Linguistic Portfolios

Volume 6 Article 5

2017

The Acoustic Correlates of the Voiceless Palatal Fricative [ʃ] in Central Minnesota English

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

Maria Bloch
Hofstra University, mbloch1@pride.hofstra.edu

THE ACOUSTIC CORRELATES OF THE VOICELESS PALATAL FRICATIVE [ʃ] IN CENTRAL MINNESOTA ENGLISH

ETTIEN KOFFI AND MARIA BLOCH[1]

ABSTRACT

Fricatives are found in all world languages (Maddieson 1984:41). Similarities in mouth geometry and in their aerodynamic characteristics have made them the favorite segments for testing claims about universals of phonetic features. Yet, there is important interspeaker variability even within the same language (Ladefoged and Maddieson 1996:139). In this paper we investigate the voiceless sibilant fricative [ʃ] produced by Central Minnesota English (CMNE) speakers to see how similar and yet different it is from those produced by speakers of other dialects of General American English (GAE). We extend the comparisons to sibilants in some non-Western languages. The acoustic correlates investigated are Center of Gravity (CoG), F2, duration, and intensity. The findings reported in this paper are based on 5,544 *measured tokens. Our results can be of interest to speech intelligibility researchers, forensic acousticians, sociophoneticians, and linguists in general, who are interested in sibilants.*

1.0 Introduction

Maddieson (1984:41) reports based on the *UCLA Phonological Segment Database* (UPSID) that fricatives are found in all world languages. Naturally, some languages have more fricatives than others. He notes on page 43 that some languages have 12 fricatives or more. It is commonly reported that English has nine fricatives [s, z, ʃ, ʒ, θ, ð, f, v, h]. In this study we focus only on the fricative [ʃ] as produced by the speakers or Central Minnesota English (CMNE). We take a variationist sociophonetic approach in this paper for three reasons. First, we want to test the claim in Ladefoged and Maddieson (1996:139) that "the acoustic structure of fricatives seems to vary widely from individual to individual." Second, we want to determine which acoustic correlate(s) is/are robust for the perception of [ʃ]. This inquiry is important because it can contribute much needed insights into the acoustics of fricatives, which is still in its infancy. Ladefoged and Maddieston (1996:173) lament the lack of serious research on fricatives, noting that "There have been surprisingly few studies of the acoustics of fricatives." The third goal is related to the first two, namely, to determine whether or not [ʃ] can be used in forensic acoustics. The Critical Bands Theory (CBT) is the framework that we rely on to interpret the acoustic measurements of [ʃ]. The acoustic correlates investigated are Center of Gravity (CoG), F2, Duration, and Intensity. The paper is subdivided into two main sections. The first provides an overview of the articulatory characteristics of [ʃ], a brief literature review, some background information about the participants, the data recording procedures, and the methodology. The second installment focuses on the various acoustic analyses and the interpretations of the various measurements.

---

[1] The first author assigned this topic to the second author as a capstone project for the fulfillment of the requirements of the BA in linguistics. They met regularly over the course of a semester to discuss reading assignments and/or to verify the accuracy of the acoustic measurements. She is recognized as the second to the extent that she collected the original data and did all the acoustic measurements.

2.0 An Articulatory Overview of English Sibilants

English has two palatal fricatives, the voiceless [ʃ] and the voiced [ʒ]. We focus on [ʃ] for the obvious reason that it occurs more frequently in English than [ʒ]. It also has a wider distribution than [ʒ]. The latter is confined for the most part to word-medial positions, and to loanwords from French. The place of articulation of [ʃ] in English is hotly debated (Ladefoged and Maddieson 1996:148). Some researchers describe it as palato-alveolar, others as palatal, and still others simply as post-alveolar. Fromkin et al. (2014:210, 242) ascribe the phonetic feature [-anterior] to [ʃ] because it is produced behind the alveolar region. Lip rounding is a secondary articulation commonly associated with [ʃ] in English. Figure 1 depicts the place of articulation of [ʃ].

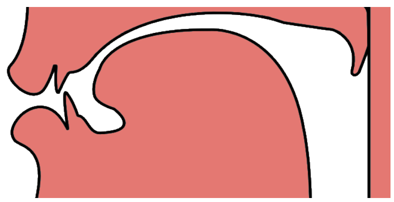

Figure 1: Place of articulation of [ʃ]

Aerodynamically, the production of [ʃ] requires that air molecules flow over a relatively large surface, from behind the alveolar ridge right to the edge of the velar area. As a result, the energy involved in the production of [ʃ] is diffuse. If we accept the proposition that [ʃ] is a palatal fricative as does Fromkin et al. (2014:204), then according to Zsiga (2013:140), the formants that correlate to its place of articulation are F2, F3, and F4. We will investigate F2 shortly, but we will not measure F3 and F4 in this paper for the following reasons. F3 correlates with lip rounding/protrusion. Much of this information is subsumed under F2. Therefore, measuring F3 would be redundant, and even pointless. As for F4, it is hardly ever used to measure fricatives directly. Instead, most phoneticians turn to Center of Gravity (CoG), as is explained in 5.0.

3.0 Succint Literature Review

This paper is inspired by the research designs in Jongman et al. (2000), Gordon et al. (2000), and Haley et al. (2010). The first two investigate fricatives comprehensively. The latter focuses narrowly on [ʃ] as we do in this paper. Jongman et al. (2000) studied eight fricatives [f, v, θ, ð, s, z, ʃ, ʒ] produced by 20 students from Cornell University: 10 males and 10 females. They investigated these segments in six vowel contexts (p. 1255). The test items in which the fricative occurred were repeated three times (p. 1255). However, their article does not specify what those words were. We only know that their corpus consisted of 144 tokens (8 x 6 x 3) for a total of 2,880 measured tokens in syllable onsets. The acoustic correlates used are CoG, F2, Duration, and Intensity. Their mean measurements are displayed in Table 1:

[ʃ]	CoG	F2	Duration	Intensity
Males + Females	4229 Hz	1982 Hz	178 ms	66 dB
Males (Only)	NA	1849 Hz	NA	NA
Females (Only)	NA	2115 Hz	NA	NA

Table 1: Summary of Relevant Correlates

The acronym "NA" stands for "Not Available," which means that for the most part the authors did not differentiate between male and female talkers for the correlates under consideration, except for F2. Gordon et al. (2000) also investigated duration and CoG of fricatives in seven native American languages. Their study is significant for investigating the acoustic correlates of [ʃ] because their languages lie outside of the Indo-European family of languages. Their measurements are displayed in Table 2:

Languages	**Duration**	**CoG**
Chickasaw	112 ms	4679 Hz
Western Apache	175 ms	4859 Hz
Gaelic	110 ms	4396 Hz
Western Aleut	NA	NA[2]
Montana Salish	178 ms	4134 Hz
Hupa	217 ms	4440 Hz
Toda	239 ms	4704 Hz
Mean for all Languages	**171 ms**	**4,535 Hz**

Table 2: Measurements of [ʃ] in Seven Languages

To the best of our knowledge, the only study that deals exclusively with the acoustics of [ʃ] is Haley et al.'s (2010). They studied it by examining its behavior in five vowel contexts and by contrasting male and female speech. Their list includes five words: <*sheet, shack, shawl, shoe, shun*> in which [ʃ] occurs in syllable onsets. The five vowel contexts before which [ʃ] occurred are [i, æ, ɔ, u, ʌ]. Their study included 10 participants: 5 males and 5 females, all whom are native speakers of American English. Five of them were from the Southeast, three from the Midwest, one from New York, and one from Florida. They were referred to in the study simply as Male 1, 2, 3, 4, 5 and Female 1, 2, 3, 4, and 5.

Speakers	**CoG**	**St.Deviation**	**Speakers**	**CoG**	**St.Deviation**
Female 1	4,900 Hz	36	Male 1	4,000 Hz	30
Female 2	5,700 Hz	21	Male 2	4,100 Hz	35
Female 3	4,600 Hz	25	Male 3	3,900 Hz	33
Female 4	5,500 Hz	24	Male 4	4,500 Hz	36
Female 5	5,300 Hz	31	Male 5	4,600 Hz	44
Female Mean	**5,200 Hz**	47	**Male Mean**	**4,200 Hz**	44
Overall Mean (female and male combined) = 4,700 Hz, Standard deviation = 66					

Table 3: CoG Measurements of [ʃ] in Haley et al.

Haley et al.'s (2010) paper offers us an incomplete view of the acoustic characteristics of [ʃ] in American English because they did not investigate F2, Intensity, and Duration. Yet, the fact that they provide male and female measurements is important for our study because it provides us with a preliminary gender-based analysis of [ʃ].

In summary, the findings in the three studies mentioned above give us some rough estimates about the acoustic correlates of [ʃ]. We know from Haley et al.'s (2010) and Jongman

[2] There is no segment [ʃ] in this language.

et al. (2000) that the CoG of [ʃ] in American English is > 4,200 Hz.. We also know from them that its duration is > 170 ms. We learn from Jongman et al. (2000) that the F2 of [ʃ] is about 2,000 Hz, and that its intensity is > 60 dB. These are the baseline measurements against which we will compare the CoG, F2, duration, and intensity of the [ʃ]s produced by the CMNE speakers in our study.

4.0 Participants, Data Set, Equipment, and Methodology

Six participants, three males and three females from Central Minnesota, were recorded for this study. Each produced 22 words containing ([ʃ]). In 11 of them, [ʃ] appears in syllable onsets, and in 11 others, it is found in syllable codas. The participants signed a consent form approved by the IRB (Institutional Review Board) at St. Cloud State University. They were recorded on an Olympus (VN-6200PC) digital recorder. They wore a noise cancellation Logitech headset microphone (G230) during the recording sessions. The recordings took place in quiet surroundings, but not in sound treated rooms or anechoic chambers. The sound files from the recordings were converted to WAV files and transferred to Praat for acoustic phonetic measurements (Boersma and Weenink 2010). All the recordings were sampled at 44100 Hz at a rate of 16 bits per sample. The words recorded for the study are listed as follows:

N0	Initial [ʃ]	Transcription	Final [ʃ]	Transcription
1.	sheep	[ʃip]	leash	[liʃ]
2.	ship	[ʃɪp]	fish	[fɪʃ]
3.	shape	[ʃep]	taish	[teʃ]
4.	shed	[ʃɛd]	flesh	[flɛʃ]
5.	shack	[ʃæk]	cash	[kæʃ]
6.	shop	[ʃɑp]	Josh	[d͡ʒɑʃ]
7.	shawl	[ʃɔl]	rawsh	[rɑʃ]
8.	showed	[ʃod]	skoash	[skoʃ]
9.	should	[ʃʊd]	push	[pʊʃ]
10.	shoot	[ʃut]	whoosh	[wuʃ]
11.	shut	[ʃʌt]	brushf	[brʌʃ]

Table 4: Data Set

The words <*taish, rawsh, skoash*> are nonsense words. They are included in the corpus in order to have a balance of 22 words. In the original project, seven acoustic correlates were investigated: F0/Pitch, F1, F2, F3, intensity, duration, Center of Gravity (CoG). However, due to space limitation, we will focus only on four of them, namely, CoG, F2, Intensity, and Duration.

Spectrographs were produced on the basis of the speech samples obtained from the participants. In analyzing the spectrograms, the boundaries were drawn to separate the different sound segments in each word. Each word was transcribed phonetically in accordance with the International Phonetics Alphabet guidelines, (IPA 1999:27-33). The focus was placed primarily on the prevocalic and postvocalic occurrences of [ʃ]. Boundaries were created around [ʃ] and relevant measurements were collected, as shown in Figure 2:

