Hello, who is this? The relationship between linguistic and speaker-dependent information in the acoustics of consonants

Smorenburg, B.J.L.

CHAPTER 2

Linguistic effects on the speaker-dependent variability in fricatives

Abstract

Although previous work has shown that some speech sounds are more speaker-specific than others, not much is known about the speaker information of the same segment in different linguistic contexts. The present study therefore investigated whether Dutch fricatives /s/ and /x/ from telephone dialogues contain differential speaker information as a function of syllabic position and labial co-articulation. These linguistic effects, established in earlier work on read broadband speech, were firstly

investigated. Using a corpus of Dutch telephone speech, results showed that the telephone bandwidth captures the expected effects of perseverative and anticipatory labialization for dorsal fricative /x/, for which spectral peaks fall within the telephone band, but not for coronal fricative /s/, for which the spectral peak falls outside the telephone band. Multinomial logistic regression shows that /s/ contains slightly more speaker information than /x/ in telephone speech and that speaker information is distributed across the speech signal in a systematic way; even though differences in classification accuracy were small, codas and tokens with labial neighbors yielded higher scores than onsets and tokens with non-labial neighbors for both /s/ and /x/. These findings indicate that speaker information in the same speech sound is not the same across linguistic contexts.

This chapter was published:

Smorenburg, L., & Heeren, W. (2020). The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. *Journal of the Acoustical Society of America, 147*(2), 949-960. doi: 10.1121/10.0000674

2.1 Introduction

Speakers' voices convey idiosyncratic information. In everyday communication, listeners make use of this information while interpreting what they hear and, in forensic phonetics, speech analyst use this information to acoustically characterize speakers. Although previous research has already shown that some speech sounds convey more speaker information than others (e.g., Kavanagh, 2012; Van den Heuvel, 1996), not much is known about how speaker information in the same speech sound interacts with its linguistic environment. The present study investigated the speaker-dependency of the same speech sound in different linguistic contexts. Specifically, we examined whether the speaker-dependency of Dutch fricatives varied as a function of syllabic position and labial co-articulation. Additionally, the aim was to determine which segment and which specific (combinations of) acoustic features are most successful in characterizing speakers. Contrary to many previous studies that used read speech, the present study used spontaneous telephone dialogues to investigate speaker variation.

Investigating the distribution of speaker information is relevant for forensic speech science because the role of the speaker in speech production is still largely unclear. It is known that speaker-dependent information conveys all kinds of meanings (e.g., gender identity) and that these meanings are also perceived by listeners. However, it is not clear where in the speech signal speakers have the articulatory freedom to convey speaker information, or if there are such distributional limitations. Additionally, this study may be particularly relevant for forensic speaker comparisons, where often low-quality speech samples are assessed in terms of the typicality and similarity of the speaker-dependent features they contain. In forensic phonetics, speaker-specificity is defined as the ratio of between- to within-speaker variation. The present work contributes to both fields by checking whether previously reported linguistic effects for fricatives are present in spontaneous telephone dialogues, which is a relevant speech style and channel both for everyday communication and forensic speaker comparisons, and whether these linguistic effects interact with the

amount of speaker information for two highly frequent fricatives in Dutch.

2.1.1 Within-speaker variability in fricative production

2.1.1.1 Labialization

Within speakers, it has been shown that fricative acoustics vary systematically as a function of phonetic context. Predominantly, anticipatory lip-rounding has repeatedly been shown to lower resonance frequencies in fricatives (e.g., Bell-Berti & Harris, 1979; Koenig et al., 2013). Anticipatory lip-rounding lowers the resonance frequencies in fricatives because the lip protrusion associated with the lip movement lengthens the anterior cavity. Notably, neighboring labial consonants such as English bilabial /w/ and /p/ also seem to display a lowering effect on /s/ spectra (Munson, 2004), even though the lip movement for /p/ is better described as lip closure rather than lip-rounding. This implies that labial closure also lengthens the anterior cavity to some extent.

