruit: bruitage/débruitage*. http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=155578

Morrill, T. H., Dilley, L. C., & McAuley, J. D. (2014). Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure. *Journal of Phonetics*, *46*, 68–85. https://doi.org/10.1016/j.wocn.2014.06.001

Nakamura, K., Hashimoto , K., Nankaku, Y., & Tokuda, K. (2014). Integration of Spectral Feature Extraction and Modeling for HMM-Based Speech Synthesis. *IEICE Transactions on Information and Systems*, *E97.D*(6), 1438–1448. https://doi.org/10.1587/transinf.e97.d.1438

Patel, R., & Schell, K. W. (2008). The Influence of Linguistic Content on the Lombard Effect. *Journal of Speech, Language, and Hearing Research*, *51*(1), 209–220. https://doi.org/10.1044/1092-4388(2008/016)

Reilly, K. J. (2020). Vowel and Sibilant Production in Noise: Effects of Noise Frequency and Phonological Similarity. *Journal of Speech Language and Hearing Research*, *63*(4), 1002–1017. https://doi.org/10.1044/2020_jslhr-19-00345

Rostolland, D. (1982). Phonetic structure of shouted voice. *Acta Acustica united with Acustica*, *51*(2), 80-89.

Schulman, R. G. (1989). Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America*, *85*(1), 295–312. https://doi.org/10.1121/1.397737

Shadle, C. (1985). The acoustics of fricative consonants. *Mit.edu*. http://hdl.handle.net/1721.1/15252

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics*, *19*(3-4), 409–424. https://doi.org/10.1016/s0095-4470(19)30332-8

Shinnosuke Takamichi, Tomoki Koriyama, & Saruwatari, H. (2017). Sampling-based speech parameter generation using moment-matching networks. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1704.03626

Stevens, K. N. (1971). Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations. *The Journal of the Acoustical Society of America*, *50*(4B), 1180–1192. https://doi.org/10.1121/1.1912751

Stevens, K. N. (2000). *Acoustic phonetics*. MIT Press.

Stowe, L. M., & Golob, E. J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *The Journal of the Acoustical Society of America*, *134*(1), 640–647. https://doi.org/10.1121/1.4807645

Stuart-Smith, J. (2020). Changing perspectives on /s/ and gender over time in Glasgow. *Gla.ac.uk*. https://eprints.gla.ac.uk/188216/8/188216.pdf

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, *107*(6), 3438–3451. https://doi.org/10.1121/1.429414

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, *5*(2), 99. https://doi.org/10.2307/3001913

Valin, J. M., & Skoglund, J. (2019, May). LPCNet: Improving neural speech synthesis through linear prediction. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5891-5895). IEEE.

Appendix A.1 Participant Distribution

Participant	Sex	Age	Additional languages spoken; Fluency (1-5)
1	F	25	N/A
2	M	22	Somali; 2
3	M	22	Arabic; 5
4	M	22	Spanish; 4
5	F	18	N/A
6	M	22	Classical Arabic; 3
7	F	20	Farsi; 5 French; 1
8	F	20	French; 5
9	F	24	Spanish; 5 French; 3.5 German; 3
10	M	29	French; 4 Italian; 3 Irish; 2

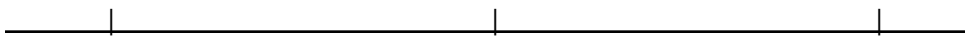
Appendix A.2. Average Participant Age

	Average Age		
Sex	Mean	Median	Mode
F	21.4 years	20	20
M	23.4 years	22	22

Appendix B. Background Questionnaire

Participant Background Questionnaire

1. **Name:** _____
2. **Date of Birth (mm/dd/year):** _____/_____/_____
3. **Location of Birth:** _____
4. **Sex:** _____
5. **Is English your first language?** *Yes / No
   *participation in this study is contingent on your status as a native English speaker
6. **Do you speak any other languages? If so, please indicate which languages and your level of proficiency from 1-5:**



1 **3** **5**

Basic **Conversational** **Native-fluency**

(i) **Language:** _____ **Proficiency:** _____

(ii) **Language:** _____ **Proficiency:** _____

(iii) **Language:** _____ **Proficiency:** _____

7. **How did you learn these languages (ie. through family, standard school curriculum, independent study, etc.)? Please indicate for each language other than English.**