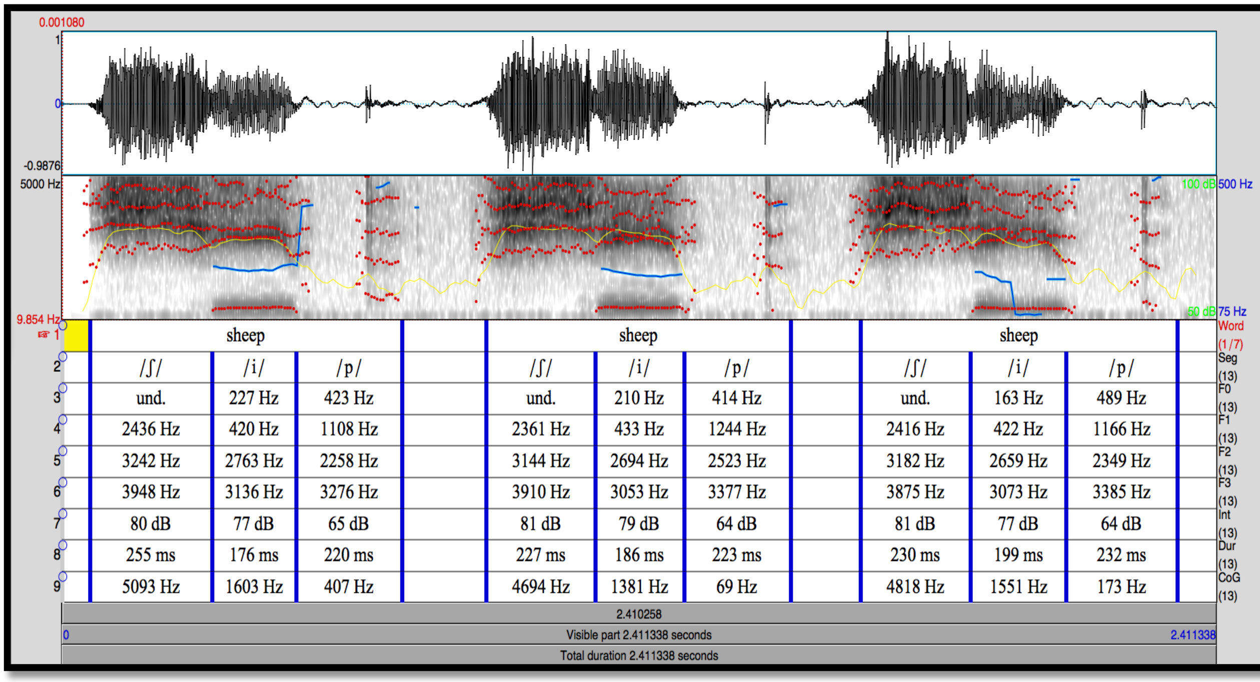

Figure 2: Annotate Spectrograph of <sheep> by Female 1

Figure 2 is a prototypical spectrograph. The total corpus consists of 22 such spectrographs (11 x 2). The findings reported in this study are based on 924 tokens (11 x 2 x 6 x7) and 5,544 measured tokens.[3] Praat Version 5.13.16 was used to make a total of 132 separate annotated spectrographs.

4.1 The Critical Band Theory

Results from spectrographs such as the one in Figure 2 provide an enormous amount of data that can be interpreted differently depending on the issues under consideration. However, as Ladefoged and Maddieson (1996:139) advised, we must focus on "what is linguistically and *perceptually* most relevant in the acoustics of fricatives" [Italics added for emphasis]. In so doing, we we turn to the Critical Band Theory (CBT). In 1940, physicist Harvey Fletcher postulated theoretically and demonstrated experimentally that different portions of the basilar membrane specialize in perceiving different frequencies. Another physicist, von Békésy (1947), spent most of his academic life proving clinically that that was indeed the case. His tireless efforts earned him a Nobel Prize in Medicine in 1961. Now, it is widely accepted that the human auditory spectrum is divided into frequency bands, ranging from 20 to 20,000 Hz. However, the linguistically relevant frequencies that are encoded in speech signals range from 60 to 5,000 Hz. The literature on this topic is highly technical and even esoteric for the average linguist. Koffi (2016:115-134) has provided an accessible summary of CBT in the perception of vowels. Suffice it to say here that in speech acoustics, there are five formants that are perceptually and linguistically relevant. Fundamental frequency, abbreviated as F0, ranges from 60 to 500 Hz. It encodes pitch information. The first formant or F1, ranges from about 300 Hz to 1,000 Hz and provides information about mouth aperture. The second formant or F2 correlates with tongue movements: front, central, retraction. It is in the 2,000 Hz frequency range. The third formant or F3 deals with lip movements: lip rounding, lip protrusion, and lip spreading. It is in the 3,000 Hz frequency

[3] Due to limitation of space, we display only the spectrograph in Figure 2.

range. The fourth formant or F4 has to do with the state of the glottis during phonation. It is in the 4,000 Hz frequency range. Other formants exists, but phoneticians focus mostly on these four in describing speech. According to Zsiga (2013:140), F2, F3, and F4 are all pertinent to the study of the palatal fricative [ʃ].

CBT has allowed phoneticians to determine with precision the frequency bands that are significant in speech perception. Within each frequency band, there are specific thresholds at which speech signals are or are not salient. They are given various names: perception limen, Just Noticeable Difference (JND), or reference level. These terms are all synonyms and describe the same acoustic reality. We prefer JND because it is short. As a rule of thumb, the ear can detect a frequency change of 1 Hz between two speech signals on the F0 frequency band. On the F1 frequency band, the JND is 60 Hz. On the F2 frequency band, it is 200 Hz; on the F3, it is 400 Hz, and on the F4, it is 630 Hz.[4] These are JNDs that are linguistically or perceptually salient (Everest and Pohlmann 2015:13-15). As noted earlier, the frequencies that are of interest to us are those inherent in CoG and F2. Additionally, we will investigate the durational and intensity characteristics of [ʃ].

5.0 The Relevance of the CoG Measurement

F4 is in the 4,000 Hz range. Though it plays a role in determining the place of articulation of palatals, it hardly ever figures directly in the description of fricatives. Instead, phoneticians rely on CoG (Ladefoged and Maddieson 1996:139). CoG is a concept borrowed from physics. The *Encyclopedia Britannica* (2015) defines it as an "imaginary point in a body of matter where, for convenience in certain calculations, the total weight of the body may be thought to be concentrated." A simpler way of explaining CoG is to compare it to a violent storm. When a hurricane is approaching, meteorologists estimate its force and destructive power by measuring the energy in "the eye" of the hurricane. It is where the maximum power of the storm is concentrated. It is also from where the energy radiates outward. This analogy is useful in describing fricatives in general, and sibilant fricatives in particular. Fricatives have acoustic power. Their power is concentrated in the "eye" of the fricative. The force of the fricative is a function of where the eye is located. If it is located in the dental or alveolar area, the fricative is very powerful. If it is located between the lower lip and the upper teeth, or between the upper and lower teeth, it is less powerful. If it is located between the alveolar ridge and the velum, it is somewhat powerful. Acousticians have come up with mathematical formulas to pinpoint the CoG of fricatives. The formula in Praat is used to calculate the CoG of the [ʃ]s produced by the participants in our study, as displayed in Tables 5 and 6[5]:

[4] The JND of 630 Hz is a compromise between the F4 of males, which is at 600 Hz, and that of females, which is at 700 Hz. See Stevens (2000:154, 300) for additional information.

[5] We express our gratitude to the staff of the Statistical Consulting and Research Center at St. Cloud State University for their help with various statistical correlations.

CoG	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Male 1-Initial	4492	4438	3988	4389	4344	4080	4140	4312	4274	4316	3990	4251
Male 2-Initial	3346	3394	3718	3137	3562	2703	2788	2726	2977	3303	3219	3170
Male 3-Initial	4490	4452	4455	4135	4537	3955	4224	4006	4248	4382	4199	4280
Mean-Initial	**4109**	**4094**	**4053**	**3887**	**4147**	**3579**	**3717**	**3681**	**3833**	**4000**	**3802**	**3900**
Male 1-Final	3822	3654	4011	4061	3806	3855	3677	4146	3777	3601	3737	3831
Male 2-Final	3360	2772	3082	3022	2771	3221	2991	2822	3238	3184	3120	3053
Male 3-Final	4162	3705	3818	3903	3714	3985	3735	4176	3705	4097	3866	3896
Mean-Final	**3781**	**3377**	**3637**	**3662**	**3430**	**3687**	**3467**	**3714**	**3573**	**3627**	**3574**	**3593**
Overall Mean	**3945**	**3735**	**3845**	**3774**	**3789**	**3633**	**3592**	**3698**	**3703**	**3813**	**3688**	**3746**
St. Deviation	**476**	**585**	**412**	**512**	**571**	**501**	**537**	**659**	**478**	**475**	**393**	**480**

Table 5: CoG in Male Speech

CoG	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Female 1-Initial	4868	4155	4052	4308	4162	4050	3991	4331	4499	4411	4047	4261
Female 2-Initial	4511	4043	4043	3877	4204	3813	4099	4028	4135	4383	4074	4110
Female 3-Initial	4163	4126	4092	4223	4070	4021	4039	4310	4574	4467	4241	4211
Mean-Initial	**4514**	**4108**	**4062**	**4136**	**4145**	**3961**	**4043**	**4223**	**4402**	**4420**	**4120**	**4194**
Female 1-Final	4393	4121	4320	4034	4228	3980	3666	3774	3975	4252	3771	4046
Female 2-Final	4186	3956	4055	4010	3925	3940	3626	3856	3815	3756	3476	3872
Female 3-Final	3641	3598	3805	3553	3889	3688	4113	3652	3963	4345	3653	3809
Mean-Final	**4073**	**3891**	**4060**	**3865**	**4014**	**3869**	**3801**	**3760**	**3917**	**4117**	**3633**	**3909**
Overall Mean	**4293**	**3999**	**4061**	**4000**	**4079**	**3915**	**3922**	**3991**	**4160**	**4269**	**3877**	**4051**
St. deviation	**374**	**191**	**149**	**245**	**132**	**126**	**199**	**257**	**282**	**238**	**265**	**165**

Table 6: CoG in Female Speech

For two speech signals to be perceptually relevant, the CoG distance between them should be ≥ 630 Hz (Everest and Pohlmann 2015:13-4). In the remaining sections, we will examine the information in Tables 5 and 6 to see if CoG is perceptually salient in the pronunciation of [ʃ].

5.1 Variability in CoG by Vowel Contexts

The measurements indicate that the CoG of [ʃ] is higher in syllable onsets than in syllable codas. In male speech, the mean CoG score at the beginning of words is 3,900 Hz, and 3,746 Hz at the end of words. The difference of 154 Hz between the CoG of [ʃ] in both environments is not perceptually salient. We observe a similar pattern in female speech. In word-initial position, the mean CoG of [ʃ] is 4,194 versus 4,051 Hz at the end of words. There is a difference of 143 Hz in both positions. This difference is not perceptually salient. In male and female speech, the CoG of [ʃ] is higher before the vowels [i, ɪ, e, ɛ, æ] than the back vowels [u, ʊ, o, ɔ, ɑ]. The combined CoG score of [ʃ] for front vowels is 4,005 Hz (3,817 Hz for males and 4,193 for females, as opposed to 3,752 Hz for back vowels (3,613 Hz for males and 3,892 for females). Overall, as far as CoG is concerned, the vowel context in which [ʃ] occurs is not perceptually salient.

5.2 Speaker Variability in CoG

We will first compare the CoG of the six participants in our study with those in Jongman et al. (2000) and then compare them with those in Haley et al. (2010:550). The combined CoG of the 10 male and 10 female participants in Jongman et al. is 4,229 Hz. The one in our study is 3,901 Hz (3,746 for males and 4,051 for females). There is a difference of 328 Hz in GoG between the two studies. Since, the CoG difference is less than 630 Hz, we conclude that there is no perceptual difference between CMNE speakers and those in Jongman et al. Now, let's compare CMNE speakers with the participants in Haley et al. (2010). The CoG of the latter is 4,700 Hz.

Their aggregate CoG is 799 Hz higher than that of CMNE speakers (3,901 Hz). Since the difference is ≥ 630 Hz, we conclude that it is perceptually significant. In other words, one can rely on the CoG cue to differentiate between the participants in our study and those in Haley et al.

There are noteworthy individual differences in CoG between the three male speakers in our data. Male 1 (4,251 Hz) and Male 3 (4,280 Hz) produce their [ʃ] similarly. However, their CoG is substantially different from that of Male 2 (3,170 Hz). The mean CoG difference between the [ʃ] of the former and the latter is 1,095 Hz. This difference is both linguistically and perceptually significant because it is higher than the 630 Hz threshold. Our findings validate Ladefoged and Maddieson's (1996:139) observations that "the acoustic structure of fricatives seems to vary widely from individual to individual." It also confirms a second observation made on page 173 that "there are great discrepancies among the spectra of a given fricative as spoken by different speakers, but the differences among the spectra are consistent for a single speaker." We see that across the board, in all vowel contexts, Male 2 produces very low CoG for [ʃ]. Male 2 retracts this sound more than any of the participants in our study. He produces it somewhere between the palatal and velar areas. The CoG of [ʃ] is more homogeneous among our female talkers, as evidenced by the overall standard deviation of 165 Hz in their speech, compared with 480 Hz in male speech.

5.3 Gender-based Variations in CoG

Jongman et al. (2000) provide a combined CoG of 4,229 Hz for [ʃ]. They did not differentiate between males and females. However, Haley et al. (2010) did. They report that the CoG of [ʃ] in female speech is 5,200 Hz, compared with 4,200 Hz for male speech. The difference of 1,000 Hz is perceptually salient. In our data, the combined CoG for males in all positions is 3,747 Hz. In female speech, we have a combined CoG of 4,051 Hz in all positions. The gender-based difference in the combined positions is 304 Hz, which is not perceptually salient because it is below the 630 Hz threshold. In other words, CoG does not discriminate between the male and female talkers in our study, but it is a distinctive correlate in Haley et al.

6.0 The Relevance of the F2 Correlate

F2 correlates with the horizontal movement of the tongue and provides information as to whether a segment is fronted, centralized, or retracted. For vowels, a high F2 correlates with fronting, whereas a low F2 is an indication of retraction. For consonants, a high F2 corresponds to a [+anterior] pronunciation, whereas a low F2 value denotes a [-anterior] pronunciation. The [± anterior] features translate into the following acoustic measurement on the F2 frequency band. Generally speaking, segments whose F2 values are ≥ 2000 Hz are classified as [+front] or [+anterior], those whose values are between 1800 and 1400 Hz are considered [+central], and those whose F2 values are ≤ 1400 Hz are [+back] or [-anterior].[6] On the F2 frequency band, a JND difference of ≤ 200 Hz is not deemed to be perceptually salient. Tables 7 and 8 display the F2 measurements of [ʃ]:

[6] In binary systems such as [±anterior], segments whose F2 are lower than 1800 Hz are classified as [-anterior].

F2	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Male 1-Initial	2843	2823	2852	2909	2877	2775	2734	2712	2770	2727	2789	2801
Male 2-Initial	2642	2620	2620	2540	2543	2433	2370	2356	2405	2520	2583	2512
Male 3-Initial	2664	2609	2690	2602	2628	2433	2581	2510	2559	2545	2568	2580
Mean-Initial	**2716**	**2684**	**2720**	**2683**	**2682**	**2547**	**2561**	**2526**	**2578**	**2597**	**2646**	**2631**
Male 1-Final	2802	2720	2847	2826	2718	2653	2608	2538	2788	2487	2755	2703
Male 2-Final	2572	2423	2488	2412	2292	2467	2368	2325	2447	2405	2368	2415
Male 3-Final	2479	2568	2544	2553	2465	2422	2495	2476	2426	2433	2403	2478
Mean-Final	**2617**	**2570**	**2626**	**2597**	**2491**	**2514**	**2490**	**2446**	**2553**	**2441**	**2508**	**2532**
Overall Mean	**2667**	**2627**	**2673**	**2640**	**2587**	**2530**	**2526**	**2486**	**2565**	**2519**	**2577**	**2581**
St. Deviation	**342**	**174**	**129**	**222**	**118**	**112**	**183**	**251**	**273**	**220**	**260**	**145**

Table 7: F2 in Male Speech

F2	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	Mean
Female 1-Initial	3189	3055	3087	2981	2998	2929	2909	2993	2986	2979	3005	3010
Female 2-Initial	2860	2783	2791	2807	2829	2774	2723	2793	2733	2744	2701	2776
Female 3-Initial	2894	2768	2770	2741	2718	2612	2642	2674	2563	2668	2673	2702
Mean-Initial	**2981**	**2868**	**2882**	**2843**	**2848**	**2771**	**2758**	**2820**	**2760**	**2797**	**2793**	**2829**
Female 1-Final	2637	2628	2587	2619	2688	2660	2683	2726	2858	2823	2679	2689
Female 2-Final	2740	2818	2752	2679	2800	2787	2644	2709	2703	2733	2632	2727
Female 3-Final	2682	2767	2727	2638	2638	2523	2465	2422	2468	2482	2442	2568
Mean-Final	**2686**	**2737**	**2688**	**2645**	**2708**	**2656**	**2597**	**2619**	**2676**	**2679**	**2584**	**2661**
Overall Mean	**2833**	**2803**	**2785**	**2744**	**2778**	**2714**	**2677**	**2719**	**2718**	**2738**	**2688**	**2745**
St. Deviation	**200**	**139**	**164**	**135**	**128**	**144**	**143**	**184**	**188**	**164**	**181**	**146**

Table 8: F2 in Female Speech

Jongman et al. (2000:1259) report that the mean F2 of [ʃ] for their participants is 1982 Hz. The overall F2 for our participants is 2663 Hz. The difference between our participants and theirs is 681 Hz. This difference is perceptually salient. In other words, CMNE talkers produce [ʃ] towards the front of the mouth, whereas the participants in Jongman et al. pronounce it slightly in the back of the mouth. From the standpoint of articulatory phonetics, it can be said that CMNE speakers produce [ʃ] in the palatal area, whereas the participants in Jongman et al. produce theirs in the post-palatal area. This confirms what was said in 2.0 about the different ways in which speakers pronounce [ʃ].

6.1 Speaker Variability in F2

The data shows significant interspeaker variations in F2. Male 1, for example, fronts his [ʃ]s more forcefully than Male 2 or Male 3. The F2 difference between him (2801 Hz) and Male 2 (2512 Hz) and Male 3 (2580 Hz) is 255 Hz. This difference is perceptually salient. The same is true between Female 1 (3010 Hz) on the one hand, and Female 2 (2776 Hz) and Female 3 (2702 Hz) on the other. The F2 difference between her [ʃ] and that of the two other females is 271 Hz. This difference is also perceptually salient. In postvocalic positions, Male 1 still fronts his [ʃ] (2703 Hz) more strongly than Male 2 (2415 Hz) and Male 3 (2478 Hz). The F2 difference between him and the two other talkers is 256 Hz. The F2 difference among female talkers in postvocalic position is not salient. The F2 of [ʃ] underscores the speaker variability discussed in 5.2.

