Research Article
J Forensic Res
Volume 13:1, 2022

Journal of Forensic Research

ISSN: 2157-7145 Open Access

Speaker-specific Information in the Acoustic Characteristics of English Fricatives

Sara Carralero-Fernández*

Departamento de Lingüística, Estudios Árabes, Hebreos y de Asia Oriental, Universidad Complutense de Madrid C/Fuencarral, 160, 7º, Madrid (28010), Spain

Abstract

There is still much to learn about speakers' similarities and differences in the field of Forensic Phonetics with respect to consonant acoustics. This article analyses of acoustic features of three sibilants /s, z, ʃ/ in British English. The analyses have been carried out on twenty male speakers from the DyViS corpus focusing on static features (intensity, centre of gravity, standard deviation, skewness and kurtosis) and dynamic features (centre of gravity depending on F2 vowel onset and offset) to see if they cue speaker-specific information. The results obtained demonstrate the high speaker-specificity of centre of gravity, standard deviation and intensity. However, we must be careful with intensity because it depends on the recording circumstances. As for skewness and kurtosis, they show speaker-specificity for /z/, but results are weaker the other two. This article has shown that spectral and acoustic properties of these three sibilants in English present promising results.

Keywords: Speaker-specificity ● Sibilants ● Forensic speaker comparison ● Dynamic features ● Static features

Introduction

There is still much to learn about speakers' similarities and differences in the field of Forensic Phonetics, especially with respect to consonant acoustics. The end of the 20th century and the beginning of the 21st saw a substantial contribution from the field of Phonetics to Forensic Speech Science. Learning about properties of sounds and whether they are speaker-dependent -or not- allowed researchers and forensic linguists to use those properties for speaker comparison casework.

Properties of sounds can be divided into static and dynamic. Traditionally, researchers have studied the so called 'static' properties [1]. However, recent investigations have started to include 'dynamic' features of speech [2]. Static properties refer to the reflections of anatomical dimensions such as formant frequency or spectral peak location; whereas, dynamic properties are those referring to the movement of the individual's speech organs such as locus equations and relative amplitude. The dynamic properties carry the most important information about the speaker since the movement of those organs is speaker-specific. Static features such as the duration of a vowel or formant frequencies are also speaker-dependent but to a lesser extent [3].

Many studies have analysed the spectral characteristics of consonants and vowels [4].and a number of studies have investigated the acoustic and spectral characteristics of fricatives in different dialects of English [5] and other languages such as Swedish [6], German and Greek. Yet, there are only few studies that have investigated fricatives in English and their dynamic features in different contexts for forensic purposes [7]. It is necessary to delve into fricative acoustics since changes in the precise location and length of constriction may alter the size and shape of the cavities behind and in front of the constriction; that is, resonance of the sibilants will vary per speaker depending on the size and shape of the oral cavity. This changes the values of the acoustic features connected to the cavities that have been altered Besides, according to , the energy loci of English fricatives and duration of fricatives in specific phonological environments can be found among the features commonly considered in speaker comparisons.

Despite the fact that included fricatives in her research, she used read-speech; therefore, non-spontaneous speech. Her purpose was to explore acoustic parameters of five consonants /m, n, ŋ, l, s/ in two dialects of British English. The parameters she analysed were normalised duration, centre of gravity, standard deviation, frequency at peak amplitude and frequency at a minimum amplitude for /m, n, ŋ, l/ and skewness and kurtosis for /s/. Among other aims, she intended to discover whether the parameters analysed for these consonants showed speaker-specificity or not.

The basis of this research relies on the notion that "every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect, and the assumption that this idiolect will manifest itself through distinctive and idiosyncratic choices in texts" [8]. In fact, no one is able to repeat the exact same realization of an utterance twice [9]. It is assumed therefore that each individual presents his/her own features when it comes to speech production and that makes it possible to recognize individuals by analyzing the idiosyncratic choices. However, some features do not depend on the choices the speaker makes, but on the individual's speech organs and on anatomical dimensions.

My research therefore intends to analyses segments of simulated spontaneous speech to contribute to the field by analyzing data. Inter- and intra-speaker variation in simulated spontaneous speech is the type of data researchers and experts are likely to encounter. Hence, since it is a relatively new field of study I aim to build on Kavanagh's C [7] research and contribute to current findings by analyzing sibilant fricatives in British English with a new set of measurements.

Secondary objectives will be determining if intraspeaker variation is smaller than interspeaker variation. I also intend to give account of speaker-specific features that can be used with relatively certainty to distinguish between speakers. It is expected that the analyzed consonants can be used as speaker-specific features independently of when they are being analyzed. Furthermore, another methodological aim arises: collecting data about how the selected segments behave depending on the context and if they are constant within those contexts, analyzed how /u:/ varied depending on the context. The onset and offset of vowels or neighboring consonants will very likely affect the production of the sibilants too, that is, the spectral peak might be higher or lower, for instance.

***Corresponding author:** Sara Carralero-Fernández, Departamento de Lingüística, Estudios Árabes, Hebreos y de Asia Oriental, Universidad Complutense de Madrid C/ Fuencarral, 160, 7º, Madrid (28010), Spain, E-mail: scarrale@ucm.es*

Received 11 January, 2022, Manuscript No. jfr-22-51549; **Editor Assigned:** 13 January, 2022, PreQC No. P-51549; QC No.Q-51549; **Reviewed:** 18 January, 2022; **Revised:** 25 January, 2022, Manuscript No. R-51549; **Published:** 01 February, 2022, DOI: 10.37421/2157-7145.2022.13. 478

Methodology

Acoustic literature on fricatives

Before delving into the sibilants' literature, it is important to analyses the literature on fricatives. Fricatives are produced by forcing airflow production through one or two constrictors creating turbulence in the air and, therefore, the frication noise. The turbulence in the airflow can be produced in two different ways. One way is to produce a constriction by approximating two articulators that, if close enough, can make the airflow turbulent producing frication noise. However, turbulence in the airflow can be also produced by channeling the air at an obstacle in the vocal cavity so that frication noise happens. The first way of producing fricatives is the generation of labiodental fricatives (/f/, /v/) and the second one is used to produce sibilants (e.g. /s/, /z/).

Fricatives can be classified regarding the place of articulation. English fricatives are usually divided into four groups: labiodental /f, v/, (inter)dental /θ, ð/, alveolar /s, z/ and palato-alveolar /ʃ, ʒ/.

In addition, fricatives are also divided into two groups: sibilants and non-sibilants. Sibilance is an acoustic property which refers to strong energy in the frication noise at high frequencies [10] that is usually produced by a narrow constriction.

The consonantal segments that will be analyzed are the three English sibilants whose places of articulation are the alveolar and palato-alveolar areas, namely, /s, z, ʃ/. These three sounds (voiced and voiceless) are distinguished because of different place of articulation and vocal fold activity. The vocal fold activity carries perceptual cues that leave a trace in the spectrogram.

