Jeremy Calder* and Rebecca Wheeler

# Is Zoom viable for sociophonetic research? A comparison of in-person and online recordings for sibilant analysis

https://doi.org/10.1515/lingvan-2021-0014
Received January 27, 2021; accepted September 2, 2021; published online February 22, 2022

**Abstract:** This study is part of a larger project investigating whether Zoom is a viable data collection method for sociophonetic research, examining whether Zoom recordings yield different acoustic measurements than in-person recordings for the exact same speech for 18 speakers. In this article we analyze five spectral measures of sibilants (peak, center of gravity, standard deviation, skewness, and kurtosis) which have been shown to be conditioned by dimensions of identity like speaker gender and sexual orientation in much previous sociolinguistic research. We find that, overall, Zoom recordings yield significantly lower peak, center of gravity, and standard deviation measurements and significantly higher skewness and kurtosis values than in-person recordings for the same speech, likely due to a lower sampling rate on Zoom recordings. However, a preliminary analysis controlling for sampling rate across recording methods reveals the opposite patterns for nearly all measures, suggesting that Zoom stretches the spectral space when compared with the in-person recorder. Because the values of these measurements can lead analysts to draw social interpretations relating to a speaker's performance of gender and sexual identity, we caution against comparing across Zoom and in-person recordings, as differences in measurements may result from the recording method used to collect the data.

**Keywords:** COVID-19; methods; sibilants; sociophonetics; Zoom

## 1 Introduction

Sociophonetic work has largely relied on the ability to collect in-person speech recordings to analyze. With the advent of the COVID-19 pandemic, an airborne virus that is more easily spread when human beings are within physical proximity of one another, the ability to collect in-person recordings has been unquestionably hampered. Sociolinguists have thus been exploring other methods for data collection. Zoom, an online videoconferencing service which has all but become the primary means of social interaction for many in a post-COVID world, is one possible tool for collecting such data. But as sociolinguists are primarily interested in exploring the possible social factors that influence the linguistic variables we analyze, in order to be sure the phonetic patterns we observe can be situated within social interpretations, we want to ensure they are not merely artifacts resulting from acoustic differences between online and in-person recordings. This raises the question whether Zoom and in-person recordings yield comparable methods for phonetic measurements. And if not, how different are they? While previous important work (e.g., DeDecker and Nycz 2011; Hall-Lew and Boyd 2017, 2020) has explored the effect of data collection methods on sociophonetic measurements, the ubiquity of Zoom is relatively recent, and remains ripe for exploration as a possible method for sociophonetic data collection (e.g., Freeman and DeDecker 2020).

In a previous article we explored differences between Zoom recordings and in-person recordings for vocalic measurements of the same speech (Calder et al. in press). We found discrepancies between recording

***Corresponding author: Jeremy Calder**, Department of Linguistics, University of Colorado Boulder, 295 UCB, Boulder, CO 80309-0401, USA, E-mail: jeca9599@colorado.edu
**Rebecca Wheeler,** Department of Linguistics, University of Colorado Boulder, Hellems 290, 295 UCB, Boulder, CO 80309, USA

methods for vowel formant measurements, such that Zoom exhibited lower F1 values and higher F2 vowels than the in-person recorder. Recording methods seemed to converge around 1,200 Hz, with measurement discrepancies increasing the farther away from 1,200 Hz the measure was. Here, we expand our analysis to include acoustic measurements of a consonantal variable: spectral measurements of the sibilant /s/ in spoken English. Given that sibilants occupy a much higher-frequency range than vowel formants (i.e., occupying a range much farther from 1,200 Hz than vowel formants), we may expect to see a discrepancy between recording methods of an even larger magnitude than was found for vowel formants. We chose the sibilant /s/ as it is perhaps the most robustly studied consonantal variable that has been linked to social patterns relating to gender and sexual identities, and has even been called "the most perfect of signs" (Eckert 2017) and sociophonetics' "most iconic variable" (Calder 2020). The voiceless anterior sibilant, articulated by placing the tongue behind the upper incisors and passing air over it, has exhibited robust patterns conditioned by speaker gender, such that women have been shown to exhibit more fronted articulations of /s/ than men with the tongue closer to the upper incisors (Fuchs and Toda 2010), resulting in higher-frequency frication (Flipsen et al. 1999; Jongman et al. 2000; Stuart-Smith 2007). In addition, sexual orientation has been shown to condition /s/ realization as well, with queer speakers exhibiting different acoustic patterns from straight speakers in multiple studies (e.g., Munson et al. 2006; Podesva and Van Hofwegen 2014; Smyth and Rogers 2008).

Five spectral measurements have been shown in English to exhibit strong patterns conditioned by gender and other dimensions of social identity: peak frequency and the four spectral moments, center of gravity (COG), standard deviation, skewness, and kurtosis. COG, which represents the mean of where spectral energy of a sound is focused, correlates with frontness of articulation if all else is equal, such that a higher COG corresponds to a more fronted /s/ (though COG can also be affected by greater emphasis or a smaller constriction area). Peak frequency, which is the frequency with the highest amplitude peak in the spectrum, also corresponds with frontness of articulation, such that a higher peak corresponds with a more anterior constriction (Jongman et al. 2000). Standard deviation represents the variance in spectral energy, and while it has been argued in some work not to acoustically differentiate obstruents (e.g., Forrest et al. 1988), other work has argued that standard deviation can correlate with frontness of articulation, such that /s/ exhibits a lower standard deviation than /ʃ/ (e.g., Tomiak 1990). Standard deviation has also been shown to differentiate sibilants from non-sibilant fricatives, such that non-sibilants exhibit a greater standard deviation than sibilants (Shadle and Mair 1996). Skewness, which represents spectral tilt and corresponds with how close the mean is to the center of the frequency range, negatively correlates with /s/ frontness, such that a more negative skewness corresponds with a more fronted articulation (Forrest et al. 1988). Finally, kurtosis captures how peaked the acoustic waveform is, with higher kurtosis values representing a more peaked spectral shape, and lower kurtosis values representing a flatter spectral shape (Jongman et al. 2000).

In addition, many of these measures have been shown to pattern with gender identity in previous research, such that men (overall) have been shown to exhibit lower COG (Hazenberg 2012; Podesva and Van Hofwegen 2014), lower peak (Levon and Holmes-Elliott 2013; Stuart-Smith 2007; Stuart-Smith et al. 2003), higher skewness (Hazenberg 2012), and lower kurtosis (Jongman et al. 2000) values than women. In addition, sexual orientation (Hazenberg 2012; Podesva and Van Hofwegen 2016), visual gender presentation (Calder 2019a, 2019b), social class (Stuart-Smith et al. 2003), and ethnicity (Calder and King 2020; Pharao et al. 2014) have been shown to condition these spectral measures in multiple studies. Given the robust strands of research exploring the social dimensions of identity that condition these measures, we explore here whether recording method (Zoom vs. in-person) yields different values for the exact same speech, and whether recordings collected across the two methods are comparable with each other.

