



**HAL**  
open science

## Intra- and inter-speaker variation in eight Russian fricatives

Natalja Ulrich, François Pellegrino, Marc Allasonnière-Tang

► **To cite this version:**

Natalja Ulrich, François Pellegrino, Marc Allasonnière-Tang. Intra- and inter-speaker variation in eight Russian fricatives. *Journal of the Acoustical Society of America*, 2023, 153 (4), pp.2285-2297. 10.1121/10.0017827 . hal-04071781

**HAL Id: hal-04071781**

**<https://hal.science/hal-04071781>**

Submitted on 7 Nov 2023



**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Intra- and inter-speaker variation in eight Russian fricatives<sup>a)</sup>

Natalja Ulrich,<sup>1,b)</sup> François Pellegrino,<sup>1</sup>  and Marc Allasonnière-Tang<sup>2</sup> 

<sup>1</sup>Laboratoire Dynamique Du Langage (DDL) UMR 5596, CNRS/Université Lyon 2, Lyon, France

<sup>2</sup>Lab Ecological-Anthropology, Unité Mixte de Recherche 7206, National Museum of Natural History, Paris, France

### ABSTRACT:

Acoustic variation is central to the study of speaker characterization. In this respect, specific phonemic classes such as vowels have been particularly studied, compared to fricatives. Fricatives exhibit important aperiodic energy, which can extend over a high-frequency range beyond that conventionally considered in phonetic analyses, often limited up to 12 kHz. We adopt here an extended frequency range up to 20.05 kHz to study a corpus of 15 812 fricatives produced by 59 speakers in Russian, a language offering a rich inventory of fricatives. We extracted two sets of parameters: the first is composed of 11 parameters derived from the frequency spectrum and duration (acoustic set) while the second is composed of 13 mel frequency cepstral coefficients (MFCCs). As a first step, we implemented machine learning methods to evaluate the potential of each set to predict gender and speaker identity. We show that gender can be predicted with a good performance by the acoustic set and even more so by MFCCs (accuracy of 0.72 and 0.88, respectively). MFCCs also predict individuals to some extent (accuracy = 0.64) unlike the acoustic set. In a second step, we provide a detailed analysis of the observed intra- and inter-speaker acoustic variation.

© 2023 Acoustical Society of America. <https://doi.org/10.1121/10.0017827>

(Received 11 August 2022; revised 21 March 2023; accepted 26 March 2023; published online 17 April 2023)

[Editor: Ewa Jacewicz]

Pages: 2285–2297

### I. INTRODUCTION

The study of inter- and intra-speaker acoustic variation and speaker-specific characteristics in speech segments is important in phonetics and automatic speaker recognition, among other fields in language sciences. This acoustic information can be perceived by human listeners and automatically exploited, as has been demonstrated for vowels (McDougall and Nolan, 2007; Rose, 2007), nasals (Enzinger and Balazs, 2011; Kavanagh, 2012), and fricatives (e.g., Schwartz, 1968; Spinu *et al.*, 2018). It has further been suggested that speech segments with predominantly periodic energy, such as vowels, offer a better potential for speaker characterization than speech segments with a high degree of aperiodic energy, such as fricatives, which consist of either only aperiodic energy (voiceless fricatives) or the combination of periodic and aperiodic components (voiced fricatives). They are additionally characterized by aperiodic energy that extends in much higher frequency ranges than most other speech segments (Stevens, 1960). Yet, research on spectral aspects of fricatives is mostly limited to frequencies below 12 kHz, or even 8 kHz, thus ignoring the information encoded in higher frequencies (Forrest *et al.*, 1988; Gordon *et al.*, 2002; Jongman *et al.*, 2000; Kavanagh, 2011; Kochetov, 2017). Moreover, these analyses are generally limited to voiceless fricatives and, in particular, the alveolar [s],

making the insight into the acoustic variation among fricatives restricted to a few phonemes in a few languages.

We consider that the study of fricatives has not received sufficient attention, partly due to the limited segments investigated and the adequacy of the spectral analyses applied to them. The main objective of this article is, therefore, to characterize the acoustic variability present in fricatives and their potential in individual speaker recognition, while overcoming the aforementioned limitations. Our study focuses on Russian, a language with quite a large inventory of fricatives. For this, first, we adopt an extended frequency range up to 20.05 kHz for spectral characterization and second, we consider eight voiced or voiceless fricatives, thus providing a comprehensive overview of acoustic variation in Russian fricatives extending to high frequencies. This study, furthermore, compares two sets of parameters. The first is composed of 11 parameters derived from the frequency spectrum and duration (acoustic set) while the second consists of 13 mel frequency cepstral coefficients (MFCCs). We implemented machine learning (ML) methods to evaluate the potential of each set to predict gender<sup>1</sup> and speaker identity and providing as well a detailed analysis of the observed intra- and inter-speaker acoustic variation. This paper thus contributes to the acoustic description of high-frequency speech segments and more generally to the exploration of the high end of the speech spectrum, two goals emphasized by the guest editors of this special issue.

The article starts with an overview of the existing literature on gender and speaker variation in fricatives in Sec. II. The current research is also introduced in the same section.

<sup>a)</sup>This paper is part of a special issue on Perception and Production of Sounds in the High-Frequency Range of Human Speech.

<sup>b)</sup>Electronic mail: [ulrichnatalja@gmail.com](mailto:ulrichnatalja@gmail.com)

Section III describes the process of data collection, the dataset, and the methods used for the extraction of acoustic parameters as well as the data analyses. In Sec. IV we present the findings, followed by a discussion on gender and inter- and intra-speaker variation in Sec. V, which concludes the article.

## II. GENDER AND SPEAKER VARIATION IN FRICATIVES

A number of studies focusing on inter- and intra-speaker variation have identified a set of spectral, temporal, and amplitude parameters that contribute to the distinction between gender categories or individual speakers (Gordon *et al.*, 2002; Hughes and Halle, 1956; Jongman *et al.*, 2000; Kavanagh, 2011; Narayanan *et al.*, 1995; Newman *et al.*, 2001; Silbert and de Jong, 2008; Smorenburg and Heeren, 2020; Schindler and Draxler, 2013). As an example, the peak frequency and the spectral moments show speaker-specific patterns and are known to be correlated with the articulatory and anatomical properties of a speaker (Newman *et al.*, 2001; Schindler and Draxler, 2013; Smorenburg and Heeren, 2020).

In this context, the acoustic contrast between female and male speakers has been thoroughly studied, and it is argued to be well understood and explained by physiological and sociophonetic differences [e.g., Jongman *et al.* (2000), Ludger *et al.* (2021), and Munson *et al.* (2006)]. For instance, the production of vowels (Diehl *et al.*, 1996; Weirich and Simpson, 2014) and fricatives (Weirich and Simpson, 2015) by female speakers tends to occupy a larger phonetic space than male speakers. Several studies have reported that females exhibit clearer speech, which has been analyzed in terms of cultural, social, physiological, and perceptual factors [e.g., Eckert (1989), Henton (1995), and Labov (1990)].

In fricatives, studies on acoustic gender variation have reported higher values for female speakers in terms of center of gravity and peak frequency (Flipsen *et al.*, 1999; Gordon *et al.*, 2002; Jongman *et al.*, 2000; Kochetov, 2017; Ludger *et al.*, 2021; Newman *et al.*, 2001; Schwartz, 1968), while spectral skewness has been identified as another parameter where gender variation has been observed (Flipsen *et al.*, 1999; Ludger *et al.*, 2021; Munson *et al.*, 2006; Stuart-Smith, 2007). Differences between female and male speakers were also found in the duration examination of the singleton vs geminate contrast in Lebanese Arabic fricatives (Al-Tamimi and Khattab, 2015). Some of these studies also showed that gender variation differs across fricatives (Gordon *et al.*, 2002; Kochetov, 2017) and languages (Gordon *et al.*, 2002).

Beyond the difference between gender categories, fricatives can also encode speaker-specific patterns (Gordon *et al.*, 2002; Hughes and Halle, 1956; Kavanagh, 2011; Narayanan *et al.*, 1995; Newman *et al.*, 2001; Silbert and de Jong, 2008; Smorenburg and Heeren, 2020), underscoring the need for further investigation (Kavanagh, 2012; Schindler and Draxler, 2013). For instance, the spectral peak frequency in voiceless fricatives was found to be

highly variable among speakers, and one speaker's alveolar peaks can appear as the post-alveolar peak frequencies of another speaker (Hughes and Halle, 1956). The spectral moments also offered reliable acoustic information for speaker discrimination in [f] and [s] (Schindler and Draxler, 2013). In general, the most substantial inter-speaker variability was identified in the spectral shape of the alveolar [s] (Gordon *et al.*, 2002; Kavanagh, 2011, 2012).