Sibilants have been studied incorporating a wide variety of methods and parameters for analysis including both static and dynamic measures. Particularly, researchers have focused on spectral measures [11] on absolute and relative duration and amplitude measures. As summarizes, the most common aim of fricative acoustic investigations is to identify any possible acoustic correlates of phonetic features such as place of articulation and voicing. On another note, in the past years, researchers have kept studying fricatives and the acoustic cues they might hold. They have directed their attention towards whether social identities such as gender or social class are reflected in the acoustic properties, in particular the /s/ [12].

Production of the alveolar sibilants /s, z/

Languages commonly have at least one sibilant which is usually the voiceless alveolar /s/. In fact, Maddieson I, et al. [10] found out that 88.5% of the languages happen to have at least this sound. The preference for the voiceless alveolar sound might be due to the fact that the most frequent sounds in languages' inventories are those with great acoustic energy, [10] but it can also be due to aerodynamic reasons such as the relative ease to produce and maintain the airflow. In this case, /s/ is the fricative with the highest spectral peak, at around 4-5 kHz.

The voiced alveolar /z/ sound is the fourth most frequent fricative after its voiceless counterpart and the voiceless palato-alveolar /ʃ/ and labio-dental /f/. Usually, voiced sounds are less common than the voiceless one. The voiceless alveolar is relatively easy to pronounce, although it can present many allophones that in some languages are distinctive, i.e. Polish [10]. In order to pronounce the /s/, there must be a constriction between the tip or the blade of the tongue and the alveolar ridge, albeit some allophones produce the constriction against the teeth. The frication noise is produced when the airflow hits the upper teeth or the alveolar ridge [11].

Just as the voiceless alveolar, /z/ is pronounced by creating a constriction by placing the tip or blade of the tongue close enough to the alveolar ridge or upper teeth and generating, therefore, the frication noise which is produced when the airstream finds the upper teeth as obstacles [11]. The difference arising between these two sounds is that, when producing /s/, the glottis does not vibrate, whereas it vibrates when producing /z/. As Stuart SJ, et al. [11] showed low frequency peaks are related to resonances of the back cavity, whereas high frequency peaks are related to the front cavity. There are also values related to the noise source and its location within the vocal tract and the distance between the noise source and the constriction. Due to this dependence on the oral cavity of each individual, there might be capacity for inter-speaker variability in /s/ acoustics.

Production of the palato-alveolar sibilant /ʃ/

As for the production of the palato-alveolar sibilant, the voiceless one /ʃ/ is the second most frequent fricative after the /s/ according to Maddieson I, et al. [10]. As it has been previously mentioned, voiceless sibilants are considerably more favoured over the voiced counterparts. In fact, these authors found that the voiced palato-alveolar /ʒ/ is much less frequent. In the present study, this phoneme has been excluded due to the low number of occurrences in the corpus analysed.

Like /s, z/, the voiceless palato-alveolar is articulated with the edge of the tongue raised behind the upper side teeth so that a space is formed along the midline of the tongue creating a superficial dip. A large portion of the tip or the blade of the tongue rises to form a narrow channel with the alveolar ridge and the front of the hard palate.

Segmental acoustic literature on /s, z, ʃ/

This subsection focuses on how measures relate to (socio)linguistic dimensions. Regarding sibilants in English, the vast majority of the studies on fricatives address both sibilants and non-sibilants, the most relevant for the current investigation being those carried out by Jongman A, et al [12]. However, we shall go through most of the studies carried out so far regarding sibilants. In addition, there are studies that have studied only /s/: Stuart-Smith J, et al. [11] and Munson B [13]. Some evaluated the spectral distinction between sibilants (including the palato-alveolar fricative) through a speaker-centred approach [14]. Yet they did not do so from a forensic perspective. The results they obtained, they claim, might be useful for clinical applications since there is an intra-speaker overlap in the spectral mean in both /s/ and /ʃ/ in consonant – vowel – consonant (CVC)context.

Jongman A, et al [12] studied place of articulation of American English fricatives in CVC context in both men and women. They measured duration, mean spectral peak, absolute and relative amplitude, centre of gravity, variance, skewness and kurtosis. Overall, they identified differences between fricatives in the duration, i.e. sibilants are significantly longer than the rest of the fricatives. Regarding sociolinguistic variables, they discovered that women produced shorter fricatives, albeit they were not sure whether it was a female trend or a fricative trend. It was probably a female trend as it was later identified by Stuart-Smith J, et al. [11]. With regards to mean spectral peak and amplitude, Jongman A, et al [12] discovered that they can discriminate between the places of articulation and, furthermore, that they present gender cues. Finally, the four spectral moments showed that /s, z/ were produced with the highest amount of energy at the highest frequencies. Besides, the four spectral moments also differentiated between male and female speakers. In parallel, Stuart- Stuart-Smith J, et al. [11] kept studying in the line of Gordon and they also carried out a study focused on the differences between women and men when producing /s/. They aimed to find differences between 'sex' and 'gender'. Nevertheless, the authors concluded that "while some aspects of the acoustic signal result from anatomical differences between gender, biological sex provides an acoustic 'frame' for other social factors to work within" (2003: 1854).

Not satisfied with the previous study, Stuart-Smith J, et al. [11] data with different parameters and measurements. He modified the rate of the digitalised recordings and by doing so, he discovered that mean of spectral peaks for women were clearly higher than those of male speakers.

As for the studies taking acoustic measurements that show speaker-specificity, Hughes and Halle (1956) carried out a study focused on English fricatives /f, v, s, z, ʃ, ʒ/ in different vowel contexts and in different positions. They measured spectral peaks finding a correlation between the point of constriction and the amplitude peak location. Location of peaks was found to differ between speakers.

Almost half a century later, Gordon examined fricatives in seven

endangered languages in female and male speakers. Following Jongman A, et al [12] they measured duration, mean spectral peaks and CoG. In the line with previous studies, they found /s/ to be significantly longer than the other voiceless fricatives in five of the languages. In addition, the CoG of /s/ was higher than the rest of the fricatives in all of the languages, but one: Toda. Regarding the potential of finding inter-speaker information, Gordon found that spectral peaks of /s/ discriminated with a high degree of certainty between speakers, both female and male. As mentioned before, gender should have been a controlled factor in this study.

Kavanagh C [7] focused on different consonants (not only fricatives) from a forensic perspective. She studied male speakers simulating to be interviewed by the police so that spontaneous speech was recorded. Her results of /s/ showed that static measurements differed between speakers; however, unlike she predicted, dynamic measures did not vary much between speakers. As Kavanagh C [7] suggests female and male speakers should not be mixed in the same study or, at least, differences between them should be clearly stated. Otherwise, the speaker's gender could interfere with inter-speaker differences.