Regarding within-speaker variation, Munson (2004) hypothesized that variability in degree and timing of the labial co-articulation in /s/ would result in increased within-speaker variation. Replicating earlier research, Munson (2004) reported that /s/ has lower resonance frequencies when followed by rounded /u/ versus non-rounded /a/ and when followed by rounded /w/ versus vowels /a, u/, with labial – but not rounded – /p/ falling in-between. The results for the within-speaker variation, however, only showed increased within-speaker variation for /s/ followed by /w/ and not for /s/ followed by /u/ compared to when it is followed by /a/. It is probable that the lip-movements for /w/ versus /u/ and /p/ constitute different labial movements. Other work has shown that there are different types of labialization, e.g., different lip-area size involved in labialization for postalveolar fricatives /ʃ, ʒ/ versus approximant /w/ (Toda et al., 2003). It is therefore possible that the labial movement for /w/ is more sensitive to within-speaker variation than the

labial movements for /u/ and /p/. Alternatively, /s/ followed by /w/ may display more within-speaker variation due to differences in articulatory timing between /s/ from consonant clusters versus consonant-vowel sequences. Munson (2004) did not report on the between-speaker variation, therefore, no information on the speaker-specificity of fricatives in labialized context is available. Given that the degree and timing of labial co-articulation in fricatives might vary between speakers (Perkell & Matthies, 1992), fricatives with labialized context might also constitute relatively speaker-specific locations.

2.1.1.2 Speech effort

Articulatory strengthening (hyperarticulation) or weakening (hypoarticulation) also affect fricative acoustics within speakers. Generally speaking, it has been shown that there are articulatory strong and weak locations in speech. Whereas the initial edges of prosodic domains such as phrases and words are generally found to be locations of articulatory strengthening (Cho & McQueen, 2005; Fougeron, 2001), the final edges of syllables, i.e., codas, are generally found to be locations of articulatory weakening compared to syllable onsets (Ohala & Kawasaki, 1984). For fricatives as a group, American English coda fricatives are found to be less identifiable (Redford & Diehl, 1999), and to have a lower intensity and a delayed and lower air pressure peak than onset fricatives (Solé, 2003). However, studies that consider different fricatives separately show inconsistent results with regards to coda reduction for /s/ specifically; Redford & Diehl (1999) found coda reduction in duration in American English /s/, but not in intensity or spectral mean. Furthermore, they reported that, whereas consonant classification using linear discriminant analysis overall showed more accurately classified onsets than codas, this was not the case for /s/, where there was a reverse tendency. This lack of coda reduction for /s/ was replicated for German, where spectral mean for codas was not lower, but slightly higher than for onsets (Cunha & Reubold, 2015). Although there was no reduction effect for German /s/ in coda position, Cunha & Reubold (2015) found that codas display higher variability than onsets and that /s/ in de-accented syllables displays higher variability than /s/ in

accented syllables. In other words, they reported more variability, but no reduction, in articulatory weak locations. Overall, reports on reduction in fricative acoustics are inconsistent, particularly with regards to /s/, but studies generally report more variability for articulatory weak positions. It is unclear whether that variability is within- or between-speakers.

2.1.1.3 Segmental effects

From the somewhat conflicting results reported above, it seems that not all fricatives reduce in the same manner or to the same extent. Rather, reduction seems to be constrained by specific production requirements (Recasens, 2004). This means that features that have high production requirements for a particular speech sound are more resistant to co-articulation and reduction than features that have low production requirements for a particular speech sound. For example, in fricatives /s/ and /x/, the resistance to anticipatory labialization might be low because there are no production requirements for the lips in /s/ and /x/. Tongue front and dorsum in the production of /s/, on the other hand, are relatively resistant to co-articulation and reduction due to the production necessity of tongue front raising and dorsum lowering for this fricative (Recasens & Dolorspallarè, 2001). Speakers might vary in their articulatory timing, degree of co-articulation, and their reduction of specific features. This means that some speakers may be more sensitive to certain co-articulatory effects than others. As a result, the acoustic realizations of /s/ and /x/ might be more context-dependent in some speakers than others. It is therefore possible that highly context-dependent realizations, such as /s/ and /x/ in labialized context, display high between-speaker variability.