6.2 Gender-based Variations in F2

Jongman et al.'s (2000:1259) reported separate measurements for males and females for the F2 of [ʃ]. For all other fricatives, they did not differentiate between males and females. This suggests that they anticipated a gender-based variation in the production of [ʃ]. Their mean F2 value for the [ʃ] produced by males is 1,849 Hz, and 2,115 Hz for females. The difference of 266 Hz is perceptually salient. As noted earlier, Jongman et al.'s (2000) focused only on fricatives in syllable onsets. We will follow their lead and measure [ʃ] at the beginning of words to see if it reveals any gender difference among our participants. The combined F2 score for our males is 2,631 Hz, compared with 2,829 Hz for females. The difference between them is 198 Hz, that is, two 2 Hz shy of the limen of perception. We conclude, therefore, that in word-initial positions, male and female CMNE talkers produce [ʃ]s slightly differently. Female talkers front their [ʃ] more strongly than their male counterparts. At the end of word, the difference of 129 Hz between males (2,532 Hz) and females (2,661 Hz) is not perceptually salient. In other words, gender-based variations in the production of [ʃ] happen only at the beginning of words in CMNE. Our findings are not only in line with Jongman et al., but also in agreement with Stevens (2000:410).

7.0 The Relevance of the Duration Correlate

Segmental duration has been the subject of intense acoustic phonetic studies for several decades. In this section, we summarize the most relevant findings, and then compare our results with available durational data. One of the most quoted studies on the duration of English speech sounds is Klatt (1976:1213). He provides the following generalizations: 1) voiceless fricatives are usually 40 ms longer than voiced ones, 2) consonants are longer in word-initial positions than in word-final positions by 10 to 30 ms, 3) fricatives and sonorants in phrasal-final positions are as much as 40 to 100 ms longer than in other positions. Klatt does not address /ʃ/ per se, but these measurements provide us with a basis for comparisons. Crystal and House (1982:710, Table VI) give durational characteristics of English segments. They also do not address [ʃ] per se, but report that voiceless fricatives last on average 118 ms in slow speech. When they appear in strong syllables, they last 133 ms (Crystal and House 1988: 1575, Table 1). For [ʃ] specifically, Jongman et al. (2000:1260) report that its duration is 178 ms. Gordon et al.'s data (2000) examined the durational data of [ʃ] in seven non-Western languages and found that its mean duration is 171 ms. The agreed upon limen of perception of duration is 10 ms. This finding goes back to more than 50 years (Hirsh 1959). He noted on page 767 of his article that if the duration distance between two speech signals is ≥ 17 ms, they are "perceived correctly." Numerous experiments have been conducted since then, and they have all confirmed that the JND of duration is 10 ms (Miller 1989:2122, Kent and Read 2002:11, to mention only these two). With this limen of perception in mind, let's see how long the participants in our study produced their [ʃ]s.

Duration	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Male 1-Initial	235	225	231	260	213	209	215	208	256	236	215	227
Male 2-Initial	177	175	178	183	168	152	152	159	142	177	165	166
Male 3-Initial	187	217	187	189	195	178	180	210	200	202	194	194
Mean-Initial	**199**	**205**	**198**	**210**	**192**	**179**	**182**	**192**	**199**	**205**	**191**	**196**
Male 1-Final	324	280	290	263	286	260	234	297	280	286	267	278
Male 2-Final	305	268	265	293	231	253	260	256	278	274	282	269
Male 3-Final	400	381	404	425	387	302	352	333	360	446	372	378
Mean-Final	**343**	**309**	**319**	**327**	**301**	**271**	**282**	**295**	**306**	**335**	**307**	**308**
Overall Mean	**271**	**257**	**259**	**268**	**246**	**225**	**232**	**243**	**252**	**270**	**249**	**252**
St. Duration	**86**	**71**	**83**	**88**	**79**	**56**	**70**	**64**	**74**	**95**	**74**	**75**

Table 9: Duration of [ʃ] in Male Speech

Duration	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Female 1-Initial	317	340	326	306	310	348	323	323	346	355	320	328
Female 2-Initial	278	278	234	276	241	156	278	266	290	292	271	260
Female 3-Initial	304	268	253	240	253	240	243	220	277	298	265	260
Female-Mean	**299**	**295**	**271**	**274**	**268**	**248**	**281**	**269**	**304**	**315**	**285**	**282**
Female 1-Final	237	251	229	213	210	198	198	199	201	198	217	213
Female 2-Final	227	239	199	211	203	208	206	213	198	231	196	211
Female 3-Final	198	217	197	216	203	207	230	218	234	237	224	216
Female-Mean	**220**	**235**	**208**	**213**	**205**	**204**	**211**	**210**	**211**	**222**	**212**	**214**
Overall Mean	**260**	**265**	**239**	**243**	**236**	**226**	**246**	**239**	**257**	**268**	**248**	**247**
St. Deviation	**46**	**42**	**47**	**39**	**41**	**65**	**47**	**46**	**57**	**57**	**45**	**45**

Table 10: Duration of [ʃ] in Female Speech

Our findings confirm Klatt's observation that fricatives at the end of words are longer than those at the beginning of words. In our data, the mean duration of [ʃ] at the end of words is 261 ms, as opposed to 239 ms at the beginning of words. It is also worth noting that the participants in our study produce longer [ʃ]s at the beginning of words (239 ms) than those in Jongman et al.'s (2000:1260). Their [ʃ] lasted 178 ms. The durational difference between the two is 61 ms.

7.1 Variations in Duration by Gender

The mean duration of [ʃ] produced by males is 252 ms versus 247 ms by females. Since the JND in duration between genders is less than 10 ms, one may be tempted to conclude that duration is not perceptually significant. However, this would be an erroneous conclusion because when we examine the context in which [ʃ] occurs, we see that there is a clear gender demarcation. The mean duration of [ʃ] in word-initial position in male speech is 196 ms. In female speech, it is 282 ms. There is a difference of 86 ms between genders. This difference is perceptually significant. We conclude, therefore, that there is a clear gender difference when [ʃ] occurs at the beginning of words. The gender difference is also obvious at the end of words. However, here the gender roles are reversed. Males produce longer [ʃ] (308 ms) than females (214 ms). The difference between is 94 ms. The gender difference can be summarized as follows: males lengthen their [ʃ]s at the end of words, whereas women lengthen theirs at the beginning of words. In either case, the gender difference in the pronunciation of [ʃ] is clear.

8.0 The Relevance of the Intensity

It is commonly stated in the acoustic phonetic literature that the smallest intensity difference between two speech signals that the human ear can perceive is 1 dB (Ladefoged 2003:90). However, this intensity limen applies only to anechoic laboratory settings (Burg et al. 2013:8). When we listen to speech under normal circumstances, the JND in intensity is 3 dB (Moore 2007:460). This is the limen that is used in all sound level meters. The sensitivity specifications of audio products sold in the US must by law meet the 3 dB requirement. This explains why we use the JND of 3 dB to gauge the perception of the intensity of [ʃ].