Materials and Methods

Materials

The DyViS Database is a large-scale, forensically-oriented speech corpus It was developed at Cambridge University as part of the research project 'Dynamic Variability in Speech: A Forensic Phonetic Study of British English'. The corpus was completed in September 2009 and opened to public access for research. In this project 100 male speakers aged between 18 and 25 years old were recorded. They all spoke Standard Southern British English (SSBE). The recordings were made in a sound-treated room in the Phonetics Laboratory of the Department of Linguistics at the University of Cambridge using a Marantz PMD670 portable solid-state recorder with a sampling rate of 44.1 kHz. Each speaker had to use a Sennheiser ME64-K6 cardioid condenser microphone positioned approximately 20 cm from his mouth. All the participants were asked to perform were various tasks but the one chosen for this research were the police interview in which speech is constructed spontaneously using visual stimuli, including prompts to lie [15].

Regarding the segmentation process, each sound file was segmented using Praat (version 6.0.35) and the target segment and word boundaries were marked in a TextGrid file. Once the segments of speech were delimited, they were labelled with the appropriate marker ('s', 'z', 'ʃ' and the neighbouring vowels)

Measurements

For each of the three segments seven acoustic features were analysed, both dynamic and static properties. Following Jongman A, et al [12] and Kavanagh C [7], segments have been measured by their duration, centre of gravity, standard deviation, skewness, kurtosis and locus equations and F2 onset values.

The measures presented below were taken by using two Praat scripts that captured different windows of each segment and each parameter. Windows are small periods of time of the selected segments that help capture differences within the spectrum thereof. They allow us to obtain very specific information of each moment of the segment under analysis and, if it is the case, relate it to the neighboring context. In order to avoid window overlapped if the token was very short, we decided to obtain 20-ms windows. In fact, parameters have been taken at three different windows for static parameters (50%, 75% and 100% of the sibilants' duration) and taken at five different windows for the dynamic ones (20%, 40%, 60%, 80% and 100% of the vowels' duration). In the following result section, parameters and their windows would be indicated as follows: parameter + number.

I have also examined the dynamics of fricatives in (inter)vocalic structures, such as VC, CV and VCV structures and investigated the spectral transition within the selected fricatives.

Duration

Noise duration has been used so far to differentiate sibilants from non-sibilants. Considering that speaker's speech varied in speed both in their own discourse and compared to the rest of speakers, this measure cannot be used to gather speaker-specific information. Nonetheless, it has been measured to check if data was normally distributed.

Intensity

Intensity (dB) was measured at the different windows of the segment since high noise intensity is one of the most distinctive features about sibilants as a class (Basile and Diehl, 1994). Furthermore, there is also distinction between voiced and voiceless sibilants.

Centre of gravity

Centre of gravity of sibilants is a measure of the concentration of energy in the spectrum [7]. This parameter, also known as mean, shows the frequency at which the distribution of the energy in the spectrum is even on either side.

Standard deviation

Similar to CoG, standard deviation (SD) measures the distribution of energy in the spectrum. Particularly, it measures the dispersion or bandwidth of energy surrounding the CoG [11]. SD is calculated by measuring the square root of the second spectral moment, also known as variance [7]. If the energy is dispersed across a wider frequency range, this will result in high SD values, while energy concentrated around the CoG will give low SD values.

Locus equations and F2 onset values

Locus equations measure dynamic properties of speech sounds, since they relate points in the speech signal to F2. "Locus" was first defined by Delattre PC, et al. [16] "a place on the frequency scale at which a transition begins or to which it may be assumed to point".

According to Sussman HM [17] locus equations are calculated by "making straight line regression [that] fits to data points formed by plotting onset frequencies (at the first glottal pulse) of F2 transitions along the y axis and their corresponding mid-vowel (nuclei) frequencies along the X-axis". For Lindblom, locus equations represent and quantify the context-dependent correlation existing between the onset of the vowel and the vowel, depending on the previous or following consonant.

F2 locus has been used for stops and despite the successful results of it; there are not many studies on fricatives. Yet, these are some that have been carried out so far in which F2 locus has been measured [18]. These studies are contradictory: some of them obtained good classifications of fricatives [19], while others Wilde L [18] and Sussman HM [17] showed results in which the fricatives' loci were overlapping.

We obtained the F2 locus from the preceding and following vowels of the sibilants to check how they might affect their centre of gravity and how this varies between speakers. Thus, F2 was measured at vowel onset "starting at the first glottal pulse following cessation of the fricative" [12]. F2 was also measured at vowel offset ending at the last glottal pulse preceding frication noise of the sibilant. Likewise, the script written in Praat automatically analysed the F2 value every 20% of the vowel in order to capture the path the vowel's F2 follows to make the transition towards the sibilant and find out whether that path towards the sibilant is speaker-specific or not. It needs to be highlighted that there were not enough vowels following /z/ as to obtain significant results and there were not enough vowels preceding /ʃ/ to carry out the statistical analysis either.

Skewness and kurtosis

Both skewness and kurtosis provide results about the shape of the spectral energy of the fricative. Skewness constitutes the third spectral moment which measures the symmetry of the distribution of energy in the spectrum of a sound [7]. Results of skewness can either be zero, positive or negative: a zero value represents a perfectly symmetrical distribution; a positive skewness shows that the distribution in which the right tail is longer than the left, whereas a negative

skewness shows the left tail being longer than the right [12].

Kurtosis is the fourth spectral moment which measures how raised or flat the distribution of the energy is. According to Jongman A, et al. [12] positive values represent peaked energy distribution, while negative kurtosis values show relatively flat distributions. If the value is zero, then the distribution is symmetrical, namely, a normal distribution.

Skewness and kurtosis can show how curved or arched the tongue is in the production of the sibilants [7]. Thus, the shape of the sibilants' energy can provide us with information about the tendency of each speaker to place the tongue within his oral cavity which presents a specific configuration that will also determine the shape of the energy.

Linear discriminant analysis

In order to evaluate the speaker discrimination potential of acoustic parameters, the linear discriminant analysis has proved to be an effective conclusion method. LDA is a statistical method which can be used to test if an individual belongs to a group according to a set of variables, known as predictors. In the scope of FSC, this statistical method can be used to assess whether a variable is speaker-specific or not (Kavanagh, 2012).

Statistical analysis

Measures were analysed by using SPSS. As it was mentioned in previous subsections, the four spectral moments were analysed plus intensity. The different dependent variables that have been analysed throughout this research have been taken at different windows, particularly at the 50%, 75% and 100% of the consonant's duration. It is expected that the measurements are correlated at the different windows.