(i) **L:** _____ **Acquisition:** _____

(ii) **L:** _____ **Acquisition:** _____

(iii) **L:** _____ **Acquisition:** _____

8. **Do you have any clinically diagnosed hearing impairments/conditions, auditory processing disorders, or use hearing aids? If so, please specify:**

_____

_____

9. **Do you have any familiarity or past experience with linguistics experiments?**

_____

_____

Appendix C.1 Target Words

	Target items			Distractor items				
	/ s /	/ z /	/ ʃ /	/f/	/θ/	/w/	/l/	/r/
/ i /	seal	zeal	shield		theme	wheel	leer	
/ ɪ /	sip	zip	ship	fin	thin		lift	
/ u /	sue	zoo	shoe					room
/ ɜ /	said	zed	shed		them	well	leg	red
/ æ /	sag	zag	shag	fall		wall		
/ʌ /				fun				run

Appendix C.2. Sample Word Lists

Subject	Word List
1	fin, room, shag, sue, seal, zeal, fall, sag, said, them, red, well, wheel, run, thin, zag, shield, zed, leg, sip, fun, shed, shoe, leer, zoo, zip, ship, wall, lift, theme
2	them, well, sag, run, zeal, zag, sue, zed, room, sip, fin, shed, shag, leer, seal, zip, fall, wall, said, theme, red, lift, wheel, ship, thin, zoo, shield, shoe, leg, fun
3	room, sue, fin, zeal, shag, sag, seal, them, fall, well, said, run, red, zag, wheel, zed, thin, sip, shield, shed, leg, leer, fun, zip, shoe, wall, zoo, theme, ship, lift
4	well, run, them, zag, sag, zed, zeal, sip, sue, shed, room, leer, fin, zip, shag, wall, seal, theme, fall, lift, said, ship, red, zoo, wheel, shoe, thin, fun, shield, leg
5	run, zag, well, zed, them, sip, sag, shed, zeal, leer, sue, zip, room, wall, fin, theme, shag, lift, seal, ship, fall, zoo, said, shoe, red, fun, wheel, leg, thin, shield