2.1.1.4 Other linguistic effects

Speech style can also affect fricative acoustics within speakers. Maniwa et al. (2009) compared clearly spoken fricatives to fricatives in a conversational speech style in American English and found that clearly spoken fricatives had longer duration, higher resonance frequencies, and – surprisingly – lower relative amplitude. Moreover, individual speakers

used different strategies for producing clear speech, which were not related to speaker sex/gender. This implies that different patterns of within- and between-speaker variation may be expected in clearly spoken speech versus conversational speech. It therefore seems important to extend research on speaker variation to include conversational speech styles.

2.1.2 Between-speaker variability in fricative production

Between speakers, anatomical/physiological and social effects have been observed in fricative acoustics. Regarding anatomical/physiological variation, fricative acoustics can vary as a function of the shapes and sizes of the articulators and cavities (Stevens, 2000, pp. 411–412). In practice, this type of variation in fricative acoustics has often been observed between males and females; fricatives produced by females have higher resonance frequencies than by males, which is often explained as resulting from anatomical differences between female and male speakers (e.g., Jongman et al., 2000; Schwartz, 1968). This difference in production is perceivable and meaningful to listeners, as speaker sex can be perceived from isolated voiceless fricatives (Ingemann, 1968; Schwartz, 1968).

From sociolinguistics, there are known between-speaker factors that affect fricative acoustics. For example, there are well-attested effects of gender identity and sexual orientation on /s/ spectra that are not associated with anatomical/physiological differences but rather with production strategies, i.e., learned behavior (e.g., Bang et al., 2017; Fuchs & Toda, 2010; Munson et al., 2006). Social class may also affect fricative spectra; Stuart-Smith (2007) found that English working-class females could be grouped with working-class males, rather than with higher-class females, on several spectral features from /s/. When looking at social identity on a larger scale, such as ethnolect, dialect, and language communities, variation in fricative spectra is also observed. For example, the so-called 'Moroccan flavored Dutch' ethnolect is known for a retracted [s] realization that resembles [ʃ], i.e., sibilant palatalization,

in certain phonetic contexts (Mourigh, 2017). Another example is the regional variation for Dutch fricative /x/, which is produced with velar place of articulation (and thus higher resonance frequencies) in Flanders and Southern regions of the Netherlands, and with uvular place of articulation – often accompanied by uvular scrape, i.e., uvular trill – in Northern regions of the Netherlands (Van der Harst et al., 2007).

Given that group-level speaker characteristics such as sex/gender and ethnolect are associated with shared acoustic features, it seems important to eliminate as much group-level variation as possible when focused on characterizing individual speakers. Moreover, in forensic casework, it is deemed necessary to compare speakers amongst a reference population of similar speakers, i.e., speakers of the same sex/gender and dialect. This work therefore chose to limit itself to speakers from the same sex/gender and dialect.

2.1.3 Speaker-specificity and linguistic context

It is currently unclear how speaker-specificity is dependent on linguistic context. Given that speaker-specificity is a ratio of between-speaker to within-speaker variation, speech samples need high between-speaker variation and low within-speaker variation to be speaker-specific. There are some linguistic contexts that might facilitate such environments, and thus help listeners extract speaker information.

2.1.3.1 Segmental effects on speaker-specificity

Previous work has shown that some individual speech sounds are more speaker-specific than others. For example, vowels are found to be more speaker-specific than consonants (Van den Heuvel, 1996, pp. 145-146). Within the class of consonants, fricative /s/ – one of the speech sounds investigated in the present work – is found to be relatively speaker-specific. In Dutch read speech, /s/ was ranked below vowels and nasals, but above /r/ and plosives in terms of speaker-specificity (Van den Heuvel, 1996, pp. 72). In English read speech, /s/ along with nasal /m/

are ranked above nasals /n/ and /ŋ/, and liquid /l/ (Kavanagh, 2012, pp. 387-388). Studies on the speaker-specificity of fricatives that are not /s/ – such as the dorsal fricative /x/ also examined in the present work – are rare. Perceptually, differences in the amount of speaker-dependent information have also been observed. Comparing speaker sex identification between fricative sounds, Ingemann (1968) found that listeners can identify speaker sex from isolated back fricatives [h, χ, x] but not from isolated front fricatives [θ, f, ɸ]. Front fricatives [s, ʃ] broke this pattern; speaker sex identification from these sounds was also above chance.