In order to assess the speaker-specificity of each condition, univariate analyses of variance (ANOVAs) were carried out. The independent variable was the speaker, which was added as a random factor. As for the dependent variables, they were intensity, CoG, SD, skewness, and kurtosis in all the four windows. Regarding the dynamic measures, the dependent variable was CoG with vowel as covariate. It is important to note an alpha of .05 was used so that p-values below .05 indicated significance. Furthermore, F-ratio were used as measure to compare inter- and intraspeaker variation since it represents the relation between different sources of variance. A large F-ratio means that there is a high variation among group means, that is, speakers differ highly from one another.

As for dynamic measure's analysis, univariate analyses of covariance (ANCOVAs) were carried out. The independent variable was speaker as a random factor. As for the dependent variables, they were CoG at the 20% and 100% of the segment's duration. These variables were adjusted with F2 vowel formant.

It has to be highlighted that a Natural Logarithmic Transform on skewness and kurtosis measures was applied since neither of them fulfilled the normality assumption. Besides, they were transformed into their absolute values in order to compute the ANOVA. Correlation between the previous results and the transformed ones were computed by running Pearson's r and thus confirmed that they were correlated since the *r* values were close to 1.

Results

Results: /s/

A positive strong correlation was found between the three measures of intensity, r > .9, n=230, p < .001. Similar results were found for CoG, where correlation between the different values was positive and strong, r > .9, n=230, p < .001. As for SD, correlations varied: SD75 was highly correlated with both SD50 and SD100, r > .7, n=230, p < .001, however, correlation between SD50 and SD100 was weaker, r=.57, n=230, p < .001. Regarding correlation for skewness, results of the Pearson's r test showed measures taken at the different windows were not correlated at all, r < .2, n=230, p > .05. As for kurtosis, results varied: correlation between kurtosis50 and kurtosis75 was strong, r=.49, n=230, p < .001; correlation between kurtosis75 and kurtosis100 presented a very weak correlation, r=.3, n=230, p < .001. However, correlation between kurtosis50 and kurtosis100 was not found, r=.1, n=230, p=.032.

Overall, the speaker was found to be a highly significant factor (p < .001) in intensity, centre of gravity and standard deviation at the 50%, 75% and 100% of the segment's duration measurements since they are correlated. Nonetheless, the speaker has shown to be significant on skewness only at the 100% of the segment's duration, whereas kurtosis is only slightly significant at the 50%.

Results: /z/

The different dependent variables that have been analyzed throughout this research have been taken at different time windows. Correlations between them are presented below.

A positive strong correlation was found between the three measures of intensity, r > .9, n=208, p < .001. Similar results were found for CoG where correlation between the different values was positive and strong, r > .8, n=208, p < .001. As for SD, correlation varied: SD75 was highly correlated with both SD50 and SD100, r > .75, n=208, p < .001. However, correlation between SD50 and SD100 was slightly weaker but still strong enough to be significant, r=.63, n=208, p < .001. With regards to skewness, results showed a strong and positive correlation between all the windows, r > .74, n=208, p < .001. Similar results were found for kurtosis. A strong correlation between kurtosis50, kurtosis75 and kurtosis100 was found, r > .78, n=208, p < .001,

Speaker was found to be a highly significant factor (p < .05) in intensity, CoG, SD, skewness and kurtosis at the 50%, 75% and 100% of the segment's duration.

Results: /ʃ/

As previously mentioned, the different dependent variables have been taken at different time windows. A summary of correlations is presented in order to interpret across results. A Pearson's r test was computed to assess the relationship between each variable at the 50%, 75% and 100% of the segment's duration.

A positive strong correlation was found between the three measures of intensity, r > .9, n=234, p < .001. Similar results were found for CoG where correlation between the different values was positive and strong, r > .9, n=234, p < .001. Correlation for SD was similar to intensity and CoG, a strong positive correlation was found, r > .85, n=234, p < .001. As for skewness, results of the Pearson's r test showed measurements taken at the different time windows were not correlated, r < .3, n=234, p > .05. However, p value showed significance between skewness75 and the other two windows. With regards to kurtosis, results varied: correlation between kurtosis50 and kurtosis75 was strong, r=.47, n=234, p < .001; correlation between kurtosis75 and kurtosis100 presented a weaker correlation, r=.34, n=234, p < .001. However, correlation between kurtosis50 and kurtosis100 was a weak correlation, r=.1, n=234, p=.004.

Overall, the speaker was found to be a highly significant factor (p < .05) in intensity, centre of gravity and standard deviation at the 50%, 75% and 100%. As for skewness, it was significant at the 75% and 100% of the segment's duration, whereas for kurtosis it was significant at the 50% and 100%.

Dynamic measures

Apart from the static measures, CoG was measured in five smaller time windows to capture dynamic movements in the spectrum over time, similar to Kavanagh's (2012). Means of each window for each speaker are displayed in Figures 1-3.

It would be wrong to assume that values (i.e. CoG) remain constant throughout the speech sound, that is why it is important to also look at measures dynamically. As it is shown in Figures 1-3, CoG varies over the course of production of /s, z, ʃ/ for each speaker. The variability between speakers can be appreciated at the onset of production and at the offset of the token where it lies within different Hz for some speakers.

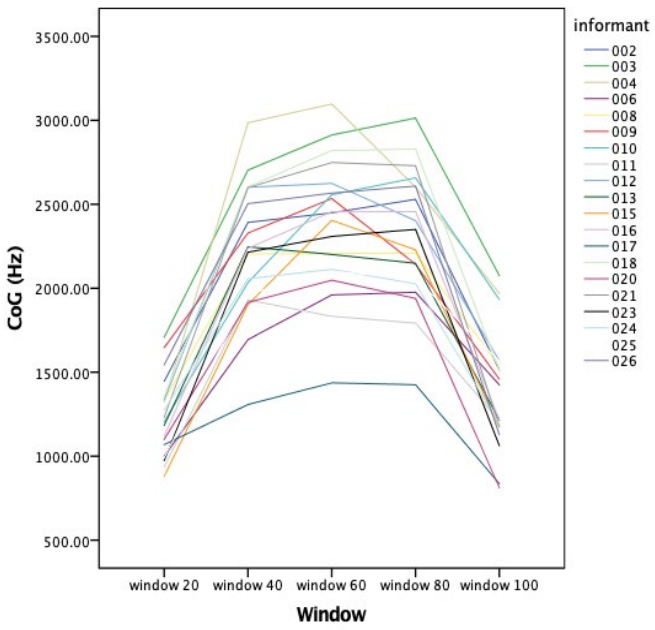

Figure 1. Mean CoG of /s/ at onset, after-onset, midpoint, before-offset, offset, showing dynamic movement.