## 2 Methods

The data are shared with a previous project exploring the effects of Zoom recordings on vowel formant measurements (Calder et al. in press). The data come from conversations between 18 graduate students enrolled in a sociophonetic analysis seminar. The speakers include six males, 11 females, and one nonbinary participant, all in their 20s and 30s, from a range of geographic backgrounds (mostly the United States, but one participant was from Spain and one from China). The speaker group is largely queer, with 14 out of 18 participants not identifying as heterosexual. However, as this is primarily an intra-speaker study, demographic information is largely incidental.

Conversations were dyadic, with pairs of participants conversing on Zoom in English. Participants recorded themselves on Zoom at the same time as they recorded themselves with in-person equipment. For in-person recorders, most speakers used the Olympus 822/823/852/853 series of portable audio recorders, a relatively inexpensive model of recorder that enables recording at a 44.1 kHz sampling rate. Three speakers used different recorders as the Olympus series was not available in their geographic locations at the time of recording: the TASCAM DR-100MKII, the Voicetracer DVT 2050, and the Philips VTR8060. However, the specific recorder used did not appear to significantly predict acoustic measurements in preliminary statistical models. All speakers used SLINT omnidirectional condenser lapel microphones affixed to their chest a few inches from their mouth. In-person equipment recorded at a sampling rate of 44.1 kHz and a bit depth of 16-bit, while Zoom recordings were taken at 32.0 kHz (the maximum sampling rate allowed by Zoom at the time of writing) and a bit depth of 16-bit. For our primary analysis, we did initially downsample the in-person recordings, as 44.1 kHz is the standard for sociophonetic analysis (DeDecker and Nycz 2013), and we wanted to test whether Zoom recordings were comparable to this standard. However, in order to explore whether Zoom and in-person recordings differed when sampling rate was controlled for, we did conduct a preliminary analysis controlling for sampling rate, which we discuss in Section 4 below. However, the bulk of our discussion will focus on our primary analysis using the unmodified recordings.

Zoom's recording settings were set to their defaults, meaning that automatic compression and gain control were enabled. We did not control for the equipment used to record on Zoom, as it is unlikely that a sociolinguist would be able to control for the computers and microphones available to their online subjects in the field. However, we do code for various aspects of the Zoom setup, including whether the speaker wore headphones during the Zoom conversation, whether the speaker recorded on Zoom with the built-in computer microphone or with an external microphone (like earbuds or a gaming headset), and whether the speaker was the host of the Zoom conversation (meaning their audio was recorded through their own computer, rather than secondhand through the other person's computer audio). All conversations were transcribed in ELAN and aligned using FAVE (Rosenfelder et al. 2015).

Praat scripts automatically extracted /s/ tokens from each sound file. We then manually accepted at least 30 tokens for each recording for each speaker, hand-correcting the interval boundaries for all accepted tokens using visual cues from the Praat wideband spectrogram (i.e., the duration of high-frequency frication minus adjacent voicing pulses, stop closures, and release bursts). We accepted exactly the same tokens for each recording for each speaker, took care to use exactly the same segmentation standards across recording methods, and excluded tokens with overlapping speech or unclear segmentation boundaries. Praat scripts were then used to automatically collect peak, COG, standard deviation, skewness, and kurtosis measurements for each accepted token of /s/ in each recording. Peak measurements were collected at the highest amplitude peak corresponding to the lowest-frequency main resonance, following previous research (e.g., Jesus and Shadle 2002; Jongman et al. 2000; Stuart-Smith 2007). The four spectral moments were measured in two ways for each token: (1) within a 40 ms Hamming window centered at the segment midpoint (e.g., Podesva and Van Hofwegen 2014, 2016); and (2) using spectral averaging (e.g., DiCanio 2013; Shadle 2012). For spectral averaging, the duration of each token was divided into six shorter intervals of 15 ms, a discrete Fourier transform (DFT) was computed for each smaller interval, and the DFTs for each token were averaged. Given the window size of 15 ms, we excluded all tokens under 56 ms to avoid windows overlapping more than 50 percent. For both types of spectral measurements, all tokens were high-pass filtered with a 1,000 Hz cutoff, in order to filter out frequencies too low to be within the range of /s/ frication. Exactly the same tokens (all eligible tokens) were analyzed in both recording methods (Zoom and in-person), with a total of 1,202 tokens measured and analyzed.

The data for each of the nine spectral measurements (peak, COG midpoint and averaging, standard deviation midpoint and averaging, skewness midpoint and averaging, and kurtosis midpoint and averaging) were fitted to mixed effects linear regression models in R, with each spectral measurement serving as a dependent variable for its own model. Fixed effects included recording method (Zoom vs. in-person), speaker gender, and log-transformed duration (which has also been shown to predictably condition spectral measurements in previous research, as discussed above). We included gender in our statistical models to ensure that our data patterned in a predictable way, given the robust gendered patterns conditioning spectral moments that have been established in the literature discussed above. Random effects included speaker, word, and the region the speaker was from. Finally, we were interested in testing which factors significantly interacted with the recording method, so we explored a number of interactions in our models, including: recording method and gender (six men, 11 women, one nonbinary); recording method and whether the speaker was the host of the Zoom conversation (nine hosts, nine non-hosts); and recording method and Zoom recording setup (13 recording with headphones and an external microphone, three recording with headphones and the computer-internal microphone, and two recording without headphones and with the computer-internal microphone).

## 3 Results

The significant effects from each of the mixed effects models are presented in Table 1, with medians and quartiles for each measure plotted in Figure 1. Importantly, recording method emerges as a significant predictor for nearly all of the spectral measurements, such that Zoom yields lower peak, COG, and standard deviation measurements (for both the midpoint and spectral averaging methods) than the in-person recorder, and Zoom yields higher skewness (spectral averaging) and kurtosis (both midpoint and spectral averaging) measurements than the in-person recorder. As peak and COG have been argued to correlate with /s/ frontness, and skewness has been argued to correlate inversely with /s/ frontness, it seems that Zoom recordings yield values for each of these measurements that could be interpreted as more retracted than the recordings from the in-person recorder. As kurtosis correlates with peakedness of spectral shape, it appears that the Zoom recording yields spectra for /s/ that are more peaked than the in-person recorder. And as standard deviation is lower in the Zoom recording than the in-person recording, this suggests that Zoom fricatives have a lower range of spectral energy than in-person recordings.