Intensity	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Male 1-Initial	81	81	81	80	80	81	80	80	81	81	80	80
Male 2-Initial	81	81	80	80	80	80	80	80	80	80	80	80
Male 3-Initial	80	80	80	79	79	79	78	79	79	79	80	79
Mean-Initial	**80.6**	**80.6**	**80.3**	**79.6**	**79.6**	**80**	**79.3**	**79.6**	**80**	**80**	**80**	**80**
Male 1-Final	80	80	80	80	79	80	78	80	80	80	79	79
Male 2-Final	78	79	79	78	77	73	75	75	76	74	77	76
Male 3-Final	80	80	79	78	79	75	73	75	74	79	76	77
Mean-Final	**79.3**	**79.6**	**79.3**	**78.6**	**78.3**	**76**	**75.3**	**76.6**	**76.6**	**77.6**	**77.3**	**77.7**
Overall Mean	**80**	**80**	**79**	**79**	**79**	**78**	**77**	**78**	**78**	**78**	**78**	**78**
St. Deviation	1	.69	.75	.89	1	2.94	2.56	2.26	2.49	2.28	1.61	1.54

Table 11: Intensity of [ʃ] in Male Speech

Intensity	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	**Mean**
Female 1-Initial	78	78	79	76	78	77	77	78	79	79	78	77
Female 2-Initial	78	77	74	75	75	75	74	73	77	75	72	75
Female 3-Initial	70	73	70	70	70	70	75	69	69	69	68	70
Mean-Initial	**75.3**	**76**	**74.3**	**73.6**	**74.3**	**74**	**75.3**	**73.3**	**75**	**74.3**	**72.6**	**74.3**
Female 1-Final	81	80	80	80	80	80	80	80	80	80	80	80
Female 2-Final	78	76	75	74	74	73	74	73	71	73	72	73
Female 3-Final	80	80	79	80	79	78	80	79	79	78	79	79
Mean-Final	**79.6**	**78.6**	**78**	**78**	**77.6**	**77**	**78**	**77.3**	**76.6**	**77**	**77**	**77.7**
Overall Mean	**77**	**77**	**76**	**75**	**76**	**75**	**76**	**75**	**75**	**75**	**74**	**75**
St. Deviation	**3.55**	**2.43**	**3.53**	**3.49**	**3.41**	**3.30**	**2.57**	**3.94**	**4.27**	**3.82**	**4.42**	**3.38**

Table 12: Intensity of postvocalic [ʃ] in Female Speech

8.1 Context-based and Gender-based Variations in Intensity

Intensity levels vary in relation to the volume of the area where a sound is made. Larger areas cause intensity to increase; whereas smaller areas cause intensity to decrease. For example, Lehiste and Peterson (1959:432) found that in American English, low vowels have a greater intensity than non-low vowels because low vowels call for the mouth to be open wider. Does the vowels that [ʃ] precedes or follows cause an increase or a decrease in intensity? The answer to this question is no. There is no perceptual difference in intensity based on the vowel context. The intensity of [ʃ] is perceptually the same before high, mid, and low vowels. It is also the same before front, central, and back vowels. There is however, an intensity difference based the position of [ʃ] in the syllable. Furthermore, this difference is gender-specific.

In male speech, [ʃ] is louder at the beginning of words (80 dB) than at the end of words (77 dB). This difference is perceptually salient because it is at least 3 dB. In female speech, the intensity difference is also perceptually salient but it is the opposite of what takes place in male speech. In female speech [ʃ] is louder at the end of words (77 dB) than at the beginning of words (74 dB). Even if the syllable context is excluded from consideration, the intensity difference between the [ʃ] produced by males (78 dB) and that produced by their female counterparts (75 dB) is perceptually salient because the difference is 3 dB. This is not surprising, given the fact that anatomical differences between males and females are uncontroversial. Stevens (1998:24) indicates that the average vocal tract volume for adult males is 170 cm^2, as opposed to 130 cm^2 for adult females. Since intensity correlates with the volume of the area where a sound is produced, this difference is totally expected.

9.0 Summary and Relevance of Acoustic Correlates

It was stated in the introduction that we were pursuing three main goals in this paper. We wanted to know if males and females produced [ʃ] differently. We wanted to determine which acoustic correlate(s) of [ʃ] could be used in forensic acoustics. Finally, we wanted to know if CMNE talkers produced [ʃ] differently from speakers from other regions of the US. Table 13 summarizes our findings as they relate to the first two goals.

No	Correlates	Males	Females	Intelligibility	Forensic
1.	CoG-Overall	3,746 Hz	4,052 Hz	No	No
2.	CoG-Onset	3,900 Hz	4,194 Hz	No	No
3.	CoG-Coda	3,593 Hz	3,909 Hz	No	Probably
4.	F2-Overall	2,581 Hz	2,745 Hz	No	No
5.	F2-Onset	2,631 Hz	2,829 Hz	Yes	Yes
6.	F2-Coda	2,532 Hz	2,661 Hz	No	No
7.	Duration-Overall	252 ms	247 ms	No	No
8.	Duration-Onset	196 ms	282 ms	Yes	Yes
9.	Duration-Coda	308 ms	214 ms	Yes	Yes
10.	Intensity-overall	78 dB	75 dB	Yes	Yes
11.	Intensity-Onset	80 dB	74 dB	Yes	Yes
12.	Intensity-Coda	77 dB	77 dB	No	No

Table 13: Summary of Acoustic Correlates

It has been discussed in the acoustic phonetic literature that fricatives can help discriminate between male and female speakers. However, this claim lacks in specificity. The information in Table 13 helps see which correlates help distinguish the [ʃ]s produced by males from those produced by females.

In our study CoG does not appear to be a robust correlate for speaker identification by gender. Yet, it may be useful for differentiating among individuals of the same gender. Since Male 1 has a very low CoG, this correlate can be used in forensic acoustics to distinguish him from a pool of individuals who otherwise produce [ʃ] identically. F2 discriminates between males and females in CMNE, but it is a robust cue only in syllable onsets. This cue is particularly strong because it is confirmed by Jongman et al. (2000:1259). According to our findings, duration can be used as a cue to determine the gender of the speaker in CMNE. However, what matters in this case is not the overall duration, but duration based on the syllable context. Females produce longer [ʃ]s in syllable onsets, while their male counterparts lengthen theirs in syllable codas. Intensity is also a strong discriminating cue. Males produce their [ʃ]s more loudly than females. This is particularly true in syllable onsets, but not in codas. There is nothing unexpected about the role that intensity can play in forensic acoustics because the glottises of males and females are anatomically different.