Appendix E. Code for Spectral Measurements

```
# Configuration form
form Spectral Moments and Peak Frequency Analysis (Using Built-in Functions)
    comment TextGrid tier to analyze (tier number):
    positive tier_number 1
    comment Analyze only non-empty intervals? (1=yes, 0=no):
    boolean non_empty_only 1
    comment Minimum frequency for peak detection (Hz):
    positive minimum_peak_frequency 2000
    comment Save results to file? (1=yes, 0=no):
    boolean save_to_file 0
    comment Output file name (if saving):
    text output_file spectral_moments_results.txt
endform

# Check that correct objects are selected
numberOfSelectedSounds = numberOfSelected("Sound")
numberOfSelectedTextGrids = numberOfSelected("TextGrid")

if numberOfSelectedSounds = 0
    exitScript: "Please select a Sound object."
elsif numberOfSelectedSounds > 1
    exitScript: "Please select only one Sound object."
endif

if numberOfSelectedTextGrids = 0
    exitScript: "Please select a TextGrid object."
elsif numberOfSelectedTextGrids > 1
    exitScript: "Please select only one TextGrid object."
endif

# Get selected objects
soundObject = selected("Sound")
textGridObject = selected("TextGrid")
soundName$ = selected$("Sound")
textGridName$ = selected$("TextGrid")

# Initialize results display
writeInfoLine: "Spectral Moments and Peak Frequency Analysis (Using Built-in Functions)"
appendInfoLine:
"============================================================================="
appendInfoLine: "Sound: ", soundName$
appendInfoLine: "TextGrid: ", textGridName$
appendInfoLine: "Tier: ", tier_number
appendInfoLine: "Spectral moments calculated with Power = 2 across entire intervals"
appendInfoLine: "Peak frequency search range: ", minimum_peak_frequency, " Hz and above"
appendInfoLine: ""

# Initialize output file if requested
if save_to_file
```

```
    writeFile: output_file$, "Sound", tab$, "TextGrid", tab$, "Interval", tab$, "Label", tab$,
"Start_Time", tab$, "End_Time", tab$, "Duration", tab$, "Peak_Frequency", tab$,
"Centre_of_Gravity", tab$, "Standard_Deviation", tab$, "Skewness", tab$, "Kurtosis", newline$
endif

# Get number of intervals in specified tier
select textGridObject
numberOfIntervals = Get number of intervals: tier_number

if numberOfIntervals = 0
    exitScript: "No intervals found in tier ", tier_number
endif

appendInfoLine: "Processing ", numberOfIntervals, " intervals..."
appendInfoLine: ""

# Column headers for Info window display
appendInfoLine: "Int.", tab$, "Label", tab$, "Start", tab$, "End", tab$, "Dur.", tab$, "Peak",
tab$, "CoG", tab$, "SD", tab$, "Skew", tab$, "Kurt"
appendInfoLine: "----", tab$, "-----", tab$, "-----", tab$, "---", tab$, "----", tab$, "----",
tab$, "---", tab$, "--", tab$, "----", tab$, "----"

# Process each interval
validIntervals = 0
for interval from 1 to numberOfIntervals
    select textGridObject

    # Get interval information
    startTime = Get start time of interval: tier_number, interval
    endTime = Get end time of interval: tier_number, interval
    label$ = Get label of interval: tier_number, interval
    duration = endTime - startTime

    # Skip empty intervals if requested
    if non_empty_only and label$ = ""
        # Skip this interval
    else
        # Check if interval has minimum duration for analysis
        if duration >= 0.001
            validIntervals += 1

            # Extract the entire interval for analysis
            select soundObject
            Extract part: startTime, endTime, "rectangular", 1, 0
            tempSound = selected("Sound")

            # Create spectrum
            To Spectrum: "yes"
            spectrum = selected("Spectrum")

            # Calculate spectral moments using built-in functions with Power = 2
            centreOfGravity = Get centre of gravity... 2
            standardDeviation = Get standard deviation... 2
```

```
            skewness = Get skewness... 2
            kurtosis = Get kurtosis... 2

            # Find peak frequency with frequency restriction
            call findPeakFrequency

            # Display results in Info window
            appendInfoLine: interval, tab$, label$, tab$, fixed$(startTime, 2), tab$,
fixed$(endTime, 2), tab$, fixed$(duration, 2), tab$, fixed$(peakFrequency, 0), tab$,
fixed$(centreOfGravity, 0), tab$, fixed$(standardDeviation, 0), tab$, fixed$(skewness, 2),
tab$, fixed$(kurtosis, 2)

            # Save to file if requested
            if save_to_file
                appendFile: output_file$, soundName$, tab$, textGridName$, tab$, interval,
tab$, label$, tab$, fixed$(startTime, 3), tab$, fixed$(endTime, 3), tab$, fixed$(duration, 3),
tab$, fixed$(peakFrequency, 2), tab$, fixed$(centreOfGravity, 2), tab$,
fixed$(standardDeviation, 2), tab$, fixed$(skewness, 4), tab$, fixed$(kurtosis, 4), newline$
            endif

            # Clean up temporary objects
            select tempSound
            Remove
            select spectrum
            Remove
        else
            appendInfoLine: interval, tab$, label$, tab$, fixed$(startTime, 2), tab$,
fixed$(endTime, 2), tab$, fixed$(duration, 2), tab$, "SKIP (too short < 1ms)"
        endif
    endif
endfor

appendInfoLine: ""
appendInfoLine: "Analysis complete!"
appendInfoLine: "Valid intervals analyzed: ", validIntervals, " of ", numberOfIntervals
appendInfoLine: "Note: Spectral moments calculated over full spectrum with Power = 2 across
entire intervals"
appendInfoLine: "Peak frequency searched only above ", minimum_peak_frequency, " Hz"

if save_to_file
    appendInfoLine: "Results saved to: ", output_file$
endif

# Restore original selection
select soundObject
plus textGridObject

# Procedure to find peak frequency with frequency restrictions
procedure findPeakFrequency
    # Get spectrum information
    numberOfBins = Get number of bins
    lowestFrequency = Get lowest frequency
    highestFrequency = Get highest frequency
```

```
    frequencyStep = Get bin width

    # Variables for peak frequency (with frequency restriction)
    maxPower = 0
    peakFrequency = minimum_peak_frequency
    peakFound = 0

    # Find peak frequency in restricted range
    for bin from 1 to numberOfBins
        frequency = lowestFrequency + (bin - 1) * frequencyStep

        # Only consider frequencies above minimum threshold
        if frequency >= minimum_peak_frequency
            realPart = Get real value in bin: bin
            imagPart = Get imaginary value in bin: bin
            power = realPart^2 + imagPart^2

            if power > maxPower
                maxPower = power
                peakFrequency = frequency
                peakFound = 1
            endif
        endif
    endfor

    # If no peak found in range, use minimum frequency
    if peakFound = 0
        peakFrequency = minimum_peak_frequency
    endif
endproc

# Notes:
# - Uses Praat's built-in spectral moment functions with Power = 2
# - Built-in functions calculate over the full spectrum (no frequency restrictions)
# - Results will match manually clicking "Get centre of gravity...", etc. with Power = 2
# - Peak frequency detection uses manual calculation for frequency range restriction
# - Empty intervals and intervals shorter than the analysis window are skipped
```

Appendix F. Code for Statistical Analysis

```
andrew_data <- read.csv("ForAnalysis_sorted.csv", header = TRUE)
library(lme4)
library(dplyr)
andrew_data <- andrew_data %>%
  mutate(sub_rank = dense_rank(sub)) %>%
  mutate(seg = ifelse(seg == "sh", "ʃ", seg))
library(lmerTest)
library(dplyr)
library(ggplot2)
library(emmeans)
library(gridExtra)

# Check what condition values exist in the data
print("Unique condition values:")
print(unique(andrew_data$cond))
print("Data structure:")
str(andrew_data$cond)

###################################COG####################
#cog model
model_cog <- lmer(cog ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_cog)
anova(model_cog)

# estimate means for statistical comparisons
emm_interaction <- emmeans(model_cog, ~ seg * cond)
emm_segment <- emmeans(model_cog, ~ seg)
emm_condition <- emmeans(model_cog, ~ cond)

# Create notched box plot using raw data
cog_interaction <- ggplot(andrew_data, aes(x = seg, y = cog, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
  labs(title = bquote("Centre of Gravity - " ~ scriptstyle(italic("Segment × Condition
Interaction"))),
       x = "Segment", y = "COG",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
cog_interaction

# pairwise comparisons
s_comparison <- emmeans(model_cog, ~ cond | seg)
s_pairs <- pairs(s_comparison)
print("Pairwise comparisons within each segment:")
print(s_pairs)

###################################SKEWNESS################
#skewness model
```

```
model_skew <- lmer(skew ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_skew)
anova(model_skew)

# marginal means for statistical comparisons
emm_interaction_skew <- emmeans(model_skew, ~ seg * cond)
emm_segment_skew <- emmeans(model_skew, ~ seg)
emm_condition_skew <- emmeans(model_skew, ~ cond)

# Create notched box plot using raw data
skew_interaction <- ggplot(andrew_data, aes(x = seg, y = skew, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
  ylim(0, 6) +
  labs(title = bquote("Skewness - " ~ scriptstyle(italic("Segment × Condition Interaction"))),
       x = "Segment", y = "Skewness",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
skew_interaction

# pairwise comparisons
s_comparison_skew <- emmeans(model_skew, ~ cond | seg)
s_pairs_skew <- pairs(s_comparison_skew)
print("Pairwise comparisons within each segment:")
print(s_pairs_skew)

####################################DURATION################
#duration model
model_dur <- lmer(dur ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_dur)
anova(model_dur)

# marginal means for statistical comparisons
emm_interaction_dur <- emmeans(model_dur, ~ seg * cond)
emm_segment_dur <- emmeans(model_dur, ~ seg)
emm_condition_dur <- emmeans(model_dur, ~ cond)

# Create notched box plot using raw data
dur_interaction <- ggplot(andrew_data, aes(x = seg, y = dur, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
  labs(title = bquote("Duration - " ~ scriptstyle(italic("Segment × Condition Interaction"))),
       x = "Segment", y = "Duration",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
dur_interaction
```

```
# pairwise comparisons
s_comparison_dur <- emmeans(model_dur, ~ cond | seg)
s_pairs_dur <- pairs(s_comparison_dur)
print("Pairwise comparisons within each segment:")
print(s_pairs_dur)

###################################PEAK################
#peak model
model_peak <- lmer(peak ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_peak)
anova(model_peak)

# marginal means for statistical comparisons
emm_interaction_peak <- emmeans(model_peak, ~ seg * cond)
emm_segment_peak <- emmeans(model_peak, ~ seg)
emm_condition_peak <- emmeans(model_peak, ~ cond)

# Create notched box plot using raw data
peak_interaction <- ggplot(andrew_data, aes(x = seg, y = peak, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
  labs(title = bquote("Peak - " ~ scriptstyle(italic("Segment × Condition Interaction"))),
       x = "Segment", y = "Peak",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
peak_interaction

# pairwise comparisons
s_comparison_peak <- emmeans(model_peak, ~ cond | seg)
s_pairs_peak <- pairs(s_comparison_peak)
print("Pairwise comparisons within each segment:")
print(s_pairs_peak)

###################################STANDARD DEVIATION################
#standard deviation model
model_sd <- lmer(sd ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_sd)
anova(model_sd)

# marginal means for statistical comparisons
emm_interaction_sd <- emmeans(model_sd, ~ seg * cond)
emm_segment_sd <- emmeans(model_sd, ~ seg)
emm_condition_sd <- emmeans(model_sd, ~ cond)

# Create notched box plot using raw data
sd_interaction <- ggplot(andrew_data, aes(x = seg, y = sd, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
```

```
  labs(title = bquote("Standard Deviation - " ~ scriptstyle(italic("Segment × Condition
Interaction"))),
       x = "Segment", y = "Standard Deviation",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
sd_interaction

# pairwise comparisons
s_comparison_sd <- emmeans(model_sd, ~ cond | seg)
s_pairs_sd <- pairs(s_comparison_sd)
print("Pairwise comparisons within each segment:")
print(s_pairs_sd)

###################################KURTOSIS################
#kurtosis model
model_kurt <- lmer(kurt ~ seg * cond + (1|sub_rank), data = andrew_data)
summary(model_kurt)
anova(model_kurt)

# marginal means for statistical comparisons
emm_interaction_kurt <- emmeans(model_kurt, ~ seg * cond)
emm_segment_kurt <- emmeans(model_kurt, ~ seg)
emm_condition_kurt <- emmeans(model_kurt, ~ cond)

# Create notched box plot using raw data
kurt_interaction <- ggplot(andrew_data, aes(x = seg, y = kurt, fill = cond)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("c1" = "black", "c2" = "blue", "c3" = "red"),
                    labels = c("c1" = "C1", "c2" = "C2", "c3" = "C3")) +
  scale_x_discrete(expand = expansion(mult = c(-0.1, 0.2))) +
  ylim(0, 25) +
  labs(title = bquote("Kurtosis - " ~ scriptstyle(italic("Segment × Condition Interaction"))),
       x = "Segment", y = "Kurtosis",
       fill = "Condition") +
  theme_minimal() +
  theme(legend.position = "bottom")
kurt_interaction

# pairwise comparisons
s_comparison_kurt <- emmeans(model_kurt, ~ cond | seg)
s_pairs_kurt <- pairs(s_comparison_kurt)
print("Pairwise comparisons within each segment:")
print(s_pairs_kurt)
```

Appendix G. Statistical Results

Note: (Intercept) = **/s/** Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Duration

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	1.030e-01	5.731e-03	1.030e+01	17.971	4.13e-09 ***
segsh	7.988e-16	2.195e-03	1.327e+03	0.000	1.00000
segz	-2.347e-02	2.195e-03	1.327e+03	-10.691	< 2e-16 ***
condc2	1.322e-02	2.203e-03	1.327e+03	6.001	2.53e-09 ***
condc3	4.485e-03	2.199e-03	1.327e+03	2.040	0.04156 *
segsh:condc2	5.008e-05	3.110e-03	1.327e+03	0.016	0.98715
segz:condc2	-8.763e-03	3.118e-03	1.327e+03	-2.811	0.00501 **
segsh:condc3	3.448e-03	3.107e-03	1.327e+03	1.110	0.26725
segz:condc3	-3.855e-03	3.104e-03	1.327e+03	-1.242	0.21457

Table 2. Linear Mixed-effects Models of Duration as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	-0.01322	0.00220	1327	-6.001	<.0001 ***
	C1-C3	-0.00449	0.00220	1327	-2.040	0.1032
	C2-C3	0.00873	0.00221	1327	3.958	0.0002 ***
/ʃ/	**C1-C2**	-0.01327	0.00220	1327	-6.044	<.0001 ***
	C1-C3	-0.00793	0.00220	1327	-3.614	0.0009 ***
	C2-C3	0.00533	0.00220	1327	2.430	0.0404 *
/z/	**C1-C2**	-0.00445	0.00221	1327	-2.018	0.1082
	C1-C3	-0.00063	0.00219	1327	-0.288	0.9554
	C2-C3	0.00382	0.00220	1327	1.736	0.1923

Table 3. Pairwise Comparisons of Noise Condition Effects on Duration Across Segment Types

Peak Frequency

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	6615.09	362.64	10.22	18.241	< .001 ***
segsh	–3249.47	134.99	1327	–24.072	< .001 ***
segz	113.77	134.99	1327	0.843	.399
condc2	–56.26	135.45	1327	–0.415	.678
condc3	271.30	135.22	1327	2.006	.045 *
segsh:condc2	124.11	191.23	1327	0.649	.516
segz:condc2	64.69	191.73	1327	0.337	.736
segsh:condc3	45.56	191.07	1327	0.238	.812
segz:condc3	–122.05	190.91	1327	–0.639	.523

Table 4. Linear Mixed-effects Models of Peak Frequency as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	56.26	135	1327	0.415	.9093
	C1-C3	–271.30	135	1327	–2.006	.1110
	C2-C3	–327.56	136	1327	–2.414	.0420 *
/ʃ/	**C1-C2**	–67.85	135	1327	–0.503	.8701
	C1-C3	–316.86	135	1327	–2.347	.0499 *
	C2-C3	–249.01	135	1327	–1.845	.1556
/z/	**C1-C2**	–8.43	136	1327	–0.062	.9979
	C1-C3	–149.25	135	1327	–1.107	.5096
	C2-C3	–140.82	135	1327	–1.039	.5521

Table 5. Pairwise Comparisons of Noise Condition Effects on Duration Across Segment Types

Centre of gravity

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	6811.260	280.956	11.276	24.243	4.39e-11 ***
segsh	–2586.360	137.691	1326.999	–18.784	< 2e-16 ***
segz	–2621.867	137.691	1326.999	–19.042	< 2e-16 ***
condc2	83.990	138.162	1327.003	0.608	0.5434
condc3	287.253	137.923	1327.000	2.083	0.0375 *
segsh:condc2	4.664	195.058	1327.001	0.024	0.9809
segz:condc2	241.993	195.565	1327.006	1.237	0.2162
segsh:condc3	–142.933	194.889	1327.000	–0.733	0.4634
segz:condc3	19.106	194.731	1327.001	0.098	0.9219

Table 6. Linear Mixed-effects Models of CoG as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	–84.0	138	1327	–0.608	0.8158
	C1-C3	–287.3	138	1327	–2.083	0.0938
	C2-C3	–203.3	138	1327	–1.469	0.3064
/ʃ/	**C1-C2**	–88.7	138	1327	–0.644	0.7959
	C1-C3	–144.3	138	1327	–1.048	0.5466
	C2-C3	–55.7	138	1327	–0.404	0.9139
/z/	**C1-C2**	–326.0	138	1327	–2.355	0.0489 *
	C1-C3	–306.4	137	1327	–2.229	0.0668
	C2-C3	19.6	138	1327	0.142	0.9889

Table 3. Pairwise Comparisons of Noise Condition Effects on COG Across Segment Types

Standard Deviation

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	1570.253	55.374	37.590	28.357	<2e-16 ***
segsh	–8.740	59.467	1327.007	–0.147	0.8832
segz	1223.033	59.467	1327.007	20.567	<2e-16 ***
condc2	17.127	59.670	1327.039	0.287	0.7741
condc3	44.052	59.567	1327.015	0.740	0.4597
segsh:condc2	–9.794	84.242	1327.023	–0.116	0.9075
segz:condc2	–167.869	84.460	1327.066	–1.988	0.0471 *
segsh:condc3	–80.552	84.169	1327.011	–0.957	0.3387
segz:condc3	194.573	84.101	1327.022	2.314	0.0208 *

Table 8 Linear Mixed-effects Models of SD as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	–17.13	59.7	1327	–0.287	0.9556
	C1-C3	–44.05	59.6	1327	–0.740	0.7400
	C2-C3	–26.92	59.8	1327	–0.450	0.8942
/ʃ/	**C1-C2**	–7.33	59.5	1327	–0.123	0.9917
	C1-C3	36.50	59.5	1327	0.614	0.8126
	C2-C3	43.83	59.5	1327	0.737	0.7415
/z/	**C1-C2**	150.74	59.8	1327	2.522	0.0316 *
	C1-C3	–238.62	59.4	1327	–4.019	0.0002 ***
	C2-C3	–389.37	59.7	1327	–6.525	<.0001 ***

Table 9. Pairwise Comparisons of Noise Condition Effects on Duration Across Segment Types

Skewness

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	0.48253	0.17179	14.64764	2.809	0.01346 *
segsh	0.99307	0.11992	1326.9976	8.281	2.94e-16 ***
segz	0.02493	0.11992	1326.9976	0.208	0.83532
condc2	-0.20300	0.12033	1327.0061	-1.687	0.09182 .
condc3	-0.38887	0.12012	1327.0000	-3.237	0.00124 **
segsh:condc2	-0.09233	0.16988	1327.0019	-0.544	0.58685
segz:condc2	-0.03543	0.17032	1327.0132	-0.208	0.83524
segsh:condc3	0.05401	0.16973	1327.0000	0.318	0.75039
segz:condc3	-0.15181	0.16959	1327.0016	-0.895	0.37088

Table 10. Linear Mixed-effects Models of Skewness as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	0.2030	0.120	1327	1.687	0.2105
	C1-C3	0.3889	0.120	1327	3.237	0.0035 *
	C2-C3	0.1859	0.121	1327	1.542	0.2716
/ʃ/	**C1-C2**	0.2953	0.120	1327	2.463	0.0370 *
	C1-C3	0.3349	0.120	1327	2.793	0.0147 *
	C2-C3	0.0395	0.120	1327	0.330	0.9419
/z/	**C1-C2**	0.2384	0.121	1327	1.978	0.1180
	C1-C3	0.5407	0.120	1327	4.516	<.0001 ***
	C2-C3	0.3023	0.120	1327	2.512	0.0325 *

Table 11. Pairwise Comparisons of Noise Condition Effects on Skewness Across Segment Types

Kurtosis

Predictor	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
(Intercept)	5.5685	0.7866	20.6295	7.080	6.11e-07 ***
segsh	0.9569	0.6881	1327.0043	1.391	0.16453
segz	-2.0924	0.6881	1327.0043	-3.041	0.00240 **
condc2	-1.7374	0.6904	1327.0202	-2.517	0.01197 *
condc3	-1.8998	0.6892	1327.0081	-2.756	0.00592 **
segsh:condc2	-0.5210	0.9747	1327.0123	-0.535	0.59306
segz:condc2	2.3455	0.9773	1327.0333	2.400	0.01653 *
segsh:condc3	0.2741	0.9739	1327.0062	0.281	0.77842
segz:condc3	-0.7676	0.9731	1327.0119	-0.789	0.43038

Table 12. Linear Mixed-effects Models of Kurtosis as a Function of Segment Type and Noise Condition

Segment	Contrast	*β*	*SE*	*df*	*t*	*Pr(>|t|)*
/s/	**C1-C2**	1.737	0.690	1327	2.517	0.0321 *
	C1-C3	1.900	0.689	1327	2.756	0.0163 *
	C2-C3	0.162	0.692	1327	0.235	0.9701
/ʃ/	**C1-C2**	2.258	0.688	1327	3.282	0.0030 **
	C1-C3	1.626	0.688	1327	2.363	0.0480 *
	C2-C3	-0.633	0.688	1327	-0.920	0.6279
/z/	**C1-C2**	-0.608	0.692	1327	-0.879	0.6536
	C1-C3	2.667	0.687	1327	3.883	0.0003 ***
	C2-C3	3.275	0.690	1327	4.743	<.0001 ***

Table 13. Pairwise Comparisons of Noise Condition Effects on Kurtosis Across Segment Types