2017

The Acoustic Correlates of Fricatives in Whispered Speech: An Idiolectal Analysis

Joshua Wallin
St. Cloud State University, ihallesnailo@gmail.com

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

THE ACOUSTIC CORRELATES OF FRICATIVES IN WHISPERED SPEECH: AN IDIOLECTAL ANALYSIS

JOSHUA WALLIN AND ETTIEN KOFFI

ABSTRACT

The purpose of this paper is to provide examples of phonetic transcriptions using acoustical data and analyze the pronunciation of fricatives in running speech. Using Praat—a phonetic analysis program—and voice recordings, a (slightly modified) standardized script was accurately transcribed into IPA, the International Phonetic Alphabet. Using the same data, an in-depth analysis on fricatives of a native speaker of General American English (GAE) showed the speaker whispered during the recordings. This is supported by numerical data of several acoustic correlates, the primary correlates being intensity and voicing. The intensity of all fricatives was lower than a standard GAE speaker by an average of 14 dB and normalized within 4 dB. On average, all fricatives were produced as voiceless or devoiced.

1.0 Introduction

This paper will be a combination of two separate, but related, topics of acoustic phonetics: a narrow International Phonetic Alphabet (IPA) transcription of a modified George Mason University (GMU) phonetic script and an in depth analysis of fricatives, both based on my pronunciation of English vowels and consonants in running speech. The first section will display and discuss the modified GMU script, which was changed to add segments of English which were missing from the original text. There will be two transcriptions: one conducted with knowledge of standard pronunciation but no phonetic data, the other conducted with phonetic data. This is done to show phonetics is a science which requires experimentation and data to be accurate. The second section will describe fricatives in General American English (GAE) as it is pronounced by a native Minnesotan speaker, myself. The focus will be on the devoicing of fricatives in both the onset and coda.

2.0 IPA Transcription

The script used throughout our course and what all of the acoustic data is based upon is a modified version of the GMU phonetic script. The following is the modified script written in parse:

Please call Stella. Ask her to bring these things with her from the store: Six good spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a foot long sandwich as a snack for her brother Bob. We also need a small plastic snake, a yellow book, a rubber duck, a paper I-pad, the dog video game, a big toy frog for the kids, but not the faked gun. She can scoop these things into three red bags, and two old backpacks, and we will go meet her, Jake, and Jenny Wednesday at the very last train station at the edge of the zoo near York's Treasure Bank.

My first attempt of transcribing the passage without the benefits of acoustic analysis is as follows:

[pʰliːz kʰal stɛlə. æsk hɹ tʰə bɹĩŋ ðiːs θĩŋs wɪθ hɹ fɹɑ̃m ðə stoɹ, sɪks gʊd spũːns əf fɹɛʃ sno pʰiːs, faif θɪk slæbs əv blu t͡ʃiːs, æ̃n meːbi ə fʊtʰ lãŋ sæ̃ndwɪt͡ʃ æs ə snæk foɹ hɹ brəðɹ baːb. Wi ɔlso niːd ə smɔl plæstɪk snek, ə jɛlo bʊk, ə ɹəbɹ dək, ə pʰepɹ aɪpʰæd, ðə daːg̃ vɪdio gẽm, ə bɪg̃ tʰoi fɹag̃ fɹ ðə kɪds, bət nat ðə fek gɑ̃n. ʃi kn̩ skʊp ðis θĩŋs ĩntʰə θɹi ɹɛd bɛgs æ̃n tʰu old bækpʰæks, æ̃n wi wɪl go miːt hɹ, d͡ʒek, æ̃n d͡ʒẽni, wɛ̃nsde ætˉ ðə veɹi læs t͡ʃɹẽn steʃn̩ ætˉ ði ɛd͡ʒ əv ðə zuː niɹ joɹks t͡ʃɹɛʒɹ bẽnk.]

