IJFL (International Journal of Forensic Linguistic)
Vol. 3, No. 1 April 2022, Page 105-115

P-ISSN: 2723-1542
E-ISSN: 2723-5254
Available Online at https://www.ejournal.warmadewa.ac.id/index.php/ijfl/index

SPEAKER SPECIFIC INFORMATION IN THE ACOUSTIC CHARACTERISTICS OF ENGLISH FRICATIVES

Sara Carralero-Fernandez
. *Universidad Complutense de Madrid*
scarrale@ucm.es

How to cite (in APA style):

Fernandez, S. C. (2022). Speaker Specific Information in the Acoustic Characteristics of English Fricatives. *IJFL (International Journal of Forensic Linguistic)*, Vol. 3 (1), 105-115. Doi: https://doi.org/10.22225/ijfl.3.1.4449.105-115

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: forensic speaker comparison, speaker-specificity, sibilants.

.

I. 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 (e.g. Stevens, 1971; Wolf, 1972). However, recent investigations have started to include 'dynamic' features of speech (Jongman *et al.*, 2000; Kavanagh, 2012). 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 (McDougall and Nolan, 2007).

Many studies have analysed the spectral characteristics of consonants and vowels (e.g. Glass and Zue, 1984; Zue, 1976; Hussain et al., 2017) and a number of studies have investigated the acoustic and spectral characteristics of fricatives in different dialects of English (Balise & Diehl, 1994; Hughes & Halle, 1956; Jongman, 1989; Tabain, 1998) and other languages such as Swedish (Shosted, 2008), German (Pouplier and Hoole, 2016) and Greek (Nirgianaki, 2014). Yet, there are only few studies that have investigated fricatives in English and their dynamic features in different contexts for forensic purposes (Kavanagh, 2012). 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 (Kavanagh, 2012). This changes the values of the acoustic features connected to the cavities that have been altered (Kavanagh, 2012). Besides, according to French et al. (2010), 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 Kavanagh (2012) 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" (Coulthard, 2004:431-432). In fact, no one is able to repeat the exact same realisation of an utterance twice (Rose, 2002). 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 analysing 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 analyse segments of simulated spontaneous speech to contribute to the field by analysing 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 (2012) research and contribute to current findings by analysing 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 analysed consonants can be used as speaker-specific features independently of when they are being analysed (Jongman *et al.*, 2000). 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, e.g. McDougall and Nolan (2007) analysed how /u:/ varied depending on the context. The onset and offset of vowels or neighbouring consonants will very likely affect the production of the sibilants too, that is, the spectral peak might be higher or lower, for instance.

II. METHODS

The DyViS Database is a large-scale, forensically-oriented speech corpus (Nolan *et al.*, 2009). 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 (Nolan *et al.*, 2009: 40). 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 (Nolan *et al.*, 2009). 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). For each of the three segments seven acoustic features were analysed, both dynamic and static properties. Following Jongman et al. (2000) and Kavanagh (2012), 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 neighbouring 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%, 100% of the vowels’ duration). In the following result section, parameters and their windows, would be indicated as follows: *parameter + number*. The researcher has 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. 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 (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 of sibilants is a measure of the concentration of energy in the spectrum (Kavanagh, 2012). This parameter, also known as *mean*, shows the frequency at which the distribution of the energy in the spectrum is even on either side. 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 (Stuart-Smith *et al.*, 2003). SD is calculated by measuring the square root of the second spectral moment, also known as *variance* (Kavanagh, 2012). 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 measure dynamic properties of speech sounds, since they relate points in the speech signal to F2. “Locus” was first defined by Delattre *et al.* (1995:769) as “a place on the frequency scale at which a transition begins or to which it may be assumed to point”. According to Sussman *et al.* (1991:1311), 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 (1963), 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, the are some that have been carried out so far (Wilde, 1993; Sussman, 1994; Yeou, 1997) in which F2 locus has been measured. These studies are contradictory: some of them (e.g. Yeou, 1997) obtained good classifications of fricatives, while others (Wilde, 1993; Sussman, 1994) showed results in which the fricatives' loci were overlapping. It is 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" (Jongman *et al.*, 2000: 1256). 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. 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 (Kavanagh, 2012). 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 (Jongman *et al.*, 2000). Kurtosis is the fourth spectral moment which measures how raised or flat the distribution of the energy is. According to Jongman *et al.* (2000), 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 (Kavanagh, 2012). 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. 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). 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.

III. RESULT

1. 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%.

2. Results: /z/

The different dependent variables that have been analysed 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.