We cannot address the third goal of our research in detail because there is practically no information on dialectal variations in the production of [ʃ]. The measurements in Jongman et al. (2000) cannot be used for dialect comparisons because they do not provide pertinent dialect information about the participants in their study. Haley et al. (2010) mention where their participants are from: five of them are from the Southeast, three from the Midwest, one from New York, and one from Florida. This information is so general that it does not allow for comparisons between dialects. To the best of our knowledge, our study is the only one that examines the pronunciation of [ʃ] from a sociophonetic perspective. It offers others the possibility to compare their findings with ours, but we cannot compare our results with any other study because, "There have been surprisingly few studies of the acoustics of fricatives" (Ladefoged and Maddieson 1996:173).

ABOUT THE AUTHORS

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.

Maria Bloch collected the data used in this paper for her capstone project to fulfill the requirements for the BA degree in Linguistics from St. Cloud State University. She is currently pursuing her graduate studies in Forensic Linguistics at Hofstra University in New York. She can be reached via e-mail at mbloch1@pride.hofstra.edu.

References

Bekesy, George V. 1947. The Variation of Phase along the Basilar Membrane with Sinusoidal Vibrations. *Journal of the Acoustical Society of America* 19 (3): 452-460.

Boersma, Paul and David Weenink. 2010. Praat: Doing Phonetics by Computer. Version 5.3.16. Available online at: http://www.praat.org/.

Crystal, Thomas H. and Arthur S. House. 1982. Segmental Durations in Connected Speech Signals: Preliminary Results. *Journal of the Acoustical Society of America* 77 (3): 705-716.

Crystal, Thomas H. and Arthur S. House. 1988. Segmental Durations in Connected Speech Signals: Syllabic Stress. *Journal of the Acoustical Society of America* 83 (4): 1574-1586.

Everest, Alton F. and Ken C. Pohlmann. 2015. *Master Handbook of Acoustics*. Sixth Edition. McGraw Hill Education: New York.

Evers, V., Reetz, H., & Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics, 26*, 345-370. Retrieved April 3, 2015, from http://www.ling-phil.ox.ac.uk/files/people/crosslinguistic_acoustic_categ.pdf

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., and McSweeny, J. (1999). Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research* 42: 663- 677

Fry, D. B. 1979. *The Physics of Speech*. Cambridge University Press: New York.

Gold, E. (n.d.). Forensic speaker comparison evidence: the international picture. Retrieved February 9, 2016, from https://www.kfs.oeaw.ac.at/publications/iafpa_abstracts/special_session/Special_session_E_Gold.pdf

Gordon, Matthew, Paul Barthmaier, and Kathy Sands. 2000. A Cross-linguistic Acoustic Study of Voiceless Fricatives. Online Version, retrieved on February 9, 2016 from http://www.linguistics.ucsb.edu/faculty/gordon/fricativeacoustics.pdf

Haley, Katarina L, Elizabeth Seelinger, Kerry C. Mandulak, and David J. Zajac. 2010. Evaluating the Spectral Distinction between Sibilant Fricatives through a Speaker-centered Approach. *Journal of Phonetics* 38:548-554

Hasen, Colin H. 2001. Fundamentals of Acoustics. *Occupational Exposure to Noise: Evaluation, Prevention, and Control*. Ed. By Bernice Goelzer, Colin H. Hasen, and Gustav A. Sherndt. Dortmund, Germany: World Health Organization. Retrieved from: http://www.who.int/occupational_health/publications/noise1.pdf

Hirsh, Ira J. 1959. Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America* 31 (6): 759-767.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A Guide to the International Phonetic Alphabet*. New York: Cambridge University Press.

Jones, M., & Knight, R. (2013). *The Bloomsbury Companion to Phonetics* (1st ed.). New York: Bloomsbury Academic.

Jongman, A., Wayland, R., Wong, S. (2000). Acoustic characteristics of English fricatives. *Acoustical Society of America* 108: 1252- 1263.

Kent, Ray D. and Charles Read. 2002. *The Acoustics Analysis of Speech*. Delmar, Cengage Learning: New York.

Klatt, Dennis H. 1976. Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence. *Journal of the Acoustical Society of America* 59 (5): 1208-1221.

Koffi, Ettien. 2016. The Acoustic Correlates of [±ATR] Vowels: An Analysis by Reference Levels of Anyi Vowels. *Linguistic Portfolios* 5:115-134.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of World Languages*. Blackwell Publishers Ltd: Malden, MA.

Lehiste, Lise and Peterson, Gordon E. 1959. Vowel Amplitude and Phonemic Stress in American English. *Journal of the Acoustical Society of America* 31 (4): 428-435.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge Studies in Speech Science and Communication. Cambridge University Press: New York.

Mattingly, I. (1998). A short history of acoustic phonetics in the U.S. Retrieved April 4, 2015, from http://www.haskins.yale.edu/Reprints/HL1144.pdf.

Miller, James D. 1989. Auditory-perceptual Interpretation of the Vowel. *Journal of the Acoustical Society of America* 85. 5. 2114-2134.

Stevens, Kenneth N. 2000. *Acoustic Phonetics.* The MIT Press: Cambridge, Massachusetts.

Stuart-Smith, J. (2007) Empirical evidence for gendered speech production: /s/ in Glaswegian. In: Cole, J. and Hualde, J.I. (eds.) *Laboratory Phonology 9.* Series: Phonology and phonetics. Mouton de Gruyter: New York, USA, pp. 65-86.

Styler, Will. 2013. Using Praat for Linguistic Research. University of Colorado at Boulder Phonetics Lab. Available online at http://savethevowels.org/praat/UsingPraatforLinguisticResearchLatest.pdf.

The Editors of Encyclopædia Britannica (2015). Centre of gravity. In: *Encyclopædia Britannica*. Retrieved April 6, 2015, from http://www.britannica.com/science/centre-of-gravity

Zsiga, Elizabeth C. 2013. *The Sounds of Language*. Wiley-Blackwell:Malden, MA.