In addition to the significant effects of recording method, gender significantly predicts peak and COG, such that men exhibit lower COG and peak than other speakers. These are patterns in the expected direction, as men have been shown to exhibit measurements suggesting more retracted articulations of /s/ (i.e., lower COG, lower peak) than other speakers in multiple studies (e.g., Flipsen et al. 1999; Jongman et al. 2000; Podesva and Van Hofwegen 2014). Duration also emerges as a significant predictor of peak, COG, and standard deviation, with longer tokens of /s/ predicting higher peak, higher COG, and lower standard deviation. In other words, longer tokens of /s/ are more fronted with a lower range of spectral energy distribution overall (which is consonant with patterns found in previous production work, e.g., Calder 2019a; Podesva and Van Hofwegen 2016).

Table 2 displays the mean values for each measure by recording method, illustrating that the difference in measurements for peak, COG, and standard deviation between the Zoom and in-person recorder is about 600 Hz in magnitude. This may suggest that, for these measures, ~600 Hz would need to be added to Zoom values to be comparable to in-person values, but for skewness and kurtosis, there is a less consistent mathematical relationship between Zoom and in-person measurements.

We now consider each of the significant interactions, starting with the interaction between the recording method and the Zoom recording setup. For the Zoom recording setup, speakers were divided into three groups: those who wore headphones and recorded on Zoom using an external microphone like earbuds or a gaming headset; those who wore headphones and recorded on Zoom using a computer-internal microphone; and those who did not wear headphones and recorded on Zoom using a computer-internal microphone. Figure 2 plots the effect of this interaction on each of the spectral measurements. For all measurements for which this interaction emerged as significant, it appears that discrepancies between recording measurements are largest for speakers not wearing headphones and recording with the computer-internal microphone during the Zoom

**Table 1:** Significant model coefficients for linear mixed effects models for spectral measurements. (see Tables 1–9 in Appendix for complete regression tables. Significance stars signify the following: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. A dot (.) signifies $p < 0.10$, suggesting the factor approaches significance but not does reach the significance threshold of $p < 0.05$).

| Measurement | Recording (Zoom) | Gender (M) | Gender (M) × Recording (Zoom) | Logged duration | Headphones and int mic × Recording (Zoom) | No headphones × Recording (Zoom) | Host (Y) × Recording |
|---|---|---|---|---|---|---|---|
| Peak (Hz) | −874.68 (***) | −1,128.55 (.) | | 530.89 (**) | | −849.10 (**) | 705.09 (***) |
| Center of gravity (midpoint; Hz) | −989.37 (***) | −1,306.72 (*) | 267.95 (*) | 454.99 (***) | 595.14 (***) | −411.66 (**) | 401.74 (***) |
| Center of gravity (averaging; Hz) | −1,024.21 (***) | −1,455.38 (*) | 403.11 (***) | 390.62 (***) | 629.76 (***) | −407.31 (*) | 391.62 (**) |
| Standard deviation (midpoint) | −564.44 (***) | | −219.51 (***) | −224.49 (***) | 351.73 (***) | −250.69 (***) | |
| Standard deviation (averaging) | −620.58 (***) | | −251.08 (***) | −172.28 (***) | 438.04 (***) | −237.92 (**) | |
| Skew (midpoint) | | | | | | 1.39 (***) | −0.47 (***) |
| Skew (averaging) | 0.56 (***) | | | | −0.56 (**) | 0.94 (***) | −0.48 (**) |
| Kurtosis (midpoint) | 1.44 (*) | | −2.23 (**) | | | | |
| Kurtosis (averaging) | 8.59 (***) | | | | −7.85 (***) | −5.67 (*) | |

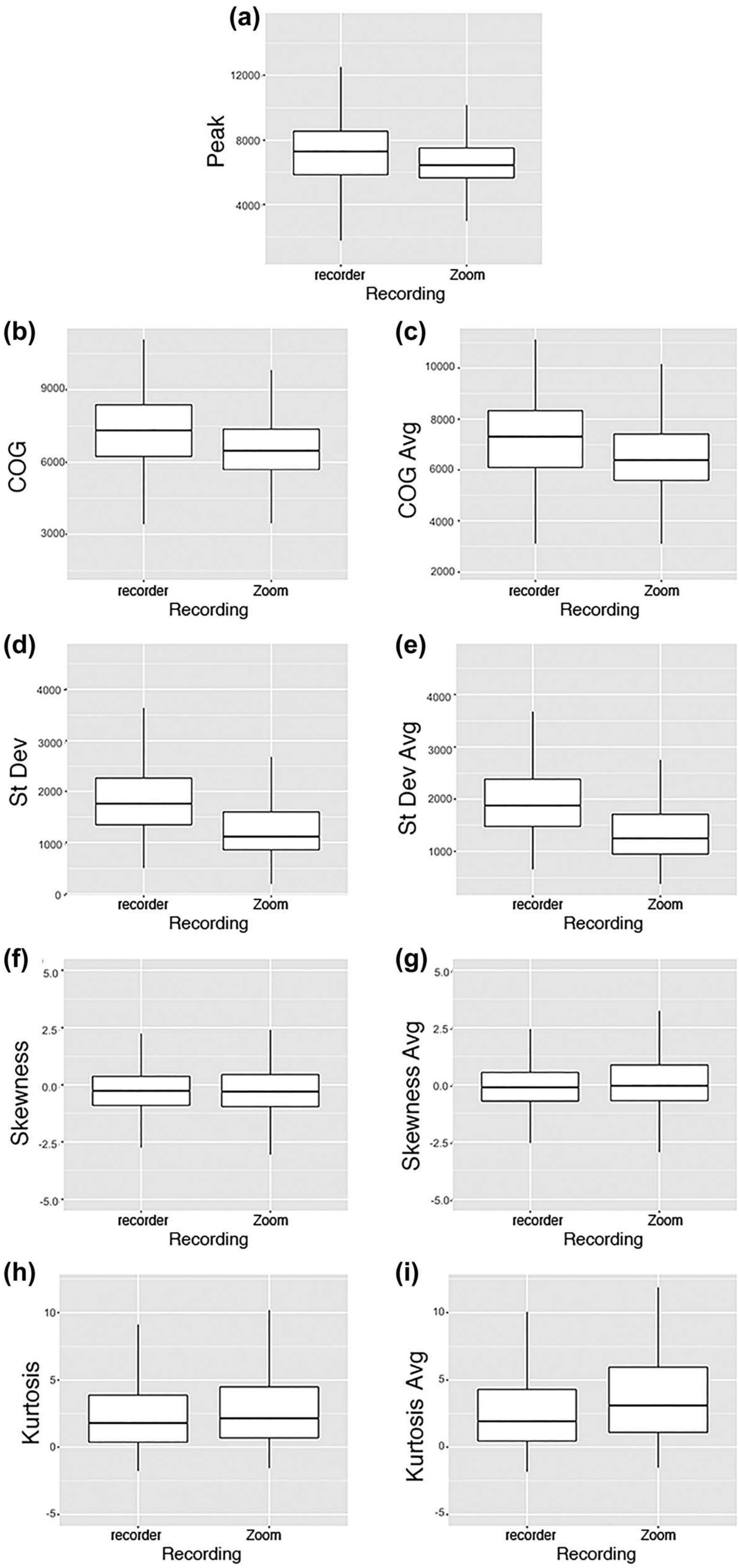

**Figure 1:** Spectral measure by recording method: (a) peak frequency; (b) center of gravity (midpoint); (c) center of gravity (averaging); (d) standard deviation (midpoint); (e) standard deviation (averaging); (f) skewness (midpoint); (g) skewness (averaging); (h) kurtosis (midpoint); (i) kurtosis (averaging).

**Table 2:** Spectral measurement means by recording method.

| Measurement | Recording (Zoom) | Recording (Recorder) |
|---|---:|---:|
| Peak (Hz) | 6,575.756 | 7,158.532 |
| Center of gravity (midpoint; Hz) | 6,615.624 | 7,274.226 |
| Center of gravity (averaging; Hz) | 6,541.507 | 7,196.291 |
| Standard deviation (midpoint) | 1,245.02 | 1854.578 |
| Standard deviation (averaging) | 1,365.694 | 1978.252 |
| Skew (midpoint) | −0.269 | −0.247 |
| Skew (averaging) | 0.192 | −0.051 |
| Kurtosis (midpoint) | 4.164 | 3.3 |
| Kurtosis (averaging) | 10.34 | 4.153 |

call. This may be due to the greater amount of sound coming out of the computer speakers during the Zoom call, which would automatically trigger Zoom's compression and gain control feature, perhaps affecting noisy segments like sibilants to a greater degree. Although we excluded tokens with overlapping speech and interfering noise, it is possible that adjacent noise or overlapping speech triggered Zoom's automatic filters, which persisted over the duration of nearby /s/ segments, even if the /s/ segments themselves did not contain overlapping noise.

Perhaps surprisingly, speakers wearing headphones and recording on Zoom with a computer-internal microphone seem to exhibit smaller differences between Zoom and in-person measurements than speakers wearing headphones and recording on Zoom with an external microphone. One possibility to explain this is that those recording with external microphones on Zoom were using gaming headsets and earbuds, whose microphones are placed relatively close to the speaker's mouth. This could mean that fricatives would be more acoustically prominent, given the close proximity of the microphone to the speech source (see, e.g., Titze and Winholtz 1993; Svec and Granqvist 2010 for a discussion of the effects of proximity on acoustic prominence of speech recordings), and this increased acoustic prominence could have triggered Zoom's noise reduction feature, lowering the intensity of high-frication noise, and thus driving down spectral measurements on Zoom even further when compared to the in-person recordings. It is also worth mentioning that only three speakers were wearing headphones and recording with a computer-internal microphone, as opposed to a group of 13 recording with headphones and an external microphone. Future work balancing the sample for different microphone types and headphone setups is needed to further probe the reasons for these differences.

Figure 3 plots significant interactions between recording method and whether or not the speaker was the host of the Zoom conversation. Overall, for those measures for which this interaction emerges as significant, it appears that those who hosted the Zoom conversation exhibit smaller discrepancies between recording methods than those who did not host the Zoom conversation. In other words, those speakers whose speech in the Zoom conversation was recorded on their own computers exhibited Zoom measurements that were closer to in-person measurements than those speakers whose speech was recorded secondhand through the other person's computer. It could be that Zoom affected the recordings of non-hosts to a greater degree since their speech was being streamed through the internet before being recorded, rather than simply being recorded directly through their own equipment, which could have reduced the sound quality of the recordings for non-hosts when compared to hosts.

Finally, Figure 4 plots significant interactions between recording method and speaker gender. Overall, it appears than men exhibit smaller discrepancies between recording methods than women for COG (midpoint and averaging) and kurtosis (midpoint), while women exhibit smaller discrepancies between recording methods for standard deviation (midpoint and averaging). As there was only one nonbinary speaker in the sample, we refrain from making any generalizations about the nonbinary speaker's patterns in relation to the men and women in the sample. Further research is needed to probe why different genders would exhibit larger discrepancies between recording methods for different variables.