This, however, is not how I actually pronounced these sounds. Using acoustic data collected for my pronunciation of consonants and vowels based on this script, it would be more appropriately transcribed as:

[pʰliːz̥ kʰal stɛlʌ. ˀæsk hɹ tʰʌ b̥ɹĩŋ θiːz̥ θĩŋz̥ wɪθ hɹ fɹɑ̃m θʌ stoɹ, sɪks gʊd spũːnz̥ ˀʌɣ fɹɛʃ no pʰiːz̥, faɪɣ θɪk slæb̥z̥ ˀʌɣ blu t͡ʃiːz̥, ˀæ̃n meːbi ˀʌ fʊt lãŋ sæ̃ndwɪt͡ʃ ˀæz̥ ˀʌ snæk foɹ hɹ b̥ɹʌθɹ b̥aːb̥. wi ɔlso niːd ˀʌ smɔl plæstɪk snek, ʌ jɛlo b̥ʊk, ˀʌ ɹʌbɹ dʌk, ˀʌ pʰepɹ aɪpʰæd, θʌ daːg̃ ɣɪdio gẽm, ˀʌ b̥ɪg̃ tʰoɪ fɹag̃ fɹ θʌ kɪdz̥, b̥ʌt nat θʌ fek gʌ̃n. ʃi kn̩ skʊp̚ θiz̥ θĩŋz̥ ĩntʰʌ θɹi ɹɛd b̥ɛgz̥ ˀæ̃n tʰu old b̥ækpʰæks, ˀæ̃n wi wɪl go miːt hɹ, d͡ʒek, ˀæ̃n d͡ʒẽni, wɛ̃nz̥de ˀætˉ θʌ ɣɛɹi læs t͡ʃɹẽn steʃn̩ ˀætˉ θi ˀɛd͡ʒ ˀʌɣ θʌ zuː niɹ joɹks t͡ʃɹɛʒ̥ɹ b̥ẽnk.]

The main emphasis here in changes is the devoicing of many consonants. All of my fricatives are devoiced, except word-initial [z]. I will be going over this in detail later during my discussion of fricatives. Also, I will note all of my [b] sounds are devoiced as well, except for intervocalically. Some other changes include the glottal stops preceding [-low] vowels in the word-initial position. The exception to this is [æ], which also has a glottal stop in word-initial positions. There were also a few anomalies which were not quite consistent with the bulk of my data, such as the deletion of [s] and [t] in a few pronunciations and a single pronunciation of a voiceless [ɹ]. I also opted for the use of [ʌ] instead of [ə] to better represent the goals of my transcription.

To gather acoustical data, I used a program called Praat, which is commonly used in the field of acoustic phonetic analysis. A sample spectrogram and annotations are as follows:
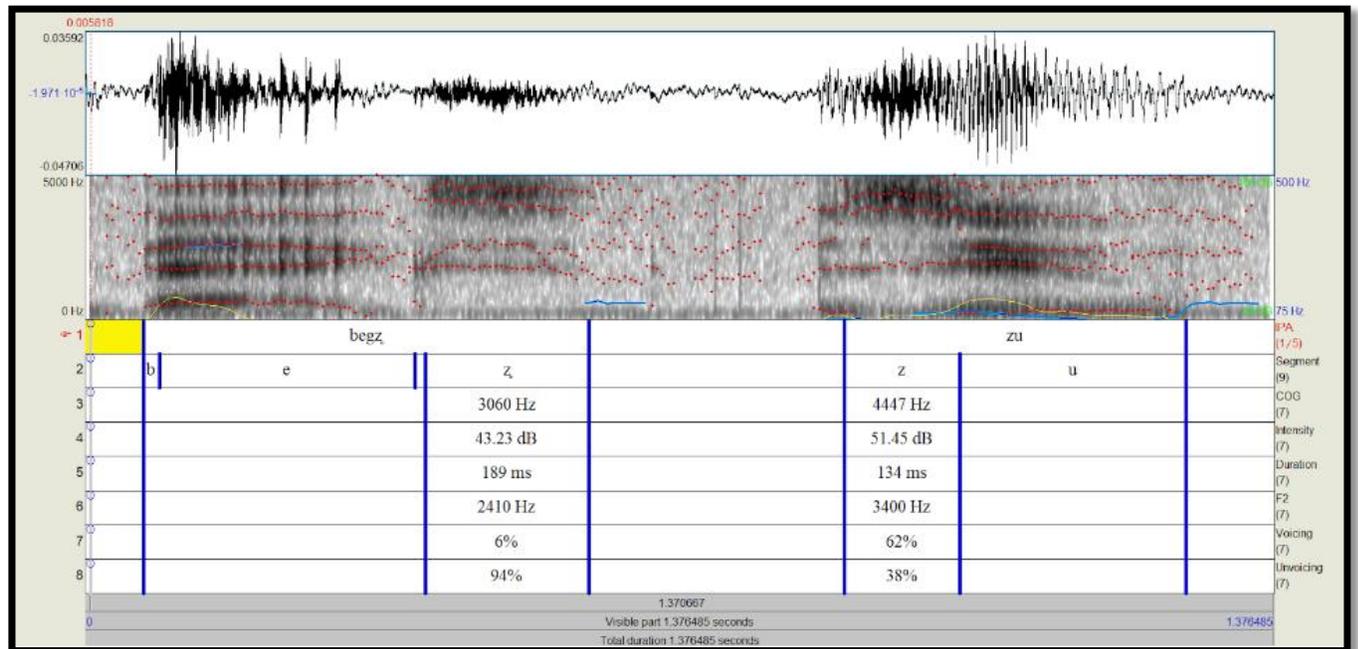
Figure 1: Spectrogram

This is an excellent example of the data which can be collected through Praat. The top is a waveform with a spectrograph below it. Beneath that are the annotations, which are labeled on the right-hand side. This same methodology was used for all of the data required for this analysis. The data displayed later on will be averages of the collected data, rather than cluttering this paper with numerous spectrograms.

3.0 Fricatives

Fricatives are defined as consonants which get their acoustic energy from air being forced through small gaps. Because of this nature, instead of resonating like other sounds, fricatives produce high pitched frequencies. This can be seen on spectrograms by their center of gravity (COG), which is focused around 3000 Hz – 6000 Hz. Depending on where the most concentrated clusters of energy are, the specific fricative being uttered can usually be determined. Whether the vocal folds are vibrating or not determines whether the fricatives are voiced or unvoiced. This will be used to analyze the devoicing of fricatives later on. First, each separate fricative in GAE must be discussed.

The first two fricatives to be discussed are quite similar in production and can easily be confused. They are [f] and [θ]. The first of the two is [f], which is a voiceless labio-dental fricative. This means it is produced by making contact with the upper teeth using the lower lip. The vocal folds do not vibrate during production. The same is for [θ], except it is produced inter-dentally. So the only difference between the two fricatives is whether the lower lip or the tongue is creating the gap with the teeth. Similarly, both of their acoustic energies are concentrated around 3000 Hz – 4000 Hz [note: Jongman et. al. (2000) states this concentration is around 5100 Hz for both, but I will use Ladefoged (2012) instead, because it is more recent]. Now, Ladefoged (2012:56) says [θ] is sometimes concentrated around 8000 Hz, but "There is often very little difference in the fricative noises of these two sounds – neither of them is very loud." This is supported by the data collected

by Jongman et al. (2000) where the intensities of [f] and [θ] are 55.7 dB and 54.7 dB, respectively. All other fricatives are above 60 dB, if not more. I will continue the discussion of intensity a bit later. The voiced counterparts of [f] and [θ] are [v] and [ð], respectively. They are almost identical to their voiceless counterparts except they are louder—which is sensible because of the vibrations of the vocal folds—and are shorter, which is similar with many voiced consonants. Altogether, this group of four fricatives make up the natural class of non-sibilant. Non-sibilants are characterized by a lower intensity and COG.

The other four fricatives are of the natural class sibilant. These consonants are [s] and [ʃ] and their voiced counterparts [z] and [ʒ], respectively. The first of this class to be discussed is [s], which is a voiceless alveolar fricative. This means it is produced by the tongue making a gap with the area just behind the top row of teeth. The other consonant pair, [ʃ] and [ʒ], are produced in the palatal region, the area just behind the alveolar region. The main difference between these two pairs of sounds is [s] and [z] have a lower intensity, but a higher COG. The COG of [s] and [z] is concentrated around 6000 Hz and extends above that, whereas [ʃ] and [ʒ] are concentrated around 3000 Hz – 4000 Hz. While this may cause people to think [ʃ] and [ʒ] could get confused with [f] and [θ], they would be greatly mistaken. In Miller and Nicely's (1955) confusion study, not once was [ʃ] and [ʒ] perceived as [f] and [θ] no matter if the talker or listener was a native speaker of GAE or not. This is because the intensities of sibilants compared to non-sibilants is drastic.

This brings us to the next topic in relation to fricatives; intensity is the most relevant correlate when discussing fricatives. Thomas (2011) and Ladefoged and Maddieson (1996) concur that, while other factors such as duration, COG, and F2 may be relevant, intensity is usually most reliable correlate. Fricatives are tricky, however, in that not one correlate can guarantee a differentiation, but a combination of all relevant correlates are required. Ladefoged and Maddieson (1996:139) express this succinctly by saying:

> ...Our best guess is that what matters for fricatives (more specifically for sibilant fricatives) is the overall   intensity, the frequency of lower cut-off point in the spectrum, and something corresponding to the   center of gravity and dispersion of spectral components above a certain threshold.

But, before moving on to the analysis of my pronunciation of fricatives, some caveats must be discussed. For the fricative [θ], a common pronunciation is [θ̪] which means it is produced dentally as a stop instead of inter-dentally as a fricative. While I did not see this in my pronunciation, it is commonplace enough to be mentioned. Also, the affricates [tʃ] and [dʒ] are similar to a stop combined with a fricative. For my pronunciation, they behave identically to the fricatives represented in their phonetic transcription. For this reason, I will not mention them specifically and, rather, will ask the reader to assume they behave in this manner for my pronunciation.

3.1 Pronunciation of Fricatives
The most surprising realization during the transcription of the modified GMU text was the pronunciation of almost all of my fricatives was either voiceless or devoiced. The only exception to this is word-initial [z], which

was mostly voiced, and sometimes word-initial [v]. To get a better grasp of this, below is a table of the averages for the pronunciation of all of my fricatives [note: I included the affricates in this table to support my claim in the previous section].

Consonant	f	v	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ
COG (Hz)	1362	1983	1126	590	5137	4015	4738	3094	4576	3010
Intensity (dB)	48	47	50	50	51	47	49	49	49	49
Duration (ms)	73	81	53	29	165	179	132	86	155	93
F2 (Hz)	2134	2062	1925	2124	2689	2934	2620	2321	2579	2626
Voicing (%)	13	38	19	0	12	25	0	11	4	4
Unvoicing (%)	87	62	81	100	88	75	100	89	96	96

Table 1: Fricatives

As we look at this table, we can see the voicing of word-initial [z] and [v] by the higher percentage of voicing compared to the other fricatives. In this case [v] is higher than [z] because there was less data, making a high voicing average, when, in actuality, [z] in word-initial positions was voiced 100% of the time. The same cannot be said for word-initial [v].

Because intensity is seen as the most relevant correlate for fricatives, it will be the first correlate discussed. After stating this, however, the highest difference in intensities is 4 dB between the voiceless sibilant [s] with 51 dB and the voiced non-sibilant [v] and voiced sibilant [z] with 47 dB. The threshold of perception for intensity in everyday listening conditions is 3 dB, which means intensity is not a very reliable correlate for my pronunciation. Another interesting point is the voiced fricatives are either equal or lower in intensity when compared to the voiceless fricatives. This is counter-intuitive and goes against Jongman's et al. (2000) findings, which state the voiced fricatives have higher intensities than their voiceless counterparts. Seeing as intensity will not be reliable for my pronunciation, we will move onto COG.

COG is considered a determining correlate for the differentiation of fricatives, but, again, it doesn't appear to be the case for my pronunciation. The non-sibilants are all much lower than what would be expected to be accepted values of COG. Also, because the COG describes the concentration of acoustic energy, which is based off of place of articulation, it is assumed the voiceless and voiced pairs would have similar COGs. This is not the case. For the lower values of COG, a threshold of around 200 Hz is enough to be perceivable, and it scales up to 400 Hz for mid-ranges and 600 Hz for high ranges. This means all voiced and voiceless pairs are perceptibly different. The difference of [f] (1362 Hz) and [v] (1983 Hz) is 621 Hz. For [θ] (1126 Hz) and [ð] (590 Hz) it is 536 Hz. For [s] (5137 Hz) and [z] (4015 Hz) it is 1122 Hz. Lastly, for [ʃ] (4738 Hz) and [ʒ] (3094 Hz) it is 1644 Hz. All of the differences are above the threshold of perception, which shows COG is probably not accurate for differentiating my fricatives. So we'll move onto duration.

Duration is also seen as a determining correlate for the discussion of fricatives. Again, this doesn't appear to be the case for my pronunciation. It is almost effective at distinguishing sibilants from non-sibilants, in that non-sibilants are shorter than 100 ms and sibilants are longer than 100 ms, but [ʒ] breaks this pattern with a duration of 86 ms. It is also poor at distinguishing voiced from voiceless fricatives in that for [f] and [v] and [s] and [z] the voiced consonants are longer, whereas for [θ] and [ð] and [ʃ] and [ʒ] the voiceless consonant are longer. Which leaves the last correlate, F2.

The second formant, F2, seems to be the most reliable correlate in terms of my pronunciation of fricatives. With a threshold of perceptibility of 200 Hz, F2 can somewhat distinguish sibilants from non-sibilants, but fails to differentiate voiced from voiceless consonants. The difference of [f] (2134 Hz) and [v] (2062 Hz) is 72 Hz. For [θ] (1925 Hz) and [ð] (2124 Hz) it is 199 Hz. For [s] (2689 Hz) and [z] (2934 Hz) it is 245 Hz. Lastly, for [ʃ] (2620 Hz) and [ʒ] (2321 Hz) it's 299 Hz. The highest value for non-sibilants is [f] with 2134 Hz and the lowest

for sibilants is [ʒ] with 2321 Hz. This gives a difference of 187 Hz, which is just under the threshold of perceptibility. Otherwise, F2 is successful at differentiating sibilants from non-sibilants. But there are still questions that come about from this data: why are almost all pronunciations of my fricatives voiceless and why is there no clear distinguishing correlate?

3.2 Analysis of Fricatives
First, we will attempt to answer why the majority of my pronunciations of fricatives are voiceless. The most likely answer is biology. Herschensohn and Young-Scholten (2013) describe the nature of fricatives by saying, "...the antagonistic articulatory requirements of voicing and frication make voiced fricatives inherently difficult to produce..." They say this because voicing requires adduction, or closing, of the glottis, whereas fricatives require sufficient airflow—and, thus, an open glottis—to be produced effectively. So, a main reason for the devoicing of fricatives in both the onset and coda would be a subconscious decision to value frication over voicing. But that still leaves word-initial [z] unaccounted for. The most reasonable explanation is the word-initial [z] wasn't word-initial at all, but, rather, intervocalic. In context and running speech, the [z] was realized as [θʌzu]. To get a better idea of what happened, a syllable diagram will give us some insight.
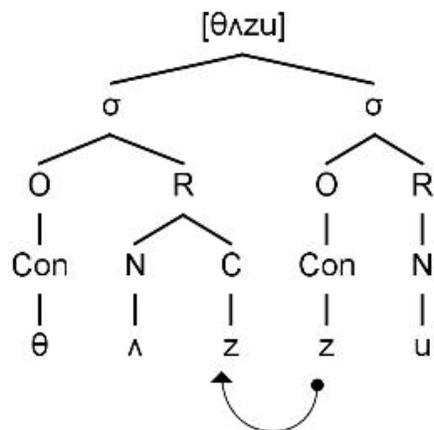