Previous studies have mostly aimed to describe acoustic variation in fricatives in terms of gender and/or identity with spectral and temporal parameters whose interpretation is quite intuitive (such as the spectral center of gravity or peak frequency). Other studies also considered the standard parameterization in the field of automatic speech processing with cepstral coefficients on the Mel scale (Antal, 2008; Kong *et al.*, 2014) or the Bark scale (Ghaffarvand Mokari and Mahdinezhad Sardhaei, 2020; Jesus and Jackson, 2008; Lilley *et al.*, 2021; Spinu *et al.*, 2018; Spinu and Lilley, 2016; Spinu *et al.*, 2012). The performance of traditional acoustic parameter sets and cepstral coefficients (CCs) have been compared in predicting the place of articulation, voicing, palatalization contrast, and speakers' gender (Ghaffarvand Mokari and Mahdinezhad Sardhaei, 2020; Jesus and Jackson, 2008; Spinu *et al.*, 2018; Spinu and Lilley, 2016). The findings on gender prediction in Azerbaijani (Ghaffarvand Mokari and Mahdinezhad Sardhaei, 2020), in Romanian (Spinu and Lilley, 2016), and with a subset of Russian fricatives (Spinu *et al.*, 2018) showed that cepstral coefficients clearly outperform traditional acoustic approach, suggesting that gender is better accounted for by a fine-grained encoding of the distribution of energy as in CCs than by more coarse-grained indices captured by traditional acoustic parameter sets.

In speaker recognition experiments, significant differences in speaker discrimination potential were observed between voiced and voiceless segments in general and between fricatives in particular in Arabic consonants (Alsulaiman *et al.*, 2017) while in English, a high recognition rate of speakers was achieved with vowels and also fricatives (Antal, 2008). In contrast, moderate performances with fricatives were reported in French compared to other phoneme classes (Ajili *et al.*, 2017). Another study in French, based on an automatic classifier trained on spectrograms rather than CCs, also found that vowels played a greater role in the identification than fricatives and nasals. The authors concluded that less speaker information is contained in these phoneme classes (Gendrot *et al.*, 2020).

Interestingly, researchers in phonetics as well as in automatic processing have both noted that the discrimination potential can also vary significantly across speakers. For instance, the relative informativeness of temporal and spectral aspects can differ across speakers (Kavanagh, 2012). In another study comparing speaker classification from spectrograms of vowels, the authors suggested that there are some *good speakers* who are well discriminated while others lead to poor results (Gendrot *et al.*, 2019). Alternatively, some other studies claimed that intra-speaker

variability in obstruents is contrast- and/or parameter-specific rather than a general talker characteristic (Harper, 2021; Romeo *et al.*, 2013).

Finally, it is worth noting that the Russian fricative inventory tends to be understudied. The Russian language has at least 12 fricatives whose phonetic characteristics are only partly understood because of the lack of large-scale studies. Several studies examined the place-of-articulation or palatalization contrasts based on the productions of a few speakers but did not specifically consider the across-speaker variation (Kochetov, 2017). Speaker and gender variation have, however, been reported for sibilant palatalized and non-palatalized fricatives produced by ten speakers (Kochetov, 2017; Spinu *et al.*, 2018) as well as some variation in vocal fold vibration for voiced fricatives from eight speakers (Barry, 1995). A recent study on the three Russian fricatives [f], [s], [ʃ], including recordings of forty speakers, concluded that the first two spectral moments are sufficient to classify the place of articulation in these three fricatives. The classification rates reached, thereby, around 98% without any effect of linguistic or non-linguistic parameters such as the vowel context, speaker, or gender (Ulrich *et al.*, 2021).

To summarize, previous studies suggest that even if fricatives are not as indicative as vowels of speaker gender and identity, they encode some discriminant information that is better captured by cepstral coefficients than by traditional acoustic parameters. The number of speakers (or fricatives) considered has, nevertheless, often been limited, and studies that articulate an automatic identification approach with a thorough acoustic analysis of the pattern of within- and across-speaker variation are rare. The present study aims to fill in these gaps by investigating inter- and intra-speaker variation in the acoustic features of voiceless and voiced fricatives in Russian.

### III. METHODS

This section first describes the data collection, annotation, and segmentation processes and techniques. The current dataset and the acoustic analysis are then introduced. Finally, the methods of data visualization and analysis are explained. The dataset, acoustic analysis methods, and ML classifiers are in part comparable to those presented in Ulrich *et al.* (2021). However, the dataset was taken from the same corpus (Ulrich, 2022) but was significantly expanded from three fricatives to eight, and the number of speakers increased from 40 to 59. The goals of the papers are also distinct since our previous study focused rather on identifying the place of the articulation of fricatives than on speaker characteristics. The acoustic analyses are similar, and two of the four classifiers previously implemented are used in this paper to estimate the predictive potential of gender and speaker by the acoustic set and MFCCs.

#### A. Participants, data collection, and segmentation

The participants were 59 students (30 females and 29 males) between 18 and 30 years old, studying at different departments of St. Petersburg University in Russia. They

were born or lived in St. Petersburg since their early childhood. No participants reported any speech or hearing impairment. All participants were first introduced to the purpose of the experiment, the expected duration, and the procedure. They were informed about their right to withdraw at any time during the experiment and provided with the contact details of a person that can answer all their questions concerning the research and their rights. The participants were compensated for their participation.

The recording sessions were conducted at the phonetic laboratory of the Phonetic Institute in St. Petersburg, in an audiometric booth using the recording program SPEECH-RECORDER version 3.28.0 (Draxler and Jänsch, 2022) at a sample rate of 44.1 kHz (16-bit encoding). For the recordings, a clip-on microphone (Sennheiser MKE 2-P) was placed at a distance of 15 cm from the speakers' mouth and connected through an audio interface (Zoom U-22) to a laptop computer.

Demographic data, such as gender and age, were collected before the experiment started. The participants were instructed to read 198 sentences presented in a random order on a computer screen. Two sentence structures were used to obtain each real-word lexemes produced in three different contexts. The first type of sentence is a so-called carrier sentence with the structure of "She said 'X' and not 'Y.'" (RU: [ana skazala salʲ, a nʲ iʃalʲ]). Minimal pairs of real words, for instance [salʲ] and [ʃalʲ], containing one of the 11 tested fricatives were placed in both X and Y positions. The second type of pre-designed sentences is a natural language sentence including each of the lexemes, for instance, "His name is Sasha [salʲ]" and "I like your [ʃalʲ]" (scarf) (RU: [jiivo zavut saʃa [salʲ], mnʲ e nraʋitsa tvaʃa [ʃalʲ]). The distribution of voiceless, voiced, and palatalized fricatives depend on several phonotactic rules [e.g., Bolla (1981) and Timberlake (2004)]. For example, voiceless fricatives can appear at the initial, medial, and word-final positions, while voiced fricatives undergo devoicing at the word-final position. Furthermore, minimal pairs do not exist for all contrastive fricatives. Consequently, different numbers of tokens were recorded for each fricative.

The raw audio files were first automatically pre-processed by applying the web service (online tool) Munich Automatic Segmentation system, MAUS (Kisler *et al.*, 2017; Schiel, 1999) available online (Schiel, 2023). Then, the files were filtered out below 80 Hz and above 20050 Hz with a smoothing of 80 Hz, and the boundaries were manually corrected using PRAAT (Boersma and Weenink, 2022). In order to determine the onset and offset of the full consonant, the broadband spectrogram was considered more informative than the start of an aperiodic waveform with rising zero crossing rates. In intervocalic fricatives, the presence of formant columns (vertical lines in the spectrogram showing glottal pulses during vowel production) was defined as the onset and offset of the fricative [following Machač and Skarnitzl (2013)].<sup>2</sup>

#### B. Dataset and acoustic analysis

The current dataset consists of 15 812 tokens, including the voiceless [f] [s] [ʃ], voiced [v] [z] [ʒ], and palatalized



TABLE I. Token count by fricative. Each speaker produced the same number of tokens for each fricative category.

Fricative	[f]	[s]	[ʃ]	[v]	[z]	[ʒ]	[sʲ]	[ç]
freq	36	67	55	29	27	24	15	15

[sʲ] [ç]<sup>3</sup> fricatives. The token counts differ between the eight fricatives and are given in Table I.<sup>4</sup>

For each token, an acoustic set consisting of 11 acoustic parameters, summarized in Table II, and 13 MFCCs were computed. The spectral peak location and the four spectral moments (center of gravity, spectral spread, skewness, and kurtosis) describe the aperiodic energy distribution. The harmonic-to-noise ratio (HNR) parameters provide an estimation of the relative distribution of periodic vs aperiodic energy. HNR mean and maximum values around zero indicate equal energy in harmonics and noise. A value of 20 indicates 99% of harmonics and 1% of noise in the signal (Boersma and Weenink, 2022). The MFCCs quantify the sound short-term power spectrum derived from a cepstral analysis performed on a non-linear filter bank (Ganchev et al., 2005). All parameters were estimated from the entire duration of the target fricatives using PRAAT (Boersma and Weenink, 2022) and standard settings. The spectral analysis (*peak, cog, sdev, skew, kurt*) was performed on 10 ms non-overlapping windows, and the means computed.

As mentioned previously, fricatives contain spectral energy in higher frequency ranges than other phoneme classes (Stevens, 1960); however, previous studies primarily focused on a frequency level up to 12 kHz or even as low as 8 kHz. To illustrate this point, Fig. 1 provides a sense of how spectral energy is distributed across 10 ms time windows in the current data. It shows that spectral energy frequently extends above 10 kHz, particularly in [f] and [s] and [sʲ], confirming the interest in performing the analysis on an extended spectral range.

TABLE II. Summary of the acoustic parameter set.

Parameter	Variable	Description
Fricative duration	<i>dur</i>	Duration of the entire segment obtained from manual segmentation
Peak frequency	<i>peak</i>	Frequency of the highest amplitude
Spectral mean	<i>cog</i>	Mean value of the distribution of spectral energy (center of gravity)
Spectral variance	<i>sdev</i>	Spectral spread of the energy around the mean
Spectral skewness	<i>skew</i>	Spectral tilt, overall asymmetry of the energy distribution
Spectral kurtosis	<i>kurt</i>	Spectral flatness of the distribution
HNR mean	<i>hmean</i>	The mean of harmonic-to-noise ratio
HNR sd	<i>hsd</i>	Standard deviation of HNR
HNR max	<i>hmax</i>	Maximum of HNR
HNR tmax	<i>htmax</i>	Time to the maximum HNR
Tilt	<i>tilt</i>	Spectral tilt. Computed by H1-H2

### C. Machine learning classifiers, data visualization, and analysis

In the current study, we implemented two classifiers based on binary recursive partitioning (Breiman et al., 1984): the first classifier generates a single *decision tree* (DT) based on the data and helps to visualize the interactions between the variables. The output of this classifier is an explicit decision tree that captures the hierarchical interactions of the variables within the dataset. The second classifier is a “random forest” (RF) (Breiman, 2001). It generates a series of 200 decision trees<sup>5</sup> analyzed as a whole and used to assess the importance of each variable with regard to correctly predicting the fricatives.

In the literature, several methods have been proposed to quantify gender and speaker acoustic variation. The statistical means of spectral and temporal measurements across speakers and phoneme classes were frequently computed and compared (Gordon et al., 2002; Kavanagh, 2011; Newman et al., 2001; Silbert and de Jong, 2008). To capture the variation within each category, the range (Kavanagh, 2011; Silbert and de Jong, 2008), standard-deviation (Newman et al., 2001), and the interquartile range (IQR) (Ferragne and Pellegrino, 2010) were also computed. For the current analysis, first, data visualization methods were used to give an overview of how male and female speakers differ in the distribution of the mean and range values of the 11 acoustic parameters across the eight fricatives. To compare this variation, Wilcoxon tests with Bonferroni multiple testing correction were implemented. Furthermore, the IQR was computed to determine the variation in the ranges of female and male speakers’ values across the 11 acoustic parameters. For each fricative and parameter, the IQR mean over gender categories was estimated and the differences were compared.

We also adopted an approach inspired by previous studies, which found that male and female speakers organize their fricative contrasts differently, as revealed by computing pairwise distances between the fricatives produced by male and female speakers (Weirich and Simpson, 2015). To assess the pairwise distances within the current dataset, two t-SNE (t-distributed stochastic neighbor embedding) representations of the segment tokens were generated (Van der Maaten and Hinton, 2008). The t-SNE method was selected to represent the high-dimensional data of acoustic parameters and MFCCs in two-dimensional spaces. For each representation, the Euclidean distance was computed between all the tokens of two contrastive fricatives in the t-SNE representations. For the comparison of female and male speakers, the measured distances were compared by gender and fricative. More precisely, the distance was computed for fricative pairs contrasted by places of articulation [f-s], [s-ʃ], [v-z], [z-ʒ], [sʲ -ç]; voicing [f-v], [s-z], [ʃ-ʒ]; and palatalization [s-sʲ], [ʃ-ç].

To further visualize the stability and the variation of the acoustic parameters across the fricatives, we generated a principal component analysis (PCA) for each fricative based

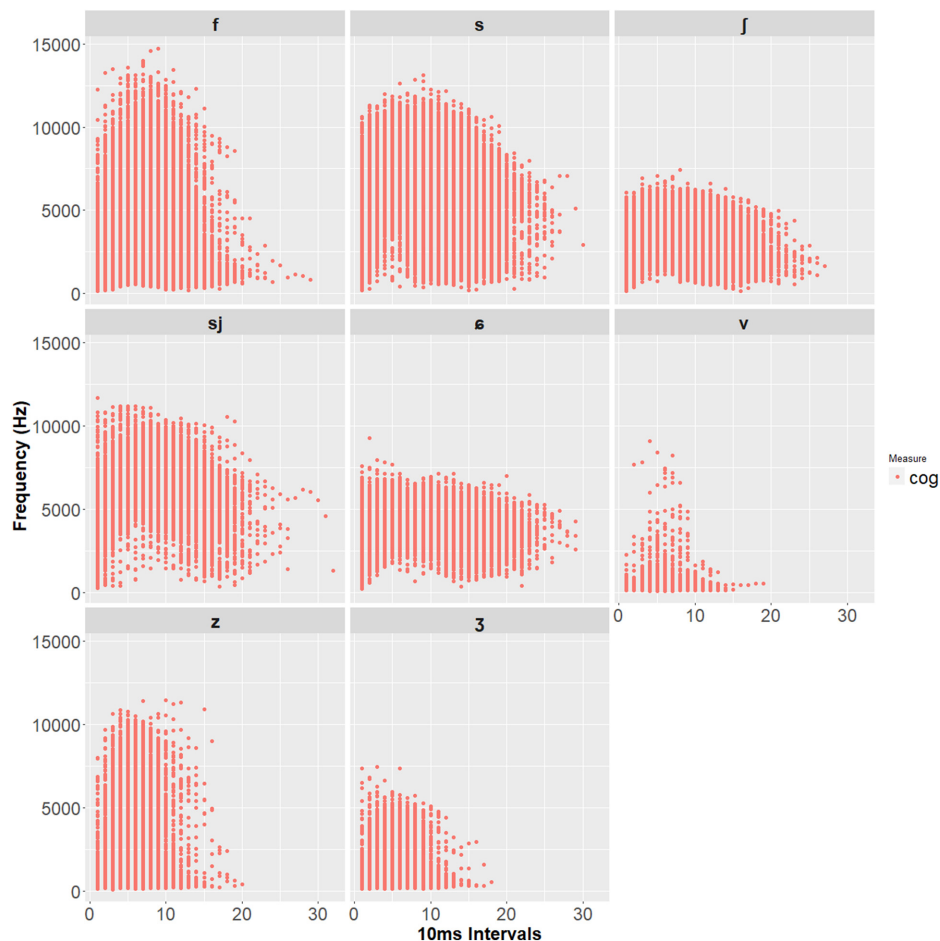


FIG. 1. (Color online) *Cog* (Center of gravity) measured for windows of 10 ms across fricatives for all speakers. The *x* axis represents the time windows and each dot represents one measurement at one-time window. The *y* axis shows the frequency in Hz. The energy extends above 10 kHz in fricatives such as [f] and [s] and [sj].

on the z-scored  $[(x - mean(x))/sd(x)]$  values of each parameter. The contribution of each parameter to the first principal component (PC1) is extracted since, by design, it captures the highest proportion of the variation present in the dataset. Then, to visualize the inter- and intra-speaker variation, we added a parameter, the SD-ratio. For each fricative, it is defined as the ratio between the overall standard deviation and the speakers' standard deviation. This parameter thus quantifies the ratio of inter- and intra-speaker variability. It is derived from a similar index (Schindler and Draxler, 2013). Next, to have an overview of the relation between PC1 and SD-ratio at the fricative level, the two values were compared and tested for correlation. A similar analysis was conducted at the speaker level. Extracting PC1 and SD-ratio by speaker hence gives an indication of which acoustic parameters are the most variable vs the most stable within speakers. A high PC1 value indicates a high variation within a speaker and a low PC1 value indicates a small variation within a speaker, while a high SD-ratio shows a high variation between speakers and a low SD-ratio indicates a low variation between speakers. Finally, to have an even more precise understanding of the variation within and between speakers, the same analysis was conducted with a data subset consisting of two fricatives ([f] and [ʃ]) produced by three female speakers, which were selected because they exhibited quite different patterns of variation.

#### IV. RESULTS

The main results of the data analysis are reported here and additional details and code can be found in supplementary materials.<sup>6</sup> The section starts with an overview of the performance in predicting the speaker's gender and identifying speakers with two ML classifiers. Then additional observations of the intra- and inter-speaker variation are presented.

##### A. Predictive power for gender and speaker

The output of the ML task predicting speakers' gender by the traditional set of acoustic parameters and MFCCs are summarized in Table III. Gender can be predicted by acoustic parameters with moderate accuracy, while the classifiers trained with MFCCs achieve better accuracy. An analysis of the importance of the variables (provided in supplementary materials)<sup>6</sup> indicates that the most relevant variables are *peak*, *cog*, *skew*, and *hmean*. These results are in line with previous findings on a subset of Russian (Spinu *et al.*, 2018), Azerbaijani (Ghaffarvand Mokari and Mahdinezhad Sardhaei, 2020), and Romanian (Spinu and Lilley, 2016) fricatives. These studies also reported that cepstral coefficients clearly outperform traditional spectral parameters. The differences in accuracy were very similar with classification rates around 60% for acoustic parameters and around 80% and higher for CCs.