3. 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, $p = .004$.

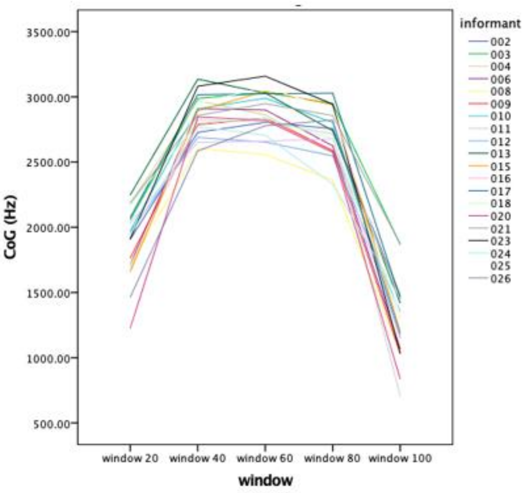
Overall, the speaker was found to be a highly significant factor ($p < .05$) in intensity, centre of gravity and standard deviation at the 50%, 75% and 100%. As for skewness, it was significant at the 75% and 100% of the segment's duration, whereas for kurtosis it was significant at the 50% and 100%.

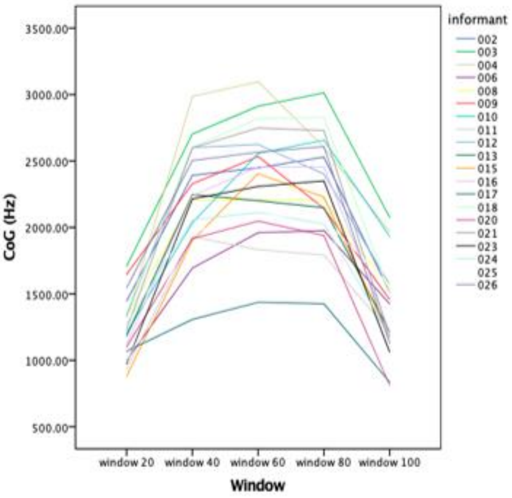
4. Dynamic measures

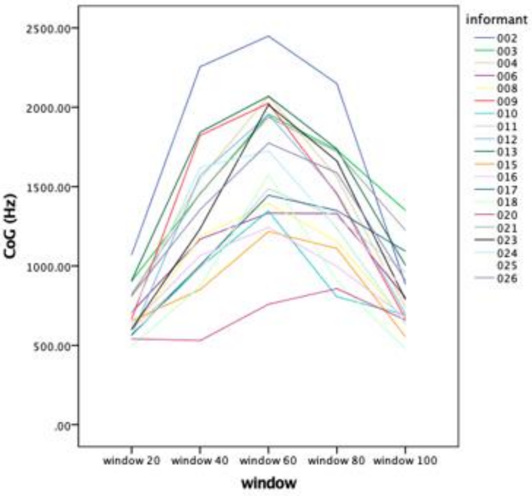
Apart from the static measures, CoG was measured in five smaller time windows to capture dynamic movements in the spectrum over time, similar to Kavanagh's (2012). Means of each window for each speaker are displayed in Figures 1-3.

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, 2 and 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.

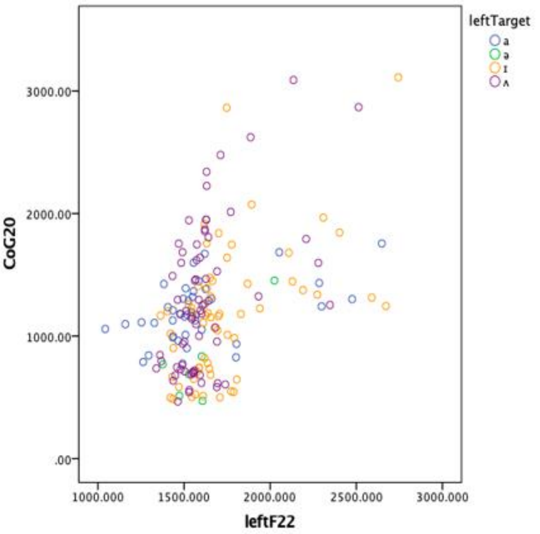
ANCOVA results for CoG at onset and offset proved to be highly significant for speaker, with the highest *F*-ratio overall for /s/ ($F(19, 153) = 3.341$, $p < .001$) at onset and ($F(19, 94) = 2.575$, $p < .001$) for onset. Hence, a main effect of speaker on CoG considering both left and right vowel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

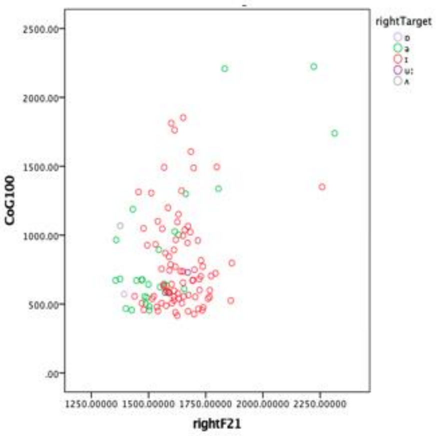


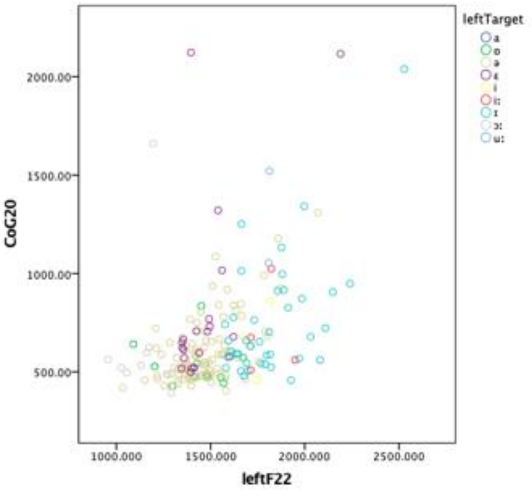




As for the frequency of appearance of vowels preceding /s/, that is, left vowel context, it is as follows (Figure 4-5): /ɪ, ʌ, a, ə/. The frequency of appearance of the vowels following /s/ –right vowel context– is /ɪ, ə/ in the first place and then there are a few tokens of /ɒ, ʌ, uː/. As it has been 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, 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.

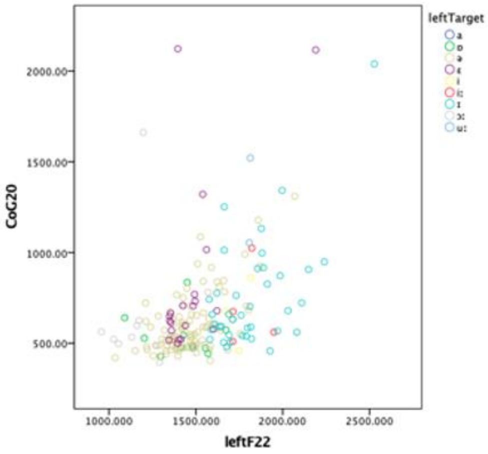






ANCOVA results for CoG at onset were significant for speaker, with the highest *F*-ratio overall for /z/ (*F*(19, 164) = 2.199, *p* = .004). However, they were not significant at offset (*F*(15, 29) = .618, *p* = .836), meaning that in the case of /z/, CoG is not particularly influenced by following 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 (Figure 6-7): /ə, ɪ, ɛ, ɔː, iː/. 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.

Based on the discussion above, it can be concluded that Digital voice forensic techniques cannot validate evidence because no standard validation is determined. Digital forensic techniques can only provide sound similarity analysis of good evidence by the suspect's voice. This shows that everyone has a different pitch value because each person's word pronunciation is different. There may be pitch values from several subjects that are almost the same.



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 (2012) 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 (e.g. Hughes & Halle, 1956; Fry, 1979; Stevens, 1968 or Kavanagh, 2012). 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 (2012) 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 (Collin, 2003). 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. 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 (1990) and Avery and Liss (1996); they obtained a greater positive skewness for /s/ than for /ʃ/. Conversely, Jongman *et al.* (2000) found negative skewness for /s/ and positive skewness for /ʃ/ like some others did too (e.g. Nittouer, 1995; Farland *et al.*, 1996).

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 (e.g. Jongman *et al.*, 2000) since these two segments present a more peaked energy distribution than /ʃ/. Furthermore, these findings are in the line of Kavanagh (2012) since she also found skewness and kurtosis to produce the lowest *F*-ratio values for /s/ despite some of them being significant for speaker.

The highest inter-speaker variation for CoG+F2 was found at the 50% of /s/ at onset. Contrary to Jongman *et al.* (2000), 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.

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 tend 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 (2012) 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 normalised to avoid the differences in the recording conditions. Skewness and kurtosis are found to be again in the line of Kavanagh (2012)'s 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 analysed 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.