Figure 2: Syllable Diagram

In English, there is a rule that prevents lax vowels from ending a syllable. To prevent this, the [z] is ambisyllabic, that is, shared as both a coda and onset, even though they are part of separate words. This would explain this anomaly. But it still doesn't explain why there isn't a definite correlate that describes the nature of my fricatives.

For almost all of my acoustic analyses, I noticed inconsistencies with my pronunciations and that of a standard GAE speaker, the most noticeable correlate being intensity. This leads to my hypothesis: I whispered during the recording of my phonetic script. This is supported by Jovicic (2008) when he states an average lowering of consonant intensity in whispered speech by 12 dB and a normalization of consonants varying no more than 3.5 dB. This is completely consistent with my data with an average lowering [compared to Jongman et al. (2000)] of 14 dB and a greatest variance of 4 dB. This would explain difficulties in proper analyses and overall trouble identifying correlates of fricatives. This would also explain why other consonants were realized as voiceless while leaving other correlates undisturbed, like voice onset time (VOT) for stops. Jovicic (2008) also states VOT and affrication duration for unvoiced consonants in whispered speech is similar to that of normal speech, but the voiced consonants are heavily affected. This could also be another possible explanation for the devoicing of so many consonants, especially fricatives.

4.0 Summary

Overall, what can be taken away from this analysis is the scientific benefits of acoustic phonetics and its ability to accurately describe and provide explanations for human speech. Using phonetic data, I was able to transcribe my pronunciation of consonants and vowels more accurately than just through my knowledge of linguistics and standard pronunciations of GAE. I discovered I have glottal stops before [æ] and [-low] vowels. I also discovered my phonetic data is more accurately described by whispered speech rather than normal running speech. This explains many abnormalities not expected from a native GAE speaker. If my hypothesis is incorrect, then I learned I devoice all of my fricatives; however, I do not believe this to be the case. Based on the lowered and normalized intensities, my recordings are akin to whispered speech, and more recordings with actual running speech should be conducted for an accurate representation of my pronunciations.

ABOUT THE AUTHORS

Joshua Wallin is an undergraduate student with an intended major of Physics Education. He has lived most of his life in Minnesota but has traveled around the United States and locations abroad, such as Egypt and Japan. He plans to teach physics or astrophysics at the high school or university level. His love for physics guided him into the field of acoustic phonetics. He can be reached via his school e-mail at: jawallin@stcloudstate.edu or his personal e-mail at: ihallesnailo@gmail.com

Ettien Koffi, Ph.D. in linguistics from Indiana University, teaches linguistics at Saint Cloud State University, MN. Author of many peer-reviewed articles on various topics in linguistics and of four books: *Language Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and Policy: Game Theoretic Solutions* (2012), *Applied English Syntax* (2010, 2015), and the *New Testament in Anyi Morofu* (2017), a task which took over 25 years. Specializing in acoustic phonetics, dialect variation, and emergent orthographies, his current research centers on speech acoustics of L2 English (within the Speech Intelligibility Framework), Central Minnesota English, and Anyi. He can be reached at enkoffi@stcloudstate.edu.

References

Ladefoged, Peter, and Disner, Sandra F. 2012. *Vowels and Consonants.*

Koffi, Ettien. 2016. *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility.* St. Cloud, MN: Unpublished Coursebook.

Herschensohn, Julia, and Young-Scholten, Martha. 2013. *The Cambridge Handbook of Second Language Acquisition.*

Jovicic, Slobodan T. 2008. Acoustic Analysis of Consonants in Whispered Speech. *Journal of Voice* Volume 22. Issue 3. Pages 263-274.