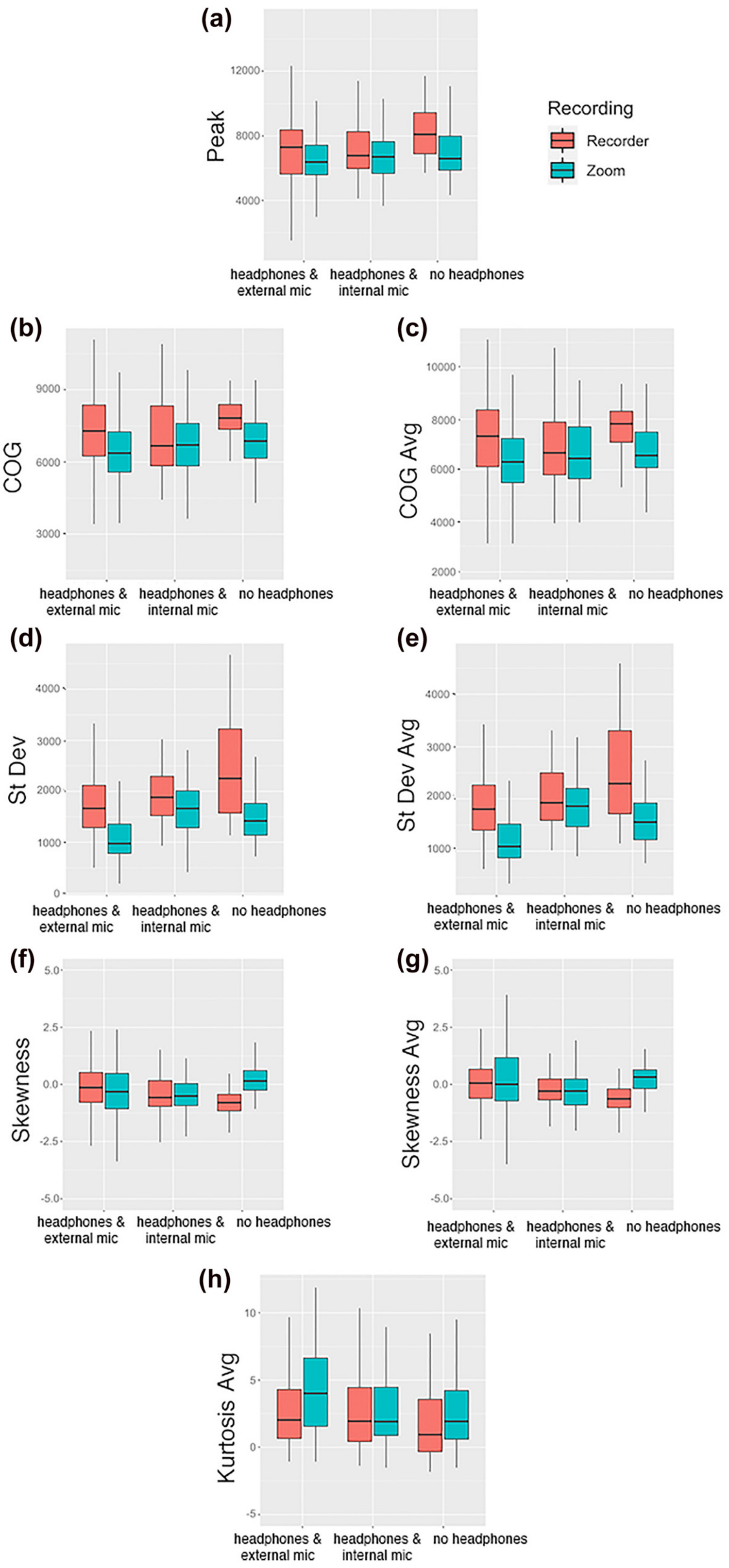

**Figure 2:** Significant interactions between recording method and Zoom setup on: (a) peak; (b) center of gravity (midpoint); (c) center of gravity (averaging); (d) standard deviation (midpoint); (e) standard deviation (averaging); (f) skewness (midpoint); (g) skewness (averaging); (h) kurtosis (averaging).

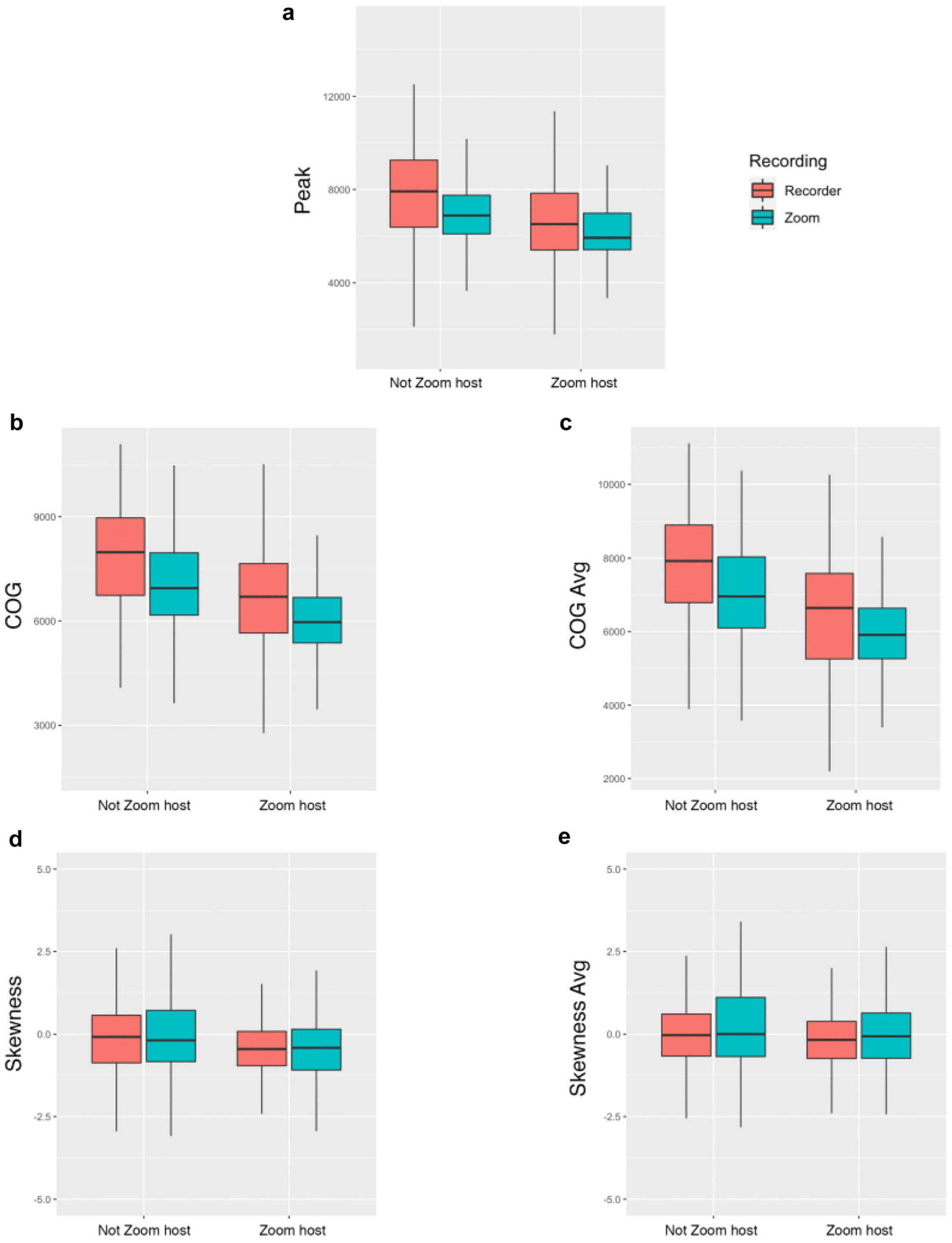

**Figure 3:** Significant interactions between recording method and Host (Y/N) on: (a) peak; (b) center of gravity (midpoint); (c) center of gravity (averaging); (d) skewness (midpoint); (e) skewness (averaging).

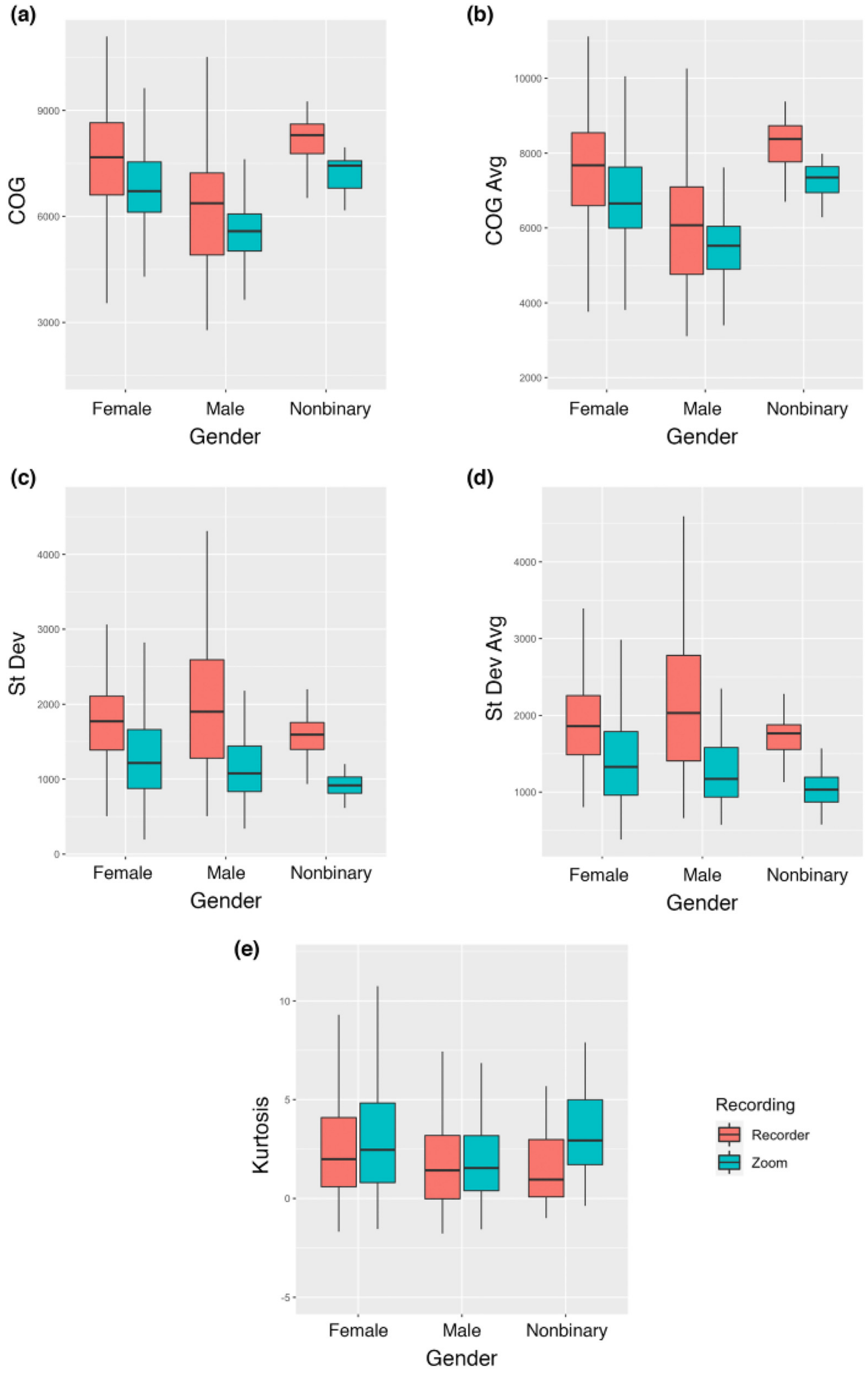

**Figure 4:** Significant interactions between recording method and gender on: (a) center of gravity (midpoint); (b) center of gravity (averaging); (c) standard deviation (midpoint); (d) standard deviation (averaging); (e) kurtosis (midpoint).

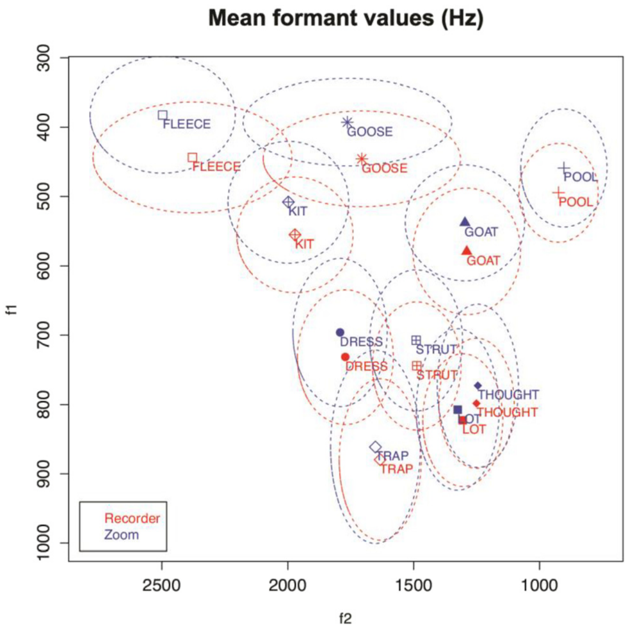

**Figure 5:** Mean vowel formant values in hertz (from Calder et al. in press).

# 4 Conclusions

In sum, Zoom recordings yield measurements that could lead to interpretations suggesting that sibilants are more retracted than they are, when compared to measurements taken from in-person recordings. Specifically, Zoom yields lower peak, lower COG, and higher skewness measurements than the in-person recorder. Zoom also yields higher kurtosis than the in-person recorder, meaning peaks are more prominent in the acoustic waveform for /s/ tokens, and lower standard deviation than the in-person recorder, meaning Zoom measurements have a lower range of spectral energy distribution than in-person measurements.

One possible explanation for the difference in measurements between the two recording methods is the difference in sampling rate. While the in-person recordings were collected at a sampling rate of 44.1 kHz, the maximum sampling rate enabled by Zoom software at the time of the data collection was 32 kHz. A lower sampling rate also results in a lower Nyquist frequency (Audacity 2020) – that is, a lower maximum frequency that can be acoustically captured by the recording. What this means is that Zoom recordings do not contain spectral information above 16,000 Hz, while in-person recordings contain information up to 22,050 Hz. Given the greater range of higher-frequency information that is possible in in-person recordings of 44.1 kHz, it is likely that this high-frequency information that is only present in the in-person recordings contributes to higher values for spectral measures (see also Shadle and Mair 1996), resulting in higher peak frequency, COG, and standard deviation and lower skewness measurements (suggesting that there is more spectral information above the mean than below the mean) than in the Zoom recordings.

However, the possibility remains that, due to Zoom's automatic compression and gain reduction features, there are effects on spectral measurements beyond those that are simply due to differences in Nyquist frequency between the recording types. As a preliminary exploration probing whether discrepancies between Zoom and in-person recordings remained when Nyquist frequency is controlled for, we collected measurements for peak and the four spectral moments at the midpoint for all tokens of /s/ in the data set (using intervals automatically generated by the forced aligner, $n$ = 12,270), and fit the data for each of these measures to statistical models like those detailed in the Methods section above.[1] Interestingly, when the hertz ceiling was capped at 16 kHz for both recording methods, many of the patterns differentiating Zoom measurements from in-person measurements patterned in the opposite direction than when the hertz ceiling was

[1] See Tables 10–14 in Appendix for complete regression tables.

not controlled for. Specifically, when measurements excluded information above 16 kHz for both recording types, Zoom yielded higher peak (estimate = 1,115.01, $p < 0.001$) and COG (estimate = 901.458, $p < 0.001$) than in-person recordings, and Zoom yielded lower standard deviation (estimate = –1,115.09, $p < 0.0001$), skewness (estimate = –0.753, $p < 0.001$), and kurtosis (estimate = –2.64, $p < 0.001$) than in-person recordings. These findings suggest that Zoom stretches the spectral range when compared to in-person recordings.

Situating these findings within the patterns found in our previous analysis of vocalic measurements across recording methods better illuminates this stretching of the spectral space on Zoom. In our analysis of vowels (Calder et al. in press), we found that Zoom yielded lower first formant (F1) values and higher second formant (F2) values than the in-person recordings for the exact same vowel tokens. The magnitude of these differences increased the lower the F1 of the vowel class was and the higher the F2 of the vowel class was. Figure 5, reproduced below from our previous work, shows that the measurements for F1 and F2 are most similar across recording methods at around 1,200 Hz. Any disparities between recording methods above or below this 1,200 Hz frequency seem to increase the farther away from this point they are; that is, Zoom measurements of formants below 1,200 Hz increasingly become lower than traditional recording measurements the farther away from 1,200 Hz they are, and Zoom measurements of formants above 1,200 Hz increasingly become higher than traditional recording measurements the farther away from 1,200 Hz they are. We see that the largest magnitude difference between Zoom and traditional measurements for vowel formants is in the lower hundreds of hertz. Sibilant measures, which occupy a frequency range much higher than those occupied by vowel formants – and are thus even farther from the 1,200 Hz frequency – exhibit even greater degrees of disparity between recording methods, with peak, COG, and standard deviation measurements in the Zoom recording being around 1,000 Hz greater than those same measurements in the traditional recording, when the hertz ceiling is the same across recording methods. In other words, because sibilants occupy a higher frequency range than the first and second vowel formants do – occupying a range much farther from 1,200 Hz – the disparity in measurements between recording methods for sibilants is even greater than it is for vowels.

It is possible that Zoom's automatic compression and gain reduction settings contribute to this stretching of the frequency range on Zoom, such that higher-frequency measures tend to be even higher in Zoom recordings. Previous work has shown that an increase in high-frequency white noise can raise acoustic measurements, while an increase in lower-pitched humming noises can lower acoustic measurements (DeDecker 2016). It is possible that Zoom's compression features increase white noise relative to speech sounds, decrease frication noise relative to ambient white noise, or decrease lower-pitched ambient noises relative to speech sounds, all of which would result in a raising of spectral measures like peak and COG, for instance. Future work is needed to explore in detail the factors that contribute to the stretching of the frequency range on Zoom, and how differences in sampling rate between the recordings interact with this spectral distortion. In sum, from our preliminary findings, it seems that the stretching of the spectral range on Zoom leads to higher-frequency information being even higher on Zoom than in the in-person recordings, but the difference in sampling rate between Zoom and in-person recordings means that there is even more spectral information above 16 kHz in the in-person recordings, which leads spectral measures to be higher in the in-person recordings when sampling rate is not controlled for.

What does this mean for using Zoom as a data source for sociophonetic analysis of sibilants? For one, if we do not control for sampling rates across recording types, we may be led to believe that speakers recorded via Zoom exhibit more retracted /s/ productions than speakers recorded in person, given the lower frequency of spectral measures on Zoom than on the in-person recordings. And when sampling rates are controlled for, we may assume that speakers recorded via Zoom exhibit more fronted /s/ productions than speakers recorded in person, given the spectral stretching resulting in higher-frequency measurements on Zoom than in the in-person recordings. The magnitude of these differences between recording types is not negligible, being about 600 Hz when sampling rate is not controlled for, and about 1,000 Hz when sampling rate is controlled for. Because of this, we cannot recommend making socially meaningful comparisons of spectral measurements between speakers across Zoom and in-person recordings. Any social interpretations of spectral patterns should be restricted to data collected within the same recording method: that is, speakers recorded on Zoom should be compared with other speakers recorded on Zoom, and speakers recorded in-person should be compared with other speakers recorded in-person.

Finally, we want to acknowledge that our results suggest that certain aspects of the Zoom recording setup can affect the disparity between Zoom and in-person recordings with respect to spectral measurements. Speakers who hosted Zoom conversations appeared to yield spectral measurements on Zoom that were more comparable to in-person measurements, than speakers who did not host Zoom conversations. And speakers who recorded on Zoom without using headphones seemed to exhibit larger discrepancies between recording methods than speakers who recorded with headphones. Even so, as Figures 2 and 3 show, there are still some apparent disparities between recording methods that remain, even with a more optimal Zoom recording setup involving headphones and the speaker recording the Zoom data on their own computer. A potential avenue for future research would be to control for the equipment setup on Zoom in comparisons between Zoom and in-person recordings, to confirm whether or not disparities still remain, as they seem to in our data. One limitation of our project was that only half of our participants recorded the Zoom conversations on their own computer; future work may explore whether measurements differ for the same speaker when recorded on the speaker's own computer versus another speaker's computer. Another limitation was that we were unable to control for the specific equipment (e.g., computers, sound cards, microphones) available to each person in the Zoom conversation, which limits our ability to more systematically explore the ways the Zoom recording setup affects measurement discrepancies. There is much potential for future work to explore the effects of different equipment setups in a more controlled way.

Overall, we are reluctant to suggest at this time that researchers make sociolinguistic comparisons between sibilants recorded on Zoom and sibilants recorded in-person. Because the values of sibilant measurements can lead analysts to draw social interpretations relating to a speaker's performance of gender and sexual identity (as well as other dimensions of identity like class and ethnicity, as discussed above), we caution against using comparisons across Zoom and in-person recordings to draw social conclusions, as differences in measurements may be an artifact of the recording method used to collect the data, rather than solely relating to a speaker's performance of identity.

## References

Audacity. 2020. Sample rates. *Audacity 2.4.2 manual*. Available at: https://manual.audacityteam.org/man/sample_rates.html.

Calder, Jeremy. 2019a. The fierceness of fronted /s/: Linguistic rhematization through visual transformation. *Language in Society* 48(1). 31–64.

Calder, Jeremy. 2019b. From sissy to sickening: The indexical landscape of /s/ in SoMa, San Francisco. *Journal of Linguistic Anthropology* 29(3). 332–358.

Calder, Jeremy. 2020. From "gay lisp" to "fierce queen": The sociophonetics of sexuality's most iconic variable. In Kira Hall & Rusty Barrett (eds.), *The Oxford handbook of language and sexuality*. Oxford: Oxford University Press.

Calder, Jeremy & Sharese King. 2020. Intersections between race, place, and gender in the production of /s/. *University of Pennsylvania Working Papers in Linguistics* 26(2).

Calder, Jeremy, Rebecca Wheeler, Sarah Adams, Daniel Amarelo, Katherine Arnold-Murray, Justin Bai, Meredith Church, Josh Daniels, Sarah Gomez, Jacob Henry, Yunan Jia, Brienna Johnson-Morris, Kyo Lee, Kit Miller, Derrek Powell, Merlin Ramsey, Sydney Rayl, Sarah Rosenau & Nadine Salvador. In press. Is Zoom viable for sociophonetic research? A comparison of in-person and online recordings for vocalic analysis. *Linguistics Vanguard*.

DeDecker, Paul. 2016. An evaluation of noise on LPC-based vowel formant estimates: Implications for sociolinguistic data collection. *Linguistics Vanguard* 2(1). 1–19.

DeDecker, Paul & Jennifer Nycz. 2011. For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics* 17(2).

DeDecker, Paul & Jennifer Nycz. 2013. The technology of conducting sociolinguistic interviews. In Christine Mallinson, Becky Childs & Gerard Van Herk (eds.), *Data collection in sociolinguistics: Methods and applications*, 123–130. New York: Routledge.

DiCanio, Christian. 2013. Spectral moments of fricative spectra script [Praat script]. https://www.acsu.buffalo.edu/~cdicanio/scripts.html (accessed 15 December 2020).

Eckert, Penelope. 2017. The most perfect of signs: Iconicity in variation. *Linguistics* 55(5). 1197–1207.

Flipsen, Peter, Jr., Shriberg Lawrence, Weismer Gary, Heather Karlsson & Jane McSweeny. 1999. Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research* 42(3). 663–667.

Forrest, Karen, Weismer Gary, Milenkovic Paul & Ronald N. Dugall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84(1). 115–123.

Freeman, Valerie & Paul DeDecker. 2020. Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *Journal of the Acoustical Society of America* 149. 1211–1223.

Fuchs, Susanne & Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In Susanne Fuchs, Martine Toda & Marzena Zygis (eds.), *An interdisciplinary guide to turbulent sounds*, 281–302. Berlin: DeGruyter Morton.

Hall-Lew, Lauren & Zac Boyd. 2017. Phonetic variation and self-recorded data. *University of Pennsylvania Working Papers in Linguistics* 23(2). 85–95.

Hall-Lew, Lauren & Zac Boyd. 2020. Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard* 6(s1).

Hazenberg, Evan. 2012. *Language and identity practice: A sociolinguistic study of gender in Ottawa, Ontario*. St. Johns: Memorial University of Newfoundland MA thesis.

Jesus, Luis M. T. & Christine H. Shadle. 2002. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30. 437–464.

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108(3). 1252–1263.

Levon, Erez & Sophie Holmes-Elliott. 2013. East End boys and West End girls: /s/-fronting in Southeast England. *University of Pennsylvania Working Papers in Linguistics* 19(2).

Munson, Benjamin, Elizabeth C. McDonald, Nancy L. DeBoe & Aubrey R. White. 2006. The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 34(2). 202–240.

Pharao, Nicolai, Marie Maegaard, Janus Spindler Moller & Tore Kristiansen. 2014. Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43. 1–31.

Podesva, Robert J. & Janneke Van Hofwegen. 2014. How conservatism and normative gender constrain variation in inland California. *University of Pennsylvania Working Papers in Linguistics* 20(2).

Podesva, Robert J. & Janneke Van Hofwegen. 2016. s/exuality in smalltown California: Gender normativity and the acoustic realization of /s. In Erez Levon & Ronald Beline Mendes (eds.), *Language, sexuality, and power*, 168–188. Oxford: Oxford University Press.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Seyfarth Scott, Kyle Gorman, Hilary Prichard & Jiahong Yuan. 2015. *FAVE (Forced Alignment and vowel extraction), version 1.1.3*. ZENODO.

Shadle, Christine H. 2012. On the acoustics and aerodynamics of fricatives. In Khalil Iskarous, Lisa Davidson, Helen M. Hanson & Christine H. Shadle (eds.), *The Oxford handbook of laboratory phonology*, 511–526. Oxford: Oxford University Press.

Shadle, Christine H. & Sheila J. Mair. 1996. Quantifying spectral characteristics of fricatives. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, 1521–1524.

Smyth, Ron & Henry Rogers. 2002. Phonetics, gender, and sexual orientation. In Sophie Burelle & Stanca Somesfalean (eds.), *Proceedings of the annual meeting of the Canadian Linguistics Association*, 299–301. https://cla-acl.ca/pdfs/actes-2002/Smyth_Rogers_2002.pdf (accessed 14 December 2021).

Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer S. Cole & Jose Ignacio Hualde (eds.), *Laboratory phonology*, Vol. 9, 65–86. New York: Mouton de Gruyter.

Stuart-Smith, Jane, Claire Timmins & Alan Wrench. 2003. Sex and gender in /s/ in Glaswegian. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1851–1854. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_1851.html (accessed 14 December 2021).

Svec, Jan G. & Svante Granqvist. 2010. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology* 19. 356–368.

Titze, Ingo R. & William S. Winholtz. 1993. Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research* 36. 1177–1190.

Tomiak, Gail R. 1990. *An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents*. Buffalo: State University of New York at Buffalo PhD dissertation.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/lingvan-2021-0014).