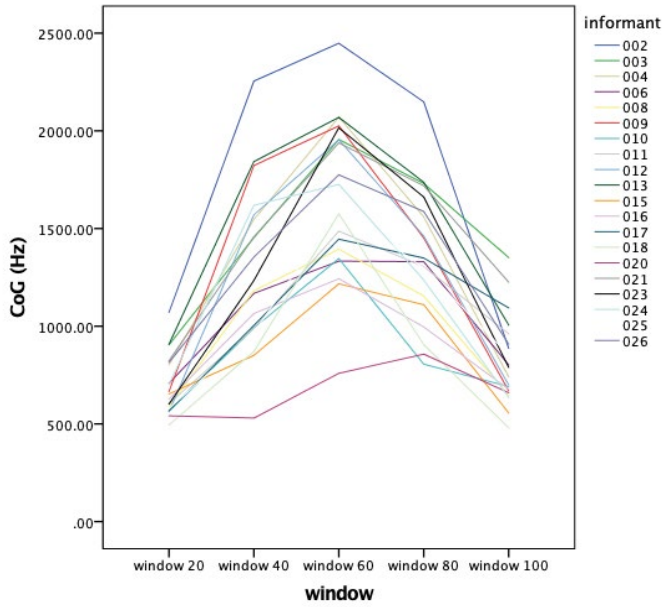

Figure 2. Mean CoG of /z/ at onset, after-onset, midpoint, before-offset, offset, showing dynamic movement throughout production.

ANCOVA results for CoG at onset and offset proved to be highly significant for speaker, with the highest F-ratio overall for /s/ ($F(19, 153)=3.341$, $p < .001$) at onset and ($F(19, 94)=2.575$, $p < .001$) for onset. Hence, a main effect of speaker on CoG considering both left and right vowel context was found. Bonferroni post-hoc comparisons showed that there is a high degree of inter-speaker variability.

As for the frequency of appearance of vowels preceding /s/, that is, left vowel context, it is as follows (Figures 4 and 5): /ɪ, ʌ, a, ə/. The frequency of appearance of the vowels following /s/ –right vowel context– is /ɪ, ə/ in the first place and then there are a few tokens of /ɒ, ʌ, u:/. As it has been hypothesised, F2 of vowels pulls down the onset of the sibilant and therefore the first window of CoG is lower than would be expected if only the measurement of the whole segment would have been taken. As can be observed from the scatter graph (Figure 4) and the descriptive statistics, /ɪ/ is the vowel that pulls it down the most. This fact turns out to be unexpected since /ɪ/ is a near close and front vowel, that is, it is close to the place of articulation of /s/. It is true;

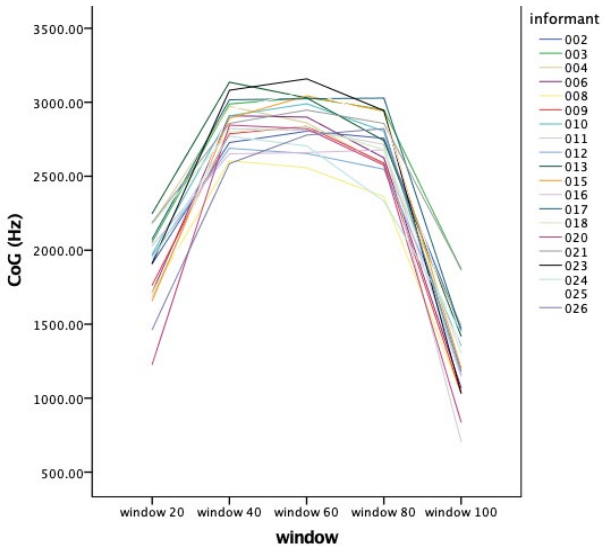

Figure 3. Mean CoG of /ʃ/ at onset, after-onset, midpoint, before-offset, offset, showing dynamic movement throughout production.

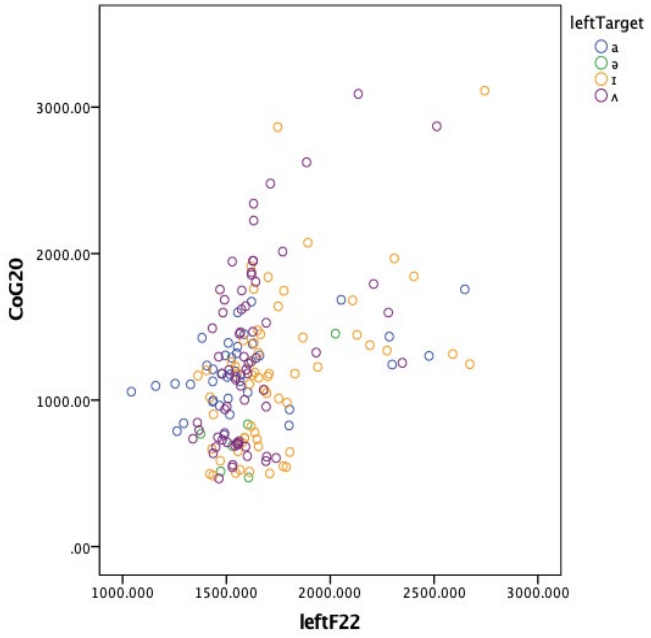

Figure 4. Influence of vowel over /s/ token's onset.

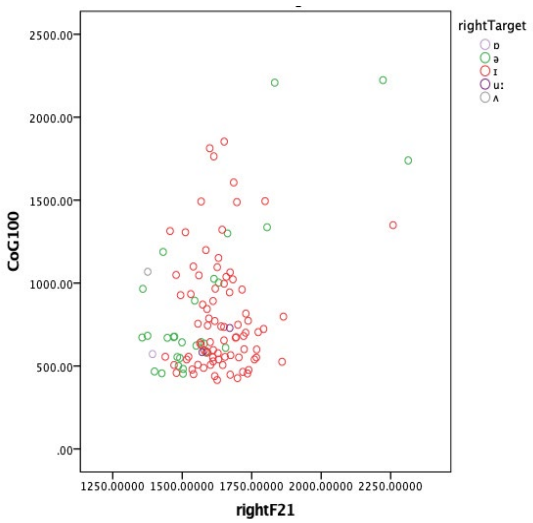

Figure 5. Influence of vowel over /s/ token's offset.

however, that this vowel is produced between 1.25-1.75 kHz and the frequency of /s/ is significantly higher. Hence, it is mostly pulled down by this vowel. Yet, /ɪ/ is more spread apart than the rest of the vowels, meaning that some of the tokens are produced at low frequencies but some others are produced at even 3 kHz. As for the vowels influencing the offset of /s/, it is noteworthy that it is /ɪ/ the one that have a bigger effect on /s/. Nevertheless, as it can be observed in the scatter graph, it appears that when followed by a vowel, CoG of /s/ is not as pulled down as it is when preceded by one.

ANCOVA results for CoG at onset were significant for speaker, with the highest F-ratio overall for /z/ ($F(19, 164)=2.199$, $p=.004$). However, they were not significant at offset ($F(15, 29)=.618$, $p=.836$), meaning that in the case of /z/, CoG is not particularly influenced by following vowels. This can be, however, due to the small number of tokens in the right vowel context. In this case, the frequency of the appearance of vowels preceding /z/ is as follows (Figures 6 and 7): /ə, ɪ, ɛ, ɔ:, i:/. The vowel that pulls down the CoG of /z/ the most is /ə/. Yet, it is not particularly relevant, since results were not significant for /z/.

Regarding /ʃ/, ANCOVA results proved to be highly significant for speaker only at offset, with the highest F-ratio overall ($F(19, 211)=2.763$, $p < .001$). ANCOVA could not be carried out at onset due to the lack vowels preceding the onset of the token. As for the relevance of the right vowel context, the ones that appear the most are /i, ɒ, ɪ/ and the one that have a bigger effect on /ʃ/ is /ɒ/ (Figure 7). This might be due to the fact that it is an open back vowel; hence, the path from the place of articulation of /ʃ/ to the place of articulation of /ɒ/ needs to produce a lower CoG so that the transition is faster from one to the other. It is also remarkable the fact that the distribution of the dots in the scatter graph is considerably different from the ones of /s/ and /z/. Despite the fact that vowels pull down the offset of /ʃ/, there many others (e.g. /i/) that are produced at frequencies between 1.7-2.2 kHz and produce the offset of /ʃ/ within a range between 500 Hz and 3 kHz.

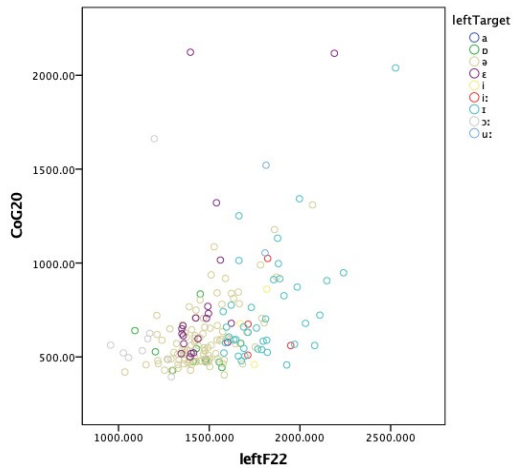

Figure 6. Influence of vowel over /z/ token's onset.

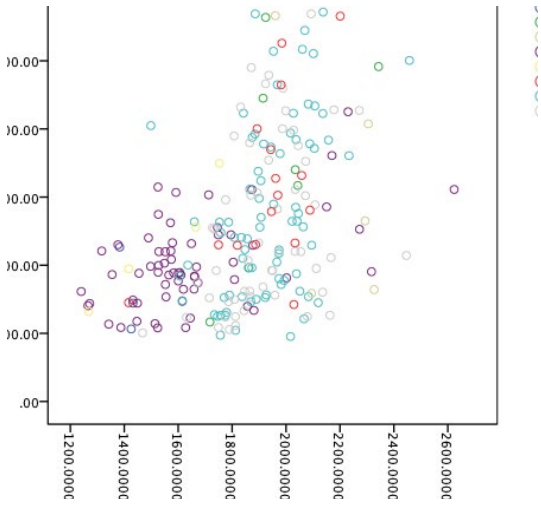

Figure 7. Influence of vowel over /ʃ/ token's offset.

Discussion

The research aims were 1) to analyse the static and dynamic acoustic features of three sibilants in spontaneous speech; 2) to collect data about how the selected segments behave depending on the context; 3) to determine if intraspeaker variation is smaller than interspeaker variation; 4) to give an account of speaker-specific features that can be used in FSC casework; and 5) to suggest a detailed methodology to follow in further studies related to Forensic Phonetics and Forensic Speaker Comparison in different languages. These aims have been tackled from an explicitly speaker-specific perspective. In order to achieve these aims, this research has segmented and described the distribution of sounds for each speaker, by evaluating the statistical effect of intensity, CoG, SD, skewness, kurtosis and F2 onset and offset of vowels over the sibilants and by testing the speaker-specificity of parameters and sibilants.

We first answer the first research question: are the static and dynamic acoustic features of three sibilants in spontaneous speech as well as the dynamics of fricatives in (inter)vocalic structures a function of speaker? For the different fricatives included in the present study, we found that their acoustic characteristics depend on the individual. The two segments which presented higher inter-speaker variability –/s, ʃ/– were also the ones including at least two parameters that were not significantly affected by speaker: skewness and kurtosis. This might be due to the fact that both parameters were highly influenced by vowels. It was only for /z/ that all measurements were found to be highly significant for the effect of speaker. These results regarding static measures for the three segments are consistent with those of Kavanagh C [7] for they also showed variation between speakers in the static measures of /s/. As for dynamic measures, the segment showing higher inter-speaker variation was /s/ both at onset and offset. Conversely, /z/ showed the least variation between speakers at offset, that is, at the right vowel context. However, /ʃ/ showed a similar trend as /s/ at offset but it missed values at onset due to the lack of tokens in this position.

One of the major sources of acoustic variability in /s/ mentioned in the literature is differences in vocal anatomy [20]. This claim coincides with the results obtained from the dynamic measurements, since formants of /s/ preceded and followed by a vowel prove to be highly significant and therefore speaker-specific, contrary to Kavanagh C [7] who found that dynamic measurements of /s/ did not vary between speakers as much as she expected.

These findings also show that /ʃ/ is the segment in which parameters happen to be more speaker-specific than /z/, despite having two parameters at two different time windows that are not significant. If we pay attention to the literature regarding the production of /ʃ/, this sound is produced with a large portion of the blade of the tongue that rises forming a narrow channel with both the alveolar ridge and the front of the hard palate. The fact that /ʃ/ needs the interaction between a large portion of the tongue and two sections of the vocal tract implies that the differences in anatomy will affect the production of the segment significantly more than the other two sibilants under investigation. In addition to anatomical features, the way those organs move and interact with each other have an effect on speaker-variability [21]. For instance, the palato-alveolar sound might be produced further to the back or further to the front, that is, it can be more palatal than alveolar and vice versa. This depends on how vocal organs move in the oral cavity, which may depend on the speaker.

Most of the acoustic characteristics of the sibilants analysed for this study were shown to depend on the speaker. Among them, intensity, CoG and SD were the most speaker-specific parameters with the highest F-ratio values. On the contrary, skewness and kurtosis were not that significant for certain segments.

Regarding skewness, our results show the greatest positive skewness for /z/, followed by /s/ and /ʃ/. This situation coincides with the report of Tomiak, Avery and Liss; they obtained a greater positive skewness for /s/ than for /ʃ/.

Conversely, Jongman A, et al [12] found negative skewness for /s/ and positive skewness for /ʃ/ like some others did too.