2.1.3.2 Speech effort and speaker-specificity

Articulatory strong locations are locations in speech that are produced with more vocal effort, e.g., onsets and stressed syllables. They are often argued to constitute canonical speech, and might therefore be characterized by low within-speaker variation. If these locations are not also characterized by low between-speaker variation, they might be relatively speaker-specific. Evidence supporting this hypothesis comes from a finding that speakers were characterized more accurately using vowels receiving sentence stress – which are generally considered to be articulatory strong locations – than vowels without sentence stress (McDougall, 2004). Other evidence that suggests that articulatory strong locations contain more speaker-dependent information can be found in Heeren (2018), who showed that the vowel /a/ sampled from spontaneous speech gave higher speaker classification scores in content than in function words. Content words are generally also found to be articulatory strong locations, which is evidenced by studies that found reduction in vowels sampled from function words relative to content words (Shi et al., 2005; Van Bergem, 1993, pp. 38–39).

Alternatively to articulatory strong locations displaying high speaker-specificity, articulatory weak locations such as codas and highly context-dependent segments, e.g., fricatives with labial neighbors, might be characterized by high between-speaker variation and may therefore also display high speaker-specificity. Based on their work on formant and intensity dynamics, He et al. (2017; 2019) hypothesize that speakers may

have more articulatory freedom in speech locations that are less constrained by articulatory targets, resulting in higher between-speaker variation in these locations. This is sometimes also referred to as variation due to target undershoot. They showed that both intensity dynamics (He & Dellwo, 2017) and formant dynamics (He et al., 2019) show more between-speaker variation in negative than in positive dynamics. Negative dynamics were defined as the intensity and formant slopes from the syllable's peak to the following trough, which are the parts of syllables associated with mouth-closing gestures. They suggest that the mouth-opening gestures (positive dynamics) might be more restricted by articulatory targets.

Previous studies thus indicate that some linguistic contexts affect the amount of within- and between-speaker variation. Namely, articulatory strong locations seem to have relatively low within-speaker variation and articulatory weak locations seem to have relatively high between-speaker variation. However, for fricatives, it is unclear how articulatory weak versus strong positions affect the speaker-specificity.

2.1.4 Fricatives in Dutch telephone speech

2.1.4.1 Dutch fricatives

The Standard Dutch fricative inventory contains eight fricatives (see Table 2.1). The present study focuses on two voiceless fricatives: the laminal alveolar /s/ and the dorsal fricative /x/ (for notation sake, the dorsal fricative – which can have a velar [x] or uvular [χ] place of articulation, will be denoted with symbol 'x'). Fricatives /s/ and /x/ were selected because they are highly frequent in syllable onsets and to a slightly lesser extent in coda position in Dutch (Baayen et al., 1993), which makes them suitable speech sounds to analyze in spontaneous speech samples.

Table 2.1: *Standard Dutch fricative inventory (cf. Gussenhoven, 1999). Fricatives in parentheses are restricted to loanwords and to alveolar fricatives with place assimilation from a following* [j] *(e.g., jas 'coat'* [jɑs]; *jasje 'little coat'* [jɑʃə]*).*

	Voiceless	Voiced
Labiodental	f	v
Alveolar	s	z
Post-alveolar	(ʃ)	(ʒ)
Dorsal	x/χ	
Glottal		ɦ

Fricative sounds are produced with a narrow constriction which results in noise generated by turbulence (Stevens, 2000, p. 379). The resonance frequencies of fricatives are mainly determined by the size of the cavity anterior to the narrow constriction (Stevens, 2000, pp. 398-403). Whereas the Dutch laminal alveolar fricative /s/ has a relatively small anterior cavity and therefore high resonance frequencies, Dutch dorsal fricative /x/ has a medium to large anterior cavity (depending on a velar or uvular place of articulation) and therefore much lower resonance frequencies. Fricative /s/ is reported to have a spectral center of gravity of around 4.8 kHz in Standard Dutch read speech (Ditewig et al., 2019) and fricative /x/ is reported to have a spectral peak of around 1.7 kHz in Standard Dutch read speech (Van der Harst et al., 2007).