IV. CONCLUSION

This article has shown that spectral and acoustic properties of the three sibilants analysed /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 analysed 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. 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

Balise, R., & Diehl, R. (1994). Some Distributional Facts about Fricatives and a Perceptual Explanation. *Phonetica, 51*(1–3), 99-110.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.35, retrieved 24 October 2017 from http://www.praat.org/

Coulthard, M. (2004). Author Identification, Idiolect and Linguistic Uniqueness. *Applied linguistics, 25*(4), 431-447.

Delattre, P.C., Liberman, A.M. & Cooper, F.S. (1995). Acoustic loci and transitional cues for consonants. *Journal of Acoustic Society of America*. *27*, 769-773.

Fry, D. (1979). *The Physics of Speech*. Cambridge: Cambridge University Press.

Glass, J., & Zue, V.W. (1984). Acoustic characteristics of nasal consonants in American English. *The Journal of the Acoustical Society of America,76*(S1), S15.

Haley, K.L., Seelinger, E., Callahan Mandulak, K. & Zajac, D.J. (2010). Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics, 38*(4), 548-554.

Hughes, G., & Halle, M. (1956). Spectral Properties of Fricative Consonants. *The Journal of the Acoustical Society of America, 28*(2), 303-310.

Hussain, Q., Proctor, M., Harvey, M., & Demuth, K. (2017). Acoustic characteristics of Punjabi retroflex and dental stops. *Journal of The Acoustical Society of America,141*(6), 4522-4542.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *The Journal of the Acoustical Society of America, 85*(4), 1718-1725.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*(3), 1252-1263.

Kavanagh, C. (2012). *New Consonantal Acoustic Parameters for Forensic Speaker Comparison,* PQDT - UK & Ireland.

Lindblom, B. (1963). The pursuit of invariance in speech signals. *Journal of Acoustic Society of America. 77*, 1199-1202.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

McDougall K. & Nolan, F. (2007) Discrimination of speakers using the formant dynamics of /u:/ in British English. In J. Trouvain and W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 6-10 August 2007, Saarbrücken, (pp. 1825-1828).

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research, 47*(1), 58-69.

Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *The Journal of the Acoustical Society of America, 135*(5), 2964-2976.

Nolan F., McDougall K., de Jong K., & Hudson T. (2009). *Dynamic variability in speech. A forensic phonetic study of British English*. http://www.ling.cam.ac.uk/dyvis/index.html

Ogden, R. (2009). *An introduction to English phonetics* (Edinburgh Textbooks on the English Language). Edinburgh: Edinburgh University Press.

Pouplier, M. & Hoole, P. (2016). Articulatory and Acoustic Characteristics of German Fricative Clusters. *Phonetica. 73*(1), 52-78.

Rose, P. (2003). *Forensic Speaker Identification.* New York: Taylor&Francis.

Shosted, R. (2008). Acoustic characteristics of Swedish dorsal fricatives. *The Journal of the Acoustical Society of America, 123*(5), 3888.

Stevens, K. N., Williams C.E., Carbonell, J.R. & Woods, B. (1968). Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America*, *44*, 1596–1607.

Stevens, K.N. (1971). Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings 7th ICPhS*, Montreal, 206-232.

Stuart-Smith, J. (2007). “Empirical evidence for gendered speech production: /s/ in Glaswegian”. In J. Cole, & J. I. Hualde (Eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter (pp. 65-86).

Stuart-Smith, J., Timmins, C., & Wrench, A. (2003). Sex and gender differences in Glaswegian /s/. *Proceedings 15th ICPhS*, Barcelona, 1851-1854.

Sussman, H. M. (1994). The phonological reality of locus equations across manner class distinctions: Preliminary observations, *Phonetica*, *51*, 119–131.

Tabain, M. (1998). Non-Sibilant Fricatives in English: Spectral Information above 10 kHz. *Phonetica, 55*(3), 107-130.

Yeou, M. (1997). Locus equations and the degree of coarticulation of Arabic consonants, *Phonetica,* 54, 187–202.

Wilde, L. (1993). Inferring articulatory movements from acoustic properties at fricative vowel boundaries, *The Journal of the Acoustical Society of America*. *94*, 1881.

Wolf, J. (1972). Efficient Acoustic Parameters for Speaker Recognition. *The Journal of the Acoustical Society of America, 51*(6B), 2044-2056.

Zue, V. (1976). Spectral characteristics of English stops in pre-stressed position. *The Journal of the Acoustical Society of America, 59*(S1), S71.