Regarding kurtosis, the highest F-ratio was found at 75% of /z/, being kurtosis of /z/ the one with the highest inter-speaker variation at all-time windows measured. Similar to results of skewness, the fact that /z/ and /s/ present greater positive values than /ʃ/ agrees with the literature. Since these two segments present a more peaked energy distribution than /ʃ/. Furthermore, these findings are in the line of Kavanagh C [7] since she also found skewness and kurtosis to produce the lowest F-ratio values for /s/ despite some of them being significant for speaker [22-28].

The highest inter-speaker variation for CoG+F2 was found at the 50% of /s/ at onset. Contrary to, F2 transition properties were found to be significant for all speakers in each segment according to the ANOVA analysis, except for /ʃ/ at onset and /z/ at offset. Participants showed a speaker-specificity in the way vowels' F2 pulled down CoG at onset or offset.

As for the second aim and as expected, vowels did affect the onset and offset of consonants and thus segments behave differently depending on the context: the lower the vowel's F2, the lower the consonant's CoG; the higher the vowel's F2, the higher the consonant's CoG as it is shown in Figures 4-7. It is indeed expected to find an effect of vowel on /s, z, ʃ/ since the mean of CoG is significantly higher than the F2 of the vowels preceding and following them. This means that a vowel's F2 pulls down the CoG of sibilants both at onset and offset. Furthermore, results show speaker-specificity of CoG when taking into account surrounding vowels meaning that not only anatomy of the vocal tract has an effect on the production of sounds, but also the way the organs move from one speech sound to another are speaker-specific [29-32].

Static and dynamic properties have been analysed and they have shown promising results. With regards to the third aim –intraspeaker variation is smaller than interspeaker variation–, F-ratio has proved to be a perfect measurement to confirm this hypothesis. In fact, the vast majority of the parameters showed a considerably high F-ratio value (between 2.5 and 10) with the exception of skewness and kurtosis of /s, ʃ/. It can be assumed then that intensity, CoG and SD present more interspeaker variation than intraspeaker variation. Skewness and kurtosis are the parameters that might pose more problems to the field of FSC since they show F-ratios close to 1.0 for the sibilants /s, ʃ/, meaning that the difference between inter- and intraspeaker variation is not that big. This is further supported by the information provided by range. Speakers showing a wide range of production of a token are considered to present high intra-speaker variation, which is not particularly good for the research since one cannot cue speaker high a high degree of certainty. However, the cases where range was smaller or located somewhere else in the boxplot –at higher or lower frequencies– are noteworthy since they demonstrate the small within-speaker variability and, therefore, the consistency of the results obtained from the ANOVA analysis.

As for the parameters presenting less intra-speaker variation, skewness stands out because some of the participants only produced positive results, meaning that they could be highlighted among different speakers from different recordings. Kurtosis tends to show small ranges and thus less intra-speaker variability. Yet it did not show high inter-speaker variability either. As for CoG, it showed similar results since for the three segments, there were speakers showing smaller ranges than other but CoG was located a similar Hz for many of them. Finally, SD is a parameter that should be analysed carefully since the correlation results of the three segments varied. The correlation between them, particularly between SD50 and SD100, proved to be slightly weak as mentioned in the introduction of this section. Normally, the three measures are highly correlated due to how close they are from each other, but in this case, we could assume one side and the other are highly influenced by the vowels surrounding them and, henceforth, the weak correlation.

Regarding the fourth of the research aims –to give account of speaker-specific features that can be used in FSC casework–, we coincide with Kavanagh C [7]) in highlighting CoG and SD as the parameters that turned out to be the most speaker-specific. Intensity proved to be a reliable parameter to use in controlled speech. In case of using it to analyse spontaneous speech, data should be normalized to avoid the differences in the recording conditions. Skewness and kurtosis are found to be again in the line of Kavanagh's C [7] results since they do not show such reliable speaker-specificity. Nonetheless, both parameters have shown greater inter-speaker variability for /z/. As for CoG+F2, more data and studies are needed to confirm whether it is a good measure to use in FSC casework or not. Besides, right vowel context for /z/ and left vowel context for /ʃ/ should be analyzed from corpora with more tokens to be statistically significant. Yet, CoG+F2 has indeed shown inter-speaker variability for /s/ and significant effect on the speaker on both vowel contexts, so this could be a start for further research. Therefore, all the parameters might be incorporated in a set of acoustic measures for FSC paying careful attention to skewness and kurtosis, which could be used only for /z/.

Lastly, the final aim of this research was to suggest a detailed methodology that could be replicated. Nevertheless, the methodology suggested has been decided after analysing the previous literature and that implies that advantages and drawbacks of other studies' methodologies have been spotted. It is of important to be meticulous from the very beginning since the transcription and annotation of the segments determine obtaining good results. It is also important to write a script for Praat that can properly obtain the measurements one needs and this should be done with the help of experts in order to avoid problems when checking the data collected. Besides, the compilation of surrounding vowels in order to analyse dynamic transitions from vowel to consonant and vice versa has been added.

Limitations

As it was mentioned in section 3.1., data used for this study was gathered by researchers of the University of Cambridge. Some students volunteered to be recorded pretending they were being interrogated by the police. This fact –although useful for research– decreases the validity of the results since it was simulated spontaneous speech, instead of natural spontaneous speech as it would be found in real forensic casework. Furthermore, there is only one recording for each speaker made in one single session, meaning that there is no non-contemporaneous data for each speaker. In FSC, researchers usually work with data recorded at different moments of the life of the speakers than can vary from a two-year-ago recording to the one made at the police station the very same day of the analysis. In addition, these recordings show a clean spectrum, meaning that there is little background noise when working with them. However, for example the ones recorded in situ by secret agents on the streets or somewhere else usually have background noise that makes the spectrum harder to analyses.

As for other limitations of this research, it would have been ideal to have tokens of /ʒ/ so that a comparison between voiced and voiceless sibilants could have been made too. Being able to analyses the voiced palato-alveolar consonant could have added more insight on the relevance of this sounds for FSC. Had we had tokens preceding the onset of /ʃ/ in the dynamic section could have had helped broaden once again the scope of this research. Yet, that can be taken into account for further research.

This study, however, intends to be an initial research that investigates acoustic parameters of one of the most speaker-specific consonants as it was mentioned in the introduction. We are well-aware of the limitations, but we hope the results of it can shed some light on the field of FSS and promote further research regarding other speech sounds that might contain speaker-specific information.

Conclusion

This article has shown that spectral and acoustic properties of the three sibilants analyzed /s, z, ʃ/ in English present promising results regarding speaker-specificity. In addition, not only the segments themselves, but also the transitions from and towards vowels are particularly speaker-dependent. This fact indicates that both static and dynamic properties should be taken into account in FSC for they reflect differences in individual variation in the articulatory trajectories followed to produce sounds and in the differences in speaker's vocal anatomy.