2.1.4.2 Telephone filter

Most acoustic reports on /s/ and /x/ are based on studio-recorded read speech. However, this speech style is not representative of everyday communication nor of forensic speaker comparisons. It is unclear

whether acoustic-phonetic and indexical information in /s/ and /x/ can be captured in spontaneous telephone dialogues. Particularly in the context of forensic speech comparisons, telephone speech is highly relevant compared to studio-recorded (read) speech, as wiretapping telephone conversations from criminal suspects is common in police investigations in the Netherlands (Odinot et al., 2010, p. 82). Using higher-quality, non-telephone speech may misrepresent what listeners may use in speech perception in daily conversation as well as what is possible for forensic speaker comparisons.

Telephone signals have a limited frequency bandwidth. For example, the landline telephone dialogues worked with in this study have a bandwidth of 340 - 3400 Hz. Given that the spectral energy for Dutch /s/ is concentrated around 4.8 kHz (Ditewig et al., 2019), this means that the spectral energy for fricative /s/ mostly resides above the upper limit of this bandwidth (see Figure 2.1a). It is therefore possible that both linguistic information and speaker information from /s/ are (partly) lost in telephone speech. The spectral energy for back fricative /x/, on the other hand, falls mostly within the telephone bandwidth (see Figure 2.1b).

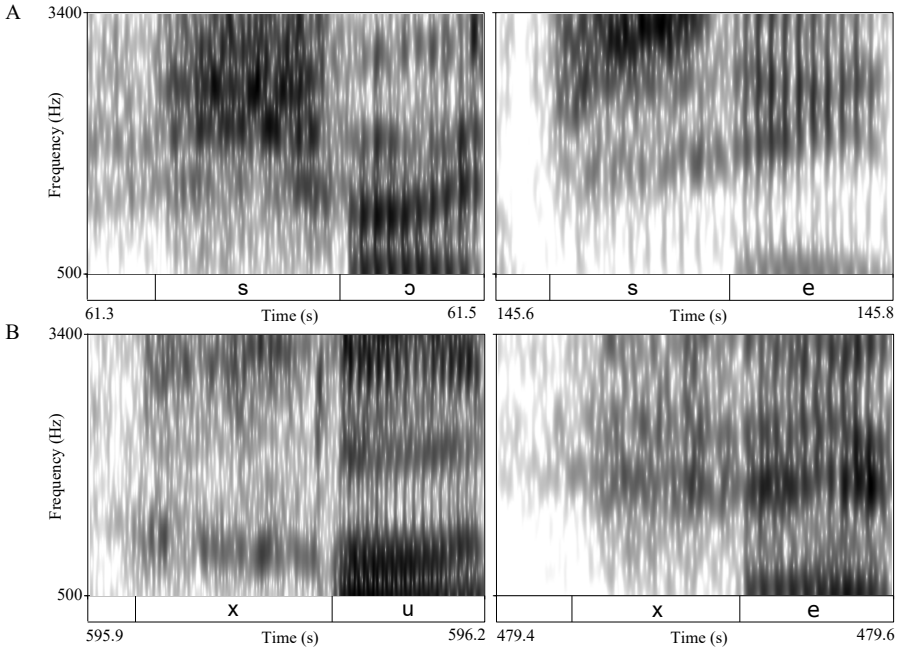

Figure 2.1: *Spectrograms for onset fricatives in labial and non-labial contexts spoken by a male speaker of Standard Dutch over a 500-3400 Hz bandwidth. A: Onset /s/ from words* soort *('sort',* /sɔːrt/) *and* cd *('cd',* /sede/*). B: Onset /x/ from words* goed *('good',* /xut/) *and* geen *('no',* /xen/*).*

Telephone speech also has other limitations that have to be considered in an acoustic analysis. Regarding signal-related transformative qualities, the lower formants may display an upward shift. Particularly F1 values display a large shift of around 14% on average, whereas higher formants generally remain unaffected (Künzel, 2001; for mobile signals, this number is 29% on average, with some F1 values rising by up to 60%: Byrne & Foulkes, 2004). Moreover, when this signal-related shift is paired with speaker-behavior such as holding the phone between the cheek and shoulder, these upwards shifts are amplified

(Jovičić et al., 2015). Additionally, the signal-related qualities of telephone speech are accompanied by distinct speech behavior. For example, speakers often increase their vocal effort, possibly to adjust for increased background noises from variable environments. This effect is generally described as the Lombard effect (e.g., Junqua, 1993).