TABLE III. The performance of the two classifiers across ten replications for predicting gender and speaker. The abbreviations are interpreted as follows: DT = single decision tree, RF = random forest, Acc = accuracy. The baseline refers to the majority baseline.

Predicting	Classifier	Set	Kappa	Acc	Baseline
Gender	DT	acoustic set	0.28	0.64	0.50
Gender	RF	acoustic set	0.45	0.72	0.50
Gender	DT	MFCCs	0.31	0.66	0.50
Gender	RF	MFCCs	0.76	0.88	0.50
Speaker	DT	acoustic set	0.00	0.01	0.02
Speaker	RF	acoustic set	0.21	0.22	0.02
Speaker	DT	MFCC	0.04	0.05	0.02
Speaker	RF	MFCC	0.64	0.64	0.02

In predicting individual speakers, the accuracy drops for both methods (Table III). The two decision tree-based algorithms were unable to identify speakers with the traditional set of acoustic parameters, as revealed by the very low accuracy rates. With MFCCs, the accuracy is much higher with around 64%, suggesting a moderate performance in predicting speaker.

### B. Acoustic gender variation

Figure 2 displays how the parameters from the acoustic set are distributed across female and male speakers. The figure shows that significant gender variation exists for most of the parameters and fricatives.

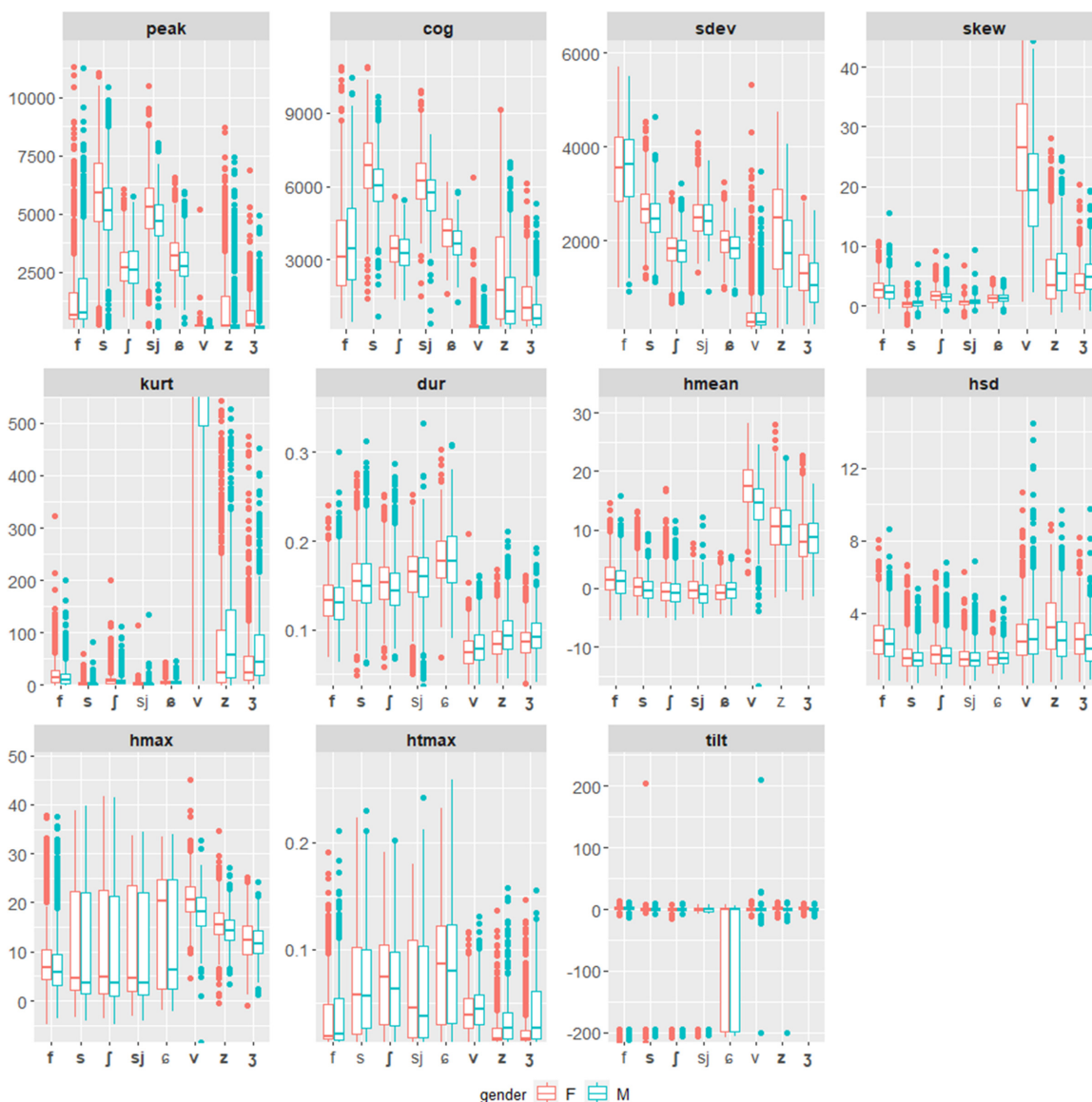


FIG. 2. (Color online) Gender variation across the eight fricatives. The phonetic symbols in bold indicate that gender variation for these fricatives is significant based on Wilcoxon tests using the Bonferroni multiple testing correction. For better visualization, each x axis scale is adapted to each parameter, and the y axes for *skew* and *kurt* were further bound.

Contrary to previous studies which noted a higher gender variation in anterior fricatives (Kochetov, 2017), our results suggest a significant variation for almost all parameters in all three places of articulation. Even though the distributions of values of both females and males are marked by many outliers in most parameters and fricatives, some tendencies can be detected, which partly confirm previous findings. First, female speakers have higher spectral energy in voiceless sibilant fricatives (Flipsen *et al.*, 1999; Jongman *et al.*, 2000; Kochetov, 2017; Ludger *et al.*, 2021; Newman *et al.*, 2001; Schwartz, 1968). We also observe the same trend for the voiceless, palatalized, and voiced sibilants, as well as for the non-sibilant [v]. The spectral energy in the voiceless non-sibilant [f], on the other hand, is lower in female speakers, as illustrated by lower values in *peak* and *cog*. In the two palatalized sibilant fricatives, the observed patterns are convergent with previous studies, which report higher spectral energy in female productions (Kochetov, 2017).

The second spectral moment (indicating spectral spread) has rarely been explored in the literature. In the current data, significant gender differences and higher values are measured for female speakers in all sibilants except [s<sup>j</sup>]. This is also theoretically expected as the spectral spread is correlated with the center of gravity, i.e., a higher *cog* leads to a higher *sdev*.

In [s], a tendency for negative skewness in female speakers and positive values or values centered near zero for male speakers are observed (Flipsen *et al.*, 1999; Ludger *et al.*, 2021). The female speakers in the present analysis show a more symmetrical distribution of energy in [s], as reflected by a skewness around zero. Male speakers generate significantly more energy at lower frequencies, with mean values around 0.5. In the representations of [f] and [ʃ], both genders exhibit an asymmetrical distribution of energy, with higher positive values for female speakers and, therefore, predominant energy at lower frequencies. The highest skewness is observed in the voiced bilabial fricative, and female speakers also produce more energy in lower frequency bands than male speakers in [v]. In the realizations of [z] and [ʒ] male speakers show higher *skew* than females.

Kurtosis describes the peakedness of the energy distribution and has not previously been investigated for gender variation. In the present data, only the spectral energy in the alveolar [s] and [s<sup>j</sup>] is normally distributed, as specified by a kurtosis of around 3. In these fricatives, female speakers show a higher kurtosis than males. All other fricatives show different degrees of peakedness, while the values decrease in both females and males with the place of articulation moving backward. Kurtosis of voiced fricatives displays a huge variation, with values over 1000 in [v], and values for females are higher than for males. Such high values observed in [v] suggest a very compact spectral distribution. In [z] the mean values are almost equal but the ranges differ between female and male speakers. In [ʒ], on the other hand, females produced lower kurtosis than males.

Duration also varies with gender in the non-palatalized fricatives, with female speakers producing longer voiceless fricatives while male speakers produce longer voiced fricatives.

Even though gender differences have widely been investigated in the spectral domain of voiceless fricatives, very limited information is available about additional parameters such as the distribution of periodic and aperiodic energy and can be quantified by HNR measures. As illustrated by *hmean* and *hmax*, the overall trend for non-palatalized fricatives is a decrease in harmonic energy as the place of articulation moves backward. Interestingly, a significant gender variation is present in some of the HNR parameters. In most fricatives, female production tends to be more harmonic as indicated by higher positive values of *hmean* in voiceless fricatives (except [ç]) and in [v] and in terms of *hmax* in all eight fricatives. Our analysis shows that the distribution of periodic and aperiodic energy generally differs between female and male speakers. Females thus seem to produce more harmonic energy in Russian fricatives. Interestingly, this gender difference in harmonicity is a distinct phenomenon from the breathy voice described as a female attribute compared to male voice quality [e.g., Klatt and Klatt (1990)].

Moreover, the findings from the t-SNE analysis (Fig. 3), comparing distances between contrastive fricatives, suggest the existence of a significant gender variation. For a better overview and an easier interpretation, only the findings of the spectral, temporal, and HNR parameters are reported in the following outline. The observations made for voiceless sibilants by previous research (Weirich and Simpson, 2015) can be extended to other fricative pairs contrasted by the place of articulation. Our analysis indicates that for both non-palatalized voiceless [f-s], [s-ʃ], and voiced pairs [v-z], [z-ʒ], contrasted by the place of articulation, female speakers produce, in general, a larger distance between the two elements of the pair. However, in the palatalized pair [s<sup>j</sup>-ç], no such gender variation is observed. In fricative pairs contrasted by voicing [f-v], [s-z], and [ʃ-ʒ], female speakers produce closer elements. Female speakers also show a larger distance between the two sibilant fricative pairs [s-s<sup>j</sup>] [ʃ-ç] contrasted by palatalization.

Additionally, a comparison of the IQR provides another way of exploring the gender-specific acoustic variation across the fricatives. These results (included in the supplementary materials)<sup>6</sup> show that female speakers tend to have higher IQR values, indicating a higher variation in female speakers. However, these tendencies cannot be generalized and gender variation seems not to be systematic across gender categories. It differs depending on the fricative and measured parameter.

### C. Inter- and intra-speaker variation

The comparison between the first principal component (PC1) of the PCA performed for each fricative is summarized in Table IV. First, the *peak* frequency and the spectral moments were the most variant parameters in [f], [v], and [z]. As an example, in [v], PC1 mostly loads on *skew* (0.32) and *kurt* (0.49), while in [f], PC1 is mostly relevant for other parameters of the spectral moments such as *cog* (0.23) or



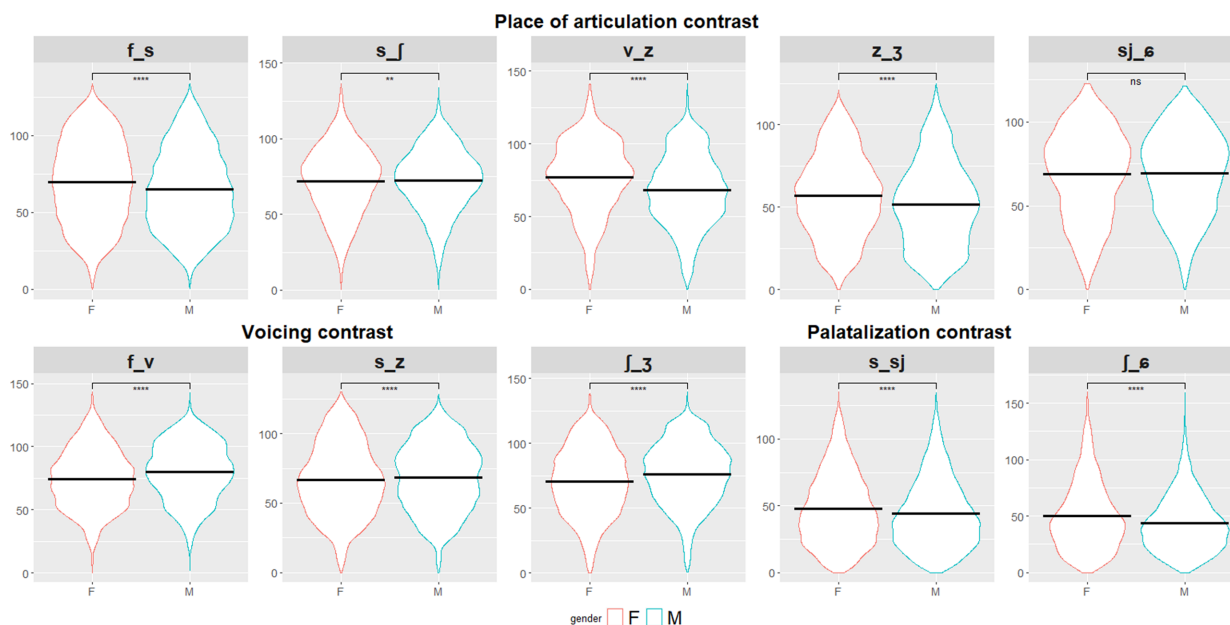


FIG. 3. (Color online) Distance within fricative pairs contrasted by the place of articulation, voicing, and palatalization. The distance is derived from the t-SNE representation and the horizontal lines indicate the mean values. Females produce more distant fricatives in pairs contrasting in terms of the place of articulation, except for [s<sup>h</sup>-ç] and in pairs contrasting in palatalization. In voicing contrast, females produce closer fricatives than male speakers.

*sdev* (0.21). In other fricatives, the spectral domain seems to be less meaningful in interpreting variation. Then, duration also explains some of the variation, but only in voiceless sibilants. For example, *dur* explains a larger portion of PC1 (0.15) for [s<sup>h</sup>] when compared with other fricatives, which generally have a value around 0.01 or 0.02. As for the HNR parameters, they were more variable than the spectral parameters in sibilant fricatives, which was unexpected. Different patterns were observed for voiceless and voiced fricatives. In voiceless sibilants, *hmax*, *htmax* and *tilt* are the most relevant in PC1. As an example, for [ʃ], *hmax*, *htmax*, and *tilt* explain 0.30, 0.18, and 0.29 of PC1 respectively. In voiced sibilants, parameters such as *hmean* and *hsd* explain a large portion of PC1. For example, for [z], *hmean* and *hsd* account for 0.12

TABLE IV. Summary of the PC1 of all eight fricatives. The “PC1 variance” indicates how much variance of the total variance is explained by the first component. The values per acoustic parameter indicate how much variation in each fricative can be explained by a certain parameter averaged over the speaker.

Parameter	[f]	[s]	[ʃ]	[s <sup>h</sup> ]	[ç]	[v]	[z]	[ʒ]
PC1 variance	0.32	0.41	0.45	0.46	0.46	0.57	0.57	0.43
peak	0.19	0.03	0.01	0.05	0.01	0.00	0.09	0.06
cog	0.23	0.01	0.00	0.03	0.01	0.01	0.15	0.10
sdev	0.21	0.03	0.02	0.03	0.01	0.06	0.18	0.13
skew	0.06	0.00	0.01	0.00	0.00	0.32	0.10	0.08
kurt	0.01	0.00	0.00	0.00	0.00	0.49	0.03	0.02
dur	0.02	0.15	0.13	0.15	0.12	0.01	0.02	0.02
hmean	0.05	0.01	0.02	0.00	0.00	0.07	0.12	0.15
hsd	0.17	0.02	0.05	0.02	0.01	0.01	0.23	0.31
hmax	0.06	0.28	0.30	0.24	0.26	0.04	0.04	0.04
htmax	0.00	0.23	0.18	0.18	0.17	0.00	0.04	0.07
tilt	0.01	0.23	0.29	0.29	0.39	0.00	0.00	0.01

and 0.23 of PC1, respectively. These results suggest that the greatest variation was detected in the distribution of periodic and aperiodic energy in the sibilant fricatives.

After comparing the variance encoded in PC1, we compare the SD-ratios (Fig. 4), which provide information on the relation between the inter- and intra-speaker variation. Figure 4 shows which parameters are highly variable across speakers. For example, in [f], the highest between-speaker variation is found in *kurt* and *tilt*. Taking another example from [f], the SD-ratios of *cog*, *sdev*, and *hsd* are extremely low, which indicates that the within-speaker variation is higher than the between-speaker variation. This, in turn, means that these three parameters provide little speaker information in the fricative [f]. An opposite example is found for [v] with *peak*, *cog*, *sdev*, and *tilt*, which have high SD-ratios. As an overview, the sibilant voiceless fricatives show very similar distributions of the SD-ratios. A higher between-speaker variation than within-speaker was found in the spectral moments, *hmean* and *hsd*, while *hmax* and *htmax* indicated higher within-speaker variation. The opposite patterns for the same parameters were observed in the voiced fricatives.

After visualizing the PC1 and the SD-ratios, we combined them in Fig. 5, which shows that in most cases, parameters with a high explanatory power of PC1 have a low SD-ratio. Consequently, the majority of parameters identified to explain a large part of the variation were produced by speakers with a high degree of within-speaker variation. Additionally, the correlation analysis of PC1 and SD-ratio found a negative correlation for almost all parameters and fricatives. These findings suggest that no parameter effectively explains variation in general within fricatives, since there is a high degree of between-speaker variation.

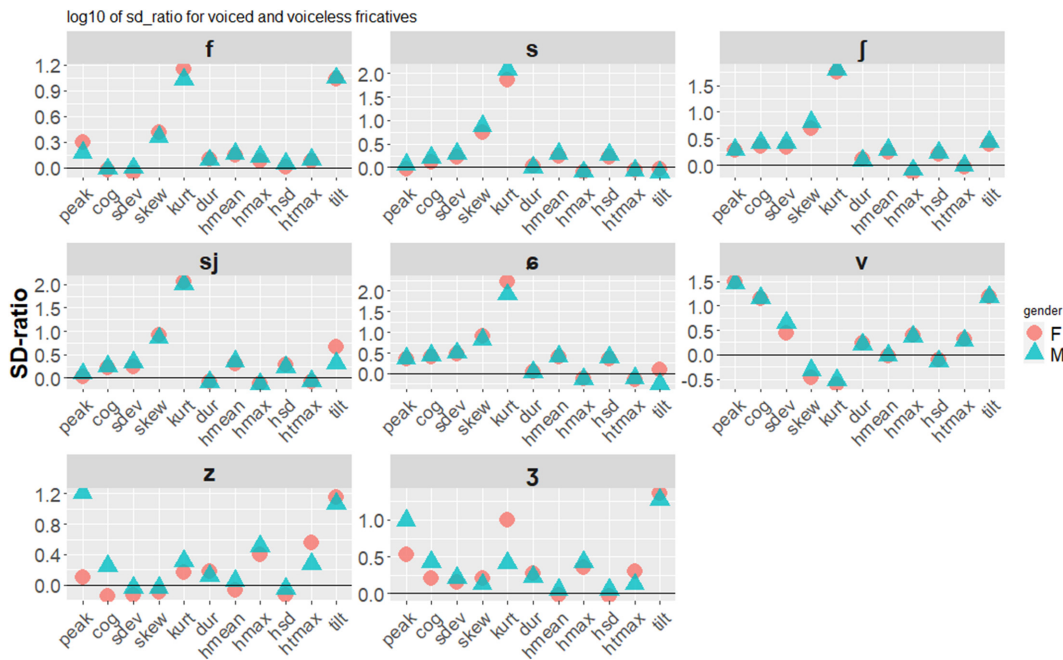


FIG. 4. (Color online) The SD-ratio averaged over all speakers by fricative and acoustic parameters. For better visualization, the log10 of an SD-ratio is used. For instance, 3 is now equal to 0.48, and an SD-ratio of 1 equals 0. The values below 0 in this Figure mean that the within-speaker variation is higher than the between-speaker variation. Values above 0 indicate higher between-speaker variation. The larger the SD-ratio, the higher the between-speaker variation and the lower the within-speaker variation, which indicates high speaker-discriminating potential.

Finally, a further illustration is given in Table V, by narrowing the comparison to three speakers and two fricatives [f] and [ʃ] only. Parameters with a low PC1 and a high SD-ratio indicate that the within-speaker variation is lower and the between-speaker variation high. It can be inferred that these parameters could potentially provide speaker-specific information.

It is striking that in both fricatives, the three speakers differed greatly in the set of parameters in which they produced the most and the least variation. The acoustics of

speaker 1 is characterized by a lower idiosyncrasy in fricatives in comparison to the other two speakers. The SD-ratios are between 1 and 2 at the highest and the PC1 variance is 0.38 in [f] and 0.46 in [ʃ]. Speaker 7 exhibits some idiosyncrasy in both fricatives, indicated by SD-ratios between 2 and 3, and higher PC1 variance values for both fricatives. Speaker 16 has the greatest degree of individual information in fricatives. The SD-ratios reach up to 15 and the PC1 variance is almost equal in both fricatives with 0.56 and 0.55.

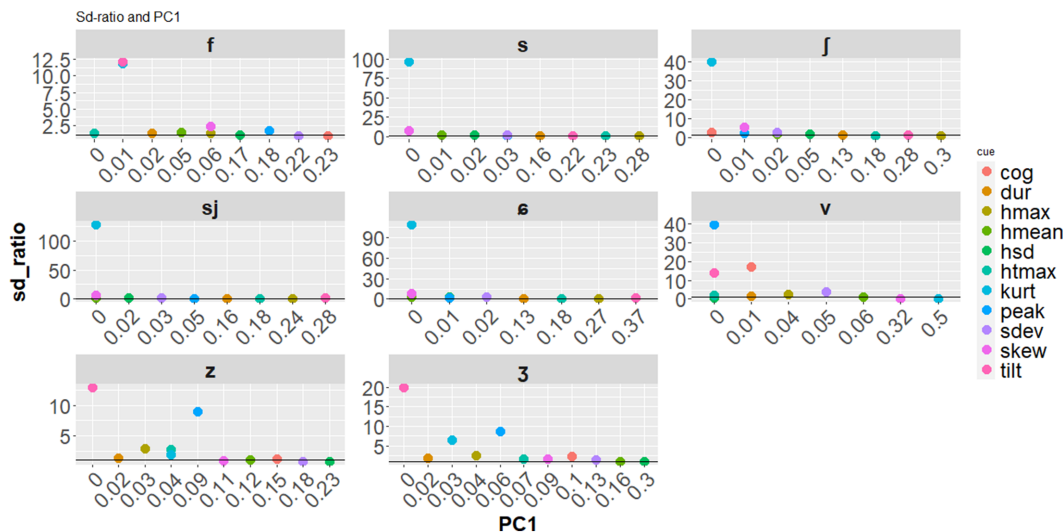


FIG. 5. (Color online) Comparison of the variance included in PC1 and the SD-ratio by acoustic parameter and fricative. A higher PC1 value shows that more variation within a fricative can be explained by this parameter. A higher SD-ratio implies a higher between-speaker variation. Values below 1 indicated by the horizontal line suggest that the within-speaker variation is higher than the variation between speakers.

TABLE V. PC1 and SD-ratio for three speakers and two fricatives [f] and [ʃ]. The three speakers show remarkable differences in the distribution of the ratio between PC1 and SD-ratio across the two fricatives.

Fricative	[f]						[ʃ]					
	1		7		16		1		7		16	
	SD	PC1	SD	PC1	SD	PC1	SD	PC1	SD	PC1	SD	PC1
peak	1.22	0.10	2.89	0.02	0.25	0.39	1.37	0.01	1.02	0.02	5.52	0.01
cog	1.17	0.13	1.75	0.10	0.48	0.23	1.54	0.01	1.24	0.01	3.05	0.00
sdev	0.91	0.15	1.27	0.19	1.00	0.02	1.81	0.01	1.35	0.01	1.43	0.02
skew	1.35	0.02	2.55	0.06	3.74	0.02	1.34	0.00	2.54	0.00	0.98	0.00
kurt	1.62	0.00	4.40	0.02	15.54	0.00	1.72	0.00	2.98	0.00	0.87	0.00
dur	1.15	0.08	1.54	0.02	1.18	0.02	1.15	0.12	1.77	0.06	1.79	0.05
hmean	1.06	0.01	1.54	0.10	2.96	0.03	1.36	0.00	1.56	0.03	1.06	0.04
hsd	0.85	0.08	1.18	0.33	1.42	0.10	0.93	0.03	1.05	0.09	0.94	0.10
hmax	0.89	0.18	0.53	0.11	0.68	0.10	1.06	0.17	1.09	0.40	1.05	0.26
htmax	0.92	0.10	0.82	0.04	1.04	0.04	0.99	0.22	1.07	0.36	1.33	0.03
tilt	0.66	0.15	1.97	0.00	0.08	0.05	0.95	0.43	1.96	0.03	0.93	0.49
PC1 variance	0.38		0.56		0.56		0.46		0.52		0.55	

V. DISCUSSION

The first objective of the current study was to predict the speaker’s gender and individual speakers from two sets of acoustic parameters. The second objective was to investigate inter- and intra-speaker variation in detail.

The results for predicting speakers’ gender indicate that MFCCs clearly outperformed spectral, temporal, and HNR parameters with an accuracy of 0.88 (vs 0.72). Thus, gender can be predicted to some extent by both sets of acoustic parameters, but speakers’ gender information is best captured by MFCCs. These findings are in line with previous investigations stating that cepstral coefficients outperform spectral parameters (Ghaffarvand Mokari and Mahdinezhad Sardhaei, 2020; Spinu et al., 2018; Spinu and Lilley, 2016). With moderate accuracy, gender can, nevertheless, be predicted by the traditional set of parameters and the most important were *peak*, *cog*, *skew*, and *hmean*.

Both classifiers failed to predict individual speakers with the acoustic parameter set. The speaker identity could, nevertheless, be partially predicted by an RF applied to MFCCs, leading to an accuracy of 0.64. These observations imply that fricatives do encode some speaker information that could probably be enhanced if transitions were considered. In the current database, all fricatives were recorded in intervocalic positions, but in the classification tasks, the transition zones were not considered indeed.

To further explore gender variation, the distributions of the acoustic parameters were compared between female and male speakers, revealing a complex pattern influenced by the fricative and the parameter considered. Furthermore, we tested whether female and male speakers organize their fricative contrasts differently in terms of distance. Previous studies, on vowels (Diehl et al., 1996; Weirich and Simpson, 2014) or sibilant voiceless fricatives (Weirich and Simpson, 2015), suggested that females produce more distance between two contrastive segments and concluded that females articulate more distinct categories. Our data

partially confirm these findings. Female speakers showed a larger distance between voiceless non-palatalized and voiced fricatives of different places of articulation, as well as in pairs contrasted by palatalization. However, they also produced a smaller distance in fricatives contrasted by voicing.

Taken together, these findings imply that the overall patterns of gender variation are less systematic across female and male speakers, but more specific to the segments and variable across acoustic parameters. Gender variation was often evaluated by previous studies measuring spectral moments in [s]. Our analysis also shows (Fig. 2) a large variation between fricatives of different places of articulation and voicing quality. Patterns found in [s] are not necessarily transferable to, for instance, [z], [sʰ], or [ʃ]. It is, therefore, recommended in future studies to consider the research of gender variation to other fricatives.

These findings could also explain why ML classifiers performed only moderately when predicting gender from spectral, temporal, and HNR parameters. The distribution of the representations shown in Fig. 2 suggests significant gender variation in most parameters, so they probably all contribute, to a certain extent, to the distinction between males and females. Nevertheless, to test the importance of the individual acoustic parameters, it may be necessary to compare separately, for example, the performance on spectral and HNR parameters. Furthermore, future analyses should probably include the comparison of voiced and voiceless fricatives.

To further explore why the ML classifiers were unable to predict speakers using spectral, temporal, and HNR parameters, the inter- and intra-speaker variation was investigated in more detail since previous research suggested some evidence that fricatives contain speaker information in the spectral moments (Kavanagh, 2011; Newman et al., 2001; Schindler and Draxler, 2013).

One explanation for why the classifiers failed to predict speakers with the traditional set of parameters in Russian

fricatives is that the dataset may be too limited to account for the intra-speaker variation correctly. The dataset has a high number of speakers (59) to be predicted and a relatively low number of tokens per fricative and speaker. On the other hand, speaker variation can possibly be encoded in a more complex way in the acoustic parameters. To explore this complexity of speaker information and individual differences, we provided a detailed description of speaker specificity in fricatives.

A PCA was performed for each fricative to identify the parameters that explain the most variation. For the analysis, the results for the first principal component were scrutinized. The findings show that spectral parameters explain the variation only in [f] across voiceless fricatives and, to some degree, in voiced fricatives. The duration and the HNR seem also to play a role in almost all fricatives with a variable contribution. Additionally, the SD-ratio was computed to determine whether the difference shown by PC1 is caused by between or within-speaker variation. The results show that most parameters identified by the PCA to be variable are characterized by an SD-ratio below 1, meaning that the within-speaker variation is higher than the inter-speaker variation.

A more detailed analysis of three speakers provided further insights into the distribution of variant and stable parameters across speakers and fricatives (Table V). The results clearly showed how largely speakers differ in the parameters exhibiting a high within-speaker variation or, on the contrary, a high constancy. The values of *peak* in [f], for instance, show that while speaker 7 has an SD-ratio of 2.89 and 0.02 in PC1, speaker 16 has the opposite pattern with an SD-ratio of 0.25 and PC1 of 0.39. This means that speaker 7 produces a very stable *peak*, and speaker 16 has a high variation within *peak* frequencies. The analysis demonstrates that speakers can potentially encode their individual information in different ways for the same segments. Also, it can be noted that no parameter seems to be consistently employed by speakers to encode individual information. Variation is higher than expected, and the process is more complex than just detecting the most stable parameter within and between speakers.

From these analyses, we can conclude that the variation differs from one parameter to the other, but that the level of individuality encoded in fricatives is highly speaker-dependent, which can, in turn, explain why the general performances aforementioned for individual recognition were poor. On the other hand, taking into account the conclusions of the identification of the place of articulation in the same dataset (Ulrich *et al.*, 2021)—which found that center of gravity and spectral spread provide sufficient information to distinguish [f, s, ʃ]—strong speaker effects were not expected to be found in the spectral domain.

To summarize, this analysis demonstrates that feature distributions exhibit large variation across fricatives. It also shows that the individuality encoded in fricatives is highly speaker-dependent. Intra- and inter-speaker variation is complex and no set of parameters seems to explain acoustic variability and stability for all fricatives and speakers.

Speakers can potentially encode some individual information in different parameters for the same fricative. To determine whether patterns between speakers exist and whether some speakers can be grouped together according to the similar distribution of variable and stable parameters, further exploration is necessary. One can also investigate to what extent a speaker's individual variation in one segment can predict the variation in another segment.

The current study has potential implications for phonetic research as well as for ASR applications. It helps us to understand how individual speaker information is distributed in fricatives across all acoustic parameters. More investigation is, nevertheless, needed to understand which underlying mechanisms define speaker-specific patterns and what influences the relative contribution of each parameter to idiosyncratic information. We also found that MFCCs contain more detailed speaker information in fricatives than the information obtained from traditional spectral, temporal, and HNR parameters. These findings suggest that the spectral domain contains information on speakers' idiosyncrasies but the spectral parameters used in phonetic research do not sufficiently capture this information.

## ACKNOWLEDGMENTS

We wish to thank our participants, the Phonetic Lab in St. Petersburg (and the sound engineer Tatiana Chukaeva in particular), the University of Zürich for its financial support, technical support and help with the design of the experiment (to Volker Dellwo in particular). N.U. was partly supported by a grant from the Doctoral Program of Linguistics of the Faculty of Arts and Social Sciences, University of Zürich, Switzerland, NU, M.A.-T. was funded by the IDEX Lyon Fellowship Grant No. 16-IDEX-0005 (2018–2021), and indirectly by the Labex ASLAN (ANR-10-LABX-0081) of the University of Lyon within the program Investissements d'Avenir (ANR-11-IDEX-0007) of the French National Research Agency (ANR). M.A.-T. is also thankful for the support of the Junior researcher grant from the French National Research Agency (ANR-20-CE27-0021). We also want to thank the Maison des Sciences de l'Homme Lyon Saint-Etienne (UAR 2005 CNRS, Université Lyon 2, Université Lyon 3, Université St-Etienne, Université Lyon 1, ENS de Lyon, Sciences Po Lyon) for its financial support for the final proofreading.

<sup>1</sup>In the literature mentioned in this section, the terms *gender* and *sex* are both utilized without reporting any discrepancies between the biological sex and the socially constructed gender across the participants. In the context of the present study, we use gender throughout the paper.

<sup>2</sup>Some speakers ended their voiceless fricatives with a somehow long and unexpected post-aspiration in intervocalic positions, but also when the fricative appeared at the end of the word and sentence. In these cases, the fricatives were segmented according to the changes in high-energy events, and the post-aspiration part was not considered. The voiced fricatives also represented a segmentation challenge, because the waveform and the spectrogram were insufficiently informative to define the onset and offset. The boundaries for these fricatives were identified according to perceptual judgments. In general, it should be noted that it is hardly possible to



standardize segmentation criteria across all fricatives and speakers. Such an investigation would also require taking sociophonetics factors into account and is beyond the scope of the present study.

<sup>3</sup>The non-palatalized post-alveolar fricatives [ʃ] and [ʒ] and the palatalized [ç] and [ʒʲ] are regarded not to be paired because [ʃ] and [ʒ] do not follow the same rules as other consonants do (become palatalized at the end of a noun in the locative singular or in the conjugation of verbs) (Timberlake, 2004). Nevertheless, in phonetic acoustic studies [ʃ] and [ç] are often treated as non-palatalized and palatalized pair (Kochetov, 2017; Spinu et al., 2018).

<sup>4</sup>The fricatives [x], [vʲ], and [zʲ] are not included in the analysis due to their low count in the data.

<sup>5</sup>The number 200 was chosen based on the stabilization point of the predictions.

<sup>6</sup>See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0017827> for all analysis and code, the output of machine learning classifiers, and results of the IQR.

Ajili, M., Bonastre, J.-F., Ben Kheder, W., Rossato, S., and Kahn, J. (2017). "Phonological content impact on wrongful convictions in Forensic Voice Comparison context," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, New Orleans, LA, pp. 2147–2151.

Alsulaiman, M., Mahmood, A., and Muhammad, G. (2017). "Speaker recognition based on Arabic phonemes," *Speech Commun.* **86**, 42–51.

Al-Tamimi, J., and Khattab, G. (2015). "Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants," *J. Acoust. Soc. Am.* **138**(1), 344–360.

Antal, M. (2008). "Phonetic speaker recognition," in *Proceedings of the 7th International Conference COMMUNICATIONS*, pp. 67–72.

Barry, S. M. E. (1995). "Variation in vocal fold vibration during voiced obstruents in Russian," *Int. J. Lang. Commun. Disord.* **30**(2), 124–131.

Boersma, P., and Weenink, D. (2022). "Praat: Doing phonetics by computer (version 6.2.14) [computer program]," <http://www.praat.org/> (Last viewed February 19, 2023).

Bolla, K. (1981). *A Conspectus of Russian Speech Sounds*, 32nd ed. (Böhlau Verlag, Köln).

Breiman, L. (2001). "Random forests," *Mach. Learn.* **45**, 5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*, Wadsworth & Brooks/Cole Statistics/Probability Series (Springer, Berlin).

Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (1996). "On explaining certain male-female differences in the phonetic realization of vowel categories," *J. Phon.* **24**(2), 187–208.

Draxler, C., and Jänsch, K. (2022). <https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/> (Last viewed February 19, 2023).

Eckert, P. (1989). "The whole woman: Sex and gender differences in variation," *Lang. Var. Change* **1**(3), 245–267.

Enzinger, E., and Balazs, P. (2011). "Speaker verification using pole/zero estimates of nasals," *Analele Univ. "Eftimie" 18*, 33–44.

Ferragne, E., and Pellegrino, F. (2010). "Formant frequencies of vowels in 13 accents of the British Isles," *J. Int. Phon. Assoc.* **40**(1), 1–34.

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., and McSweeney, J. (1999). "Acoustic characteristics of /s/ in adolescents," *J. Speech. Lang. Hear. Res.* **42**(3), 663–677.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," *J. Acoust. Soc. Am.* **84**(1), 115–123.

Ganchev, T., Fakotakis, N., and Kokkinakis, G. (2005). "Comparative evaluation of various MFCC implementations on the speaker verification task," in *Proceedings of the SPECOM*, pp. 191–194.

Gendrot, C., Ferragne, E., and Pellegrini, T. (2019). "Deep learning and voice comparison: Phonetically-motivated vs. automatically-learned features," in *ICPhS*.

Gendrot, C., Ferragne, E., and Pellegrini, T. (2020). "Informations segmentales pour la caractérisation phonétique du locuteur: Variabilité inter-et intra-locuteurs" ("Segmental information for speaker phonetic characterization: Inter- and intra-speaker variability"), *6e Conférence Conjointe Journées D'Études Sur la Parole (JEP, 33e Édition), Traitement Automatique Des Langues Naturelles (TALN, 27e Édition), Rencontre Des Étudiants Chercheurs en Informatique Pour le Traitement Automatique Des Langues (RÉCITAL, 22e Édition)*, Vol. 1: Journées

d'Études sur la Parole (Proceedings of the 6th joint conference Journées d'Études sur la Parole).

Ghaffarvand Mokari, P., and Mahdinezhad Sardhaei, N. (2020). "Predictive power of cepstral coefficients and spectral moments in the classification of Azerbaijani fricatives," *J. Acoust. Soc. Am.* **147**(3), EL228–EL234.

Gordon, M., Barthmaier, P., and Sands, K. (2002). "A cross-linguistic acoustic study of voiceless fricatives," *J. Int. Phonetic Assoc.* **32**(2), 141–174.

Harper, S. K. (2021). "Individual differences in phonetic variability and phonological representation," Ph.D. thesis, University of Southern California, Los Angeles, CA.

Henton, C. (1995). "Cross-language variation in the vowels of female and male speakers," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 4, pp. 420–423.

Hughes, G. W., and Halle, M. (1956). "Spectral properties of fricative consonants," *J. Acoust. Soc. Am.* **28**(2), 303–310.

Jesus, L. M. T., and Jackson, P. J. B. (2008). "Frication and voicing classification," *Comput. Process. Portuguese Lang.* **5190**, 11–20.

Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**(3), 1252–1263.

Kavanagh, C. (2011). *Intra-and Inter-Speaker Variability in Acoustic Properties of English/s* [International Association for Forensic Phonetics and Acoustics (IAFPA)].

Kavanagh, C. M. (2012). "New consonantal acoustic parameters for forensic speaker comparison," Ph.D. thesis, University of York, York, UK.

Kisler, T., Reichel, U., and Schiel, F. (2017). "Multilingual processing of speech via web services," *Comput. Speech Lang.* **45**, 326–347.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**(2), 820–857.

Kochetov, A. (2017). "Acoustics of Russian voiceless sibilant fricatives," *J. Int. Phon. Assoc.* **47**(3), 321–348.

Kong, Y.-Y., Mullangi, A., and Kokkinakis, K. (2014). "Classification of fricative consonants for speech enhancement in hearing devices," *PLoS One* **9**(4), e95001.

Labov, W. (1990). "The intersection of sex and social class in the course of linguistic change," *Lang. Var. Change* **2**(2), 205–254.

Lilley, J., Spinu, L., and Athanasopoulou, A. (2021). "Exploring the front fricative contrast in Greek: A study of acoustic variability based on cepstral coefficients," *J. Int. Phonetic Assoc.* **51**(3), 393–424.

Ludger, P., Fuchs, S., and Seifert, F. (2021). "Differences between male and female speakers in the production of /s/: A cross-linguistic study," *17. Phonetik und Phonologie im deutschsprachigen Raum (PP)*.

Machač, P., and Skarnitzl, R. (2009). "Principles of phonetic segmentation," *Epocha*.

McDougall, K., and Nolan, F. (2007). "Discrimination of speaker using the formant dynamics of /u:/ in British English," in *Proceedings of the International Congress of Phonetic Sciences 1825–1828*, <http://icphs2007.de/conference/Papers/1567/1567.pdf> (Last viewed February 19, 2023).

Munson, B., McDonald, E. C., DeBoe, N. L., and White, A. R. (2006). "The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech," *J. Phon.* **34**(2), 202–240.

Narayanan, S. S., Alwan, A. A., and Haker, K. (1995). "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Am.* **98**(3), 1325–1347.

Newman, R. S., Clouse, S. A., and Burnham, J. L. (2001). "The perceptual consequences of within-talker variability in fricative production," *J. Acoust. Soc. Am.* **109**(3), 1181–1196.

Romeo, R., Hazan, V., and Pettinato, M. (2013). "Developmental and gender-related trends of intra-talker variability in consonant production," *J. Acoust. Soc. Am.* **134**(5), 3781–3792.

Rose, P. (2007). "Forensic speaker discrimination with Australian English vowel acoustics," in *ICPhS XVI*, Vol. 6, No. 10.

Schiel, F. (1999). "Automatic phonetic transcription of non-prompted speech," in *ICPhS 99*.

Schiel, F. (2023). "The Munich automatic segmentation system MAUS," <https://www.bas.uni-muenchen.de/Bas/BasMAUS.html> (Last viewed August 20, 2022).

Schindler, C., and Draxler, C. (2013). "Using spectral moments as a speaker specific feature in nasals and fricatives," in *Interspeech 2013*, ISCA, pp. 2793–2796.

Schwartz, M. F. (1968). "Identification of speaker sex from isolated, voiceless fricatives," *J. Acoust. Soc. Am.* **43**(5), 1178–1179.

- Silbert, N., and de Jong, K. (2008). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," *J. Acoust. Soc. Am.* **123**(5), 2769–2779.
- Smorenburg, L., and Heeren, W. (2020). "The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues," *J. Acoust. Soc. Am.* **147**(2), 949–960.
- Spinu, L., Kochetov, A., and Lilley, J. (2018). "Acoustic classification of Russian plain and palatalized sibilant fricatives: Spectral vs. cepstral measures," *Speech Commun.* **100**, 41–45.
- Spinu, L., and Lilley, J. (2016). "A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives," *J. Phon.* **57**, 40–58.
- Spinu, L., Vogel, I., and Timothy Bunnell, H. (2012). "Palatalization in Romanian—Acoustic properties and perception," *J. Phon.* **40**(1), 54–66.
- Stevens, P. (1960). "Spectra of fricative noise in human speech," *Lang. Speech* **3**(1), 32–49.
- Stuart-Smith, J. (2007). *Empirical Evidence for Gendered Speech Production: /s/ in Glaswegian* (Mouton de Gruyter, Berlin).
- Timberlake, A. (2004). *A Reference Grammar of Russian* (Cambridge University Press, Cambridge).
- Ulrich, N. (2022). "Russian fricatives [Dataset]," <https://www.swissbase.ch/en/catalogue/studies/20152/latest/datasets/2183/2445/overview> (Last viewed February 19, 2023).
- Ulrich, N., Allasonnière-Tang, M., Pellegrino, F., and Dediu, D. (2021). "Identifying the Russian voiceless non-palatalized fricatives /t/, /s/, and /ʃ/ from acoustic cues using machine learning," *J. Acoust. Soc. Am.* **150**(3), 1806–1820.
- Van der Maaten, L., and Hinton, G. (2008). "Visualizing data using t-SNE," *J. Mach. Learn. Res.* **9**(11), 2579–2605.
- Weirich, M., and Simpson, A. P. (2014). "Differences in acoustic vowel space and the perception of speech tempo," *J. Phon.* **43**, 1–10.
- Weirich, M., and Simpson, A. P. (2015). "Gender-specific differences in sibilant contrast realizations in English and German," in *ICPhS*.