This research points out the high speaker-specificity of certain parameters of the three consonant segments. Perhaps the least speaker-specific parameters are skewness and kurtosis (except for /z/). Nonetheless, intensity, CoG and SD have proven to be parameters that can be used to discriminate speakers. Due to the promising results shown by these consonants and the parameters analyzed as well as the fact that sibilants are easy to segment in recorded speech, this kind of analysis may be included in FSC set of acoustic features. As for the segment that entails the most speaker-specificity, /ʃ/ appears to be the one. However, it is /z/ the only one in which parameter is significant, /s/ remains a speaker-discriminating segment, particularly when paying attention to F2 transitions from vowels affecting CoG.

I would also like to highlight the opportunities that this research poses for future research. First of all, researchers should incorporate new materials into the study, that is, they should intend to work with real recordings in order to check whether these parameters and segments can be indeed used for FSC casework no matter the conditions of the recordings. In fact, they could also analyses consonants or vowels of telephone recordings since the band is reduced and some sounds (e.g. sibilants) may lose information. Secondly, the analysis of these parameters and similar ones should be extended to other consonant segments paying special attention to the dynamics of them since it has been proven (e.g. Kavanagh, 2012) that they contain a lot of speaker-specific information. Lastly, I would suggest expanding this research to other languages. Research in this field has been mainly done in English, but there are other growing labs that are working with FSC and could use literature supporting their research on FSC casework.

To conclude, this research has shown that acoustic properties of sibilants contain speaker-specific information that can be used to discriminate between individuals. These pages have highlighted that there are many parameters than can be used in real forensic casework and research thereof can be expanded to other consonants or even the same but in different languages.

References

1. Stevens, Kenneth N, Carl E Williams, Jaime R Carbonell, and Barbara Woods. "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material." *J Acoust Soc Am* 44(1968): 1596-1607.
2. Jongman, Allard. Duration of frication noise required for identification of English fricatives. *J Acoust Soc Am* (1989):1718-1725.
3. McDougall, Kirsty and Francis Nolan. "Discrimination of speakers using the formant dynamics of/u:/in British English." *In* Proceedings of the International Congress of Phonetic Sciences (2007)1825-1828.
4. Hussain, Qandeel, Michael Proctor, Mark Harvey, and Katherine Demuth. "Acoustic characteristics of Punjabi retroflex and dental stops." *J Acoust Soc Am 141*(2017): 4522-4542.
5. Balise, Raymond R, Randy L Diehl. "Some distributional facts about fricatives and a perceptual explanation." *Phonetica* 51(1994): 99-110.
6. Shosted, Ryan. "Acoustic characteristics of Swedish dorsal fricatives." *J Acoust Soc Am 123*(2008): 3888.
7. Kavanagh, Colleen. "New consonantal acoustic parameters for forensic speaker comparison." PhD diss., University of York (2012).
8. Coulthard, Malcolm. "Author identification, idiolect, and linguistic uniqueness." *Applied linguistics* 25(2004): 431-447.
9. Rose, Phil, Takashi Osanai, and Y. Kinoshitaa. "Strength of forensic speaker identification evidence: Multi-speaker formant-and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold." *Forensic Linguistics 10*(2003): 179-202.
10. Maddieson,Ian, Sandra Ferrari Disner. Patterns of sounds. Cambridge University Press. 1984.
11. Stuart-Smith, Jane, Claire Timmins, and Alan Wrench. "Sex and gender differences in Glaswegian/s."In Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona (2003)1851-1854.
12. Jongman, Allard, Wayland R and Wong S. Acoustic characteristics of English fricatives. *J Acoust Soc Am* 3(2000): 1252-1263.
13. Munson, Benjamin. "Variability in/s/production in children and adults." (2004).
14. Haley, Katarina L, Elizabeth Seelinger, Kerry Callahan Mandulak and David J Zajac. "Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach." *J Phon* 38(2010): 548-554.
15. Nolan, Francis, McDougall K, de Jong K, and Hudson T. Dynamic variability in speech. A forensic phonetic study of British English. (2009).
16. Delattre,Pierre C, Alvin M Liberman, and Franklin S Cooper. "Acoustic loci and transitional cues for consonants." *J Acoust Soc Am* 27(1955): 769-773.
17. Sussman, Harvey M. "The phonological reality of locus equations across manner class distinctions: Preliminary observations." *Phonetica* 51(1994): 119-131.
18. Wilde, Lorin. "Inferring articulatory movements from acoustic properties at fricative-vowel boundaries." *J Acoust Soc Am* 94(1993): 1881-1881.
19. Yeou, Mohamed. "Locus equations and the degree of coarticulation of Arabic consonants." *Phonetica* 54(1997): 187-202.
20. Hughes, George W and Morris Halle. "Spectral properties of fricative consonants." *J Acoust Soc Am* 28(1956): 303-310.
21. Cristea, Dan, Daniela Gîfu, Mihai-Alex Moruz and Mihaela Onofrei, et al. "An Insight into the Corpus of Contemporary Romanian." Memoirs of the Scientific Sections of the Romanian Academy 40(2017).
22. Fry D. Cambridge: Cambridge University Press. The Physics of Speech (1979).
23. Glass, James R and Victor W Zue. "Acoustic characteristics of nasal consonants in American English." *J Acoust Soc Am 76*(1984): S15-S15.
24. Lisker, Leigh. "The pursuit of invariance in speech signals." *J Acoust Soc Am 77*(1985): 1199-1202.
25. Nirgianaki, Elina. "Acoustic characteristics of Greek fricatives." *J Acoust Soc Am 135*(2014): 2964-2976.
26. Ogden, Richard. Introduction to English Phonetics. Edinburgh University Press (2017).
27. Pouplier, Marianne, Philip Hoole. "Articulatory and acoustic characteristics of German fricative clusters." *Phonetica* 73(2016): 52-78.
28. Stevens, Kenneth N. "Sources of inter-and intra-speaker variability in the acoustic properties of speech sounds." *In Proceedings of the seventh International Congress of Phonetic Sciences/Actes du Septième Congrès international des sciences phonétiques* (2017): 206-232.
29. Stuart-Smith, Jane. "Empirical evidence for gendered speech production:/s/in Glaswegian." (2007): 65-86.
30. Tabain, Marija. "Non-sibilant fricatives in English: Spectral information above 10 kHz." *Phonetica* 55(1998): 107-130.
31. Wolf, Jared J. "Efficient acoustic parameters for speaker recognition." *J Acoust Soc Am* 51(1972): 2044-2056.
32. Zue, Victor W. "Spectral characteristics of English stops in prestressed position." *J Acoust Soc Am* 59(1976): 71-71.

How to cite this article: Carralero-Fernandez, Sara. "Speaker-specific Information in the Acoustic Characteristics of English Fricatives" J Forensic Res 12 (2021): 478