2.1.5 Research questions and hypotheses

The main research question of the present study is whether the amount of speaker information in speech sounds is dependent on their linguistic context. Two fricatives were selected based on their frequency of occurrence in Dutch: alveolar /s/ and dorsal /x/. These fricatives were sampled from spontaneous telephone dialogues, which are representative of everyday communication as well as forensic voice comparisons. To answer the main research question, we first tested whether linguistic context factors (syllabic position, perseverative and anticipatory labialization) – which have been shown to affect fricative acoustics in read broadband speech – can be replicated in spontaneous telephone speech. Secondly, we examined whether speaker-classification models for the two fricatives show effects of linguistic context. In this second step, the effect of the speech sound (/s/ versus /x/) and the contribution of individual acoustic measurements on speaker-classification were also examined.

2.1.5.1 Linguistic effects

Based on previous research on read broadband speech (e.g., Bell-Berti & Harris, 1979; Koenig et al., 2013), we hypothesized that perseverative and anticipatory labialization would lower fricative spectra, but that this might not be measurable for /s/ because the spectrally-defining characteristics for /s/ mostly reside over the upper limit of the telephone bandwidth. Spectrally-defining characteristics for dorsal fricative /x/, on the other hand, should fall within the telephone bandwidth. The literature is not clear on the effect of syllabic position, particularly for /s/.

2.1.5.2 Speaker classification

In the second step, we hypothesized that there might be a segmental effect in speaker classification. Namely, /x/ might be more speaker-specific than /s/, because the telephone channel cuts off most spectral energy for /s/ but not /x/. Regarding the performance of acoustic measures, previous findings report that spectral center of gravity and standard deviation were the most speaker-discriminating features (e.g., Kavanagh, 2012). We therefore predicted that most speaker-specific information might be found in spectral as opposed to temporal or amplitudinal measures. Regarding the speaker variation as a function of linguistic context, we hypothesized that articulatory strong locations (onsets and fricatives with non-labial neighbors) are characterized by low within-speaker variation and that articulatory weak locations (codas and fricatives with labial neighbors) are characterized by high between-speaker variation. However, there were no clear expectations for speaker-specificity, which equals the ratio of between- to within-speaker variation.

2.2 Methodology

2.2.1 Materials

Spontaneous telephone dialogues available in the Spoken Dutch Corpus (Oostdijk, 2000) were used to investigate the speaker-specificity in the realization of fricatives /s/ and /x/. The telephone dialogues were obtained via a switchboard. No information on the task is available, but from the recordings' content it was inferred that speakers were located in their home environment (deduced from background noises such as a crying baby or a barking dog) and were asked to converse for around ten minutes on any topic of their choosing. One to four telephone conversations (M = 1.88, SD = 0.96) – with different interlocutors – are available for each speaker in the corpus. All available conversations for a speaker were included.

Given the overrepresentation of male speakers in forensic voice comparisons, only male speakers were analyzed in this study[1]. Male speakers were included if the metadata from the corpus indicated that they were between 18 and 50 years old and if they were reported to be speakers of Standard Dutch. Speakers were excluded if the first author judged them to be speakers of non-standard Dutch. For the remaining 66 male speakers of Standard Dutch (age range = 21 - 50, M = 36.5, SD = 7.3), a total of 3,331 /s/ tokens and their adjacent contexts as well as 3,491 /x/ tokens with their adjacent contexts were first automatically segmented and provided with a broad phonetic transcription using the orthographic transcript available with the corpus. These were then manually validated by the first author. When interference such as laughter, overlapping speech from the interlocutor, or background noise showed up in the signal, tokens were excluded. Fricative tokens occurring in context with a creaky phonation were not excluded, as previous research has shown that /s/ spectra are relatively stable against creakiness (Hirson & Duckworth, 1993). Tokens were labelled as onsets (/s/: N = 1,359; /x/: N = 1,657), codas (/s/: N = 1,532; /x/: N = 1,453), or ambisyllabic (/s/: N = 440; /x/: N = 380). The latter category, containing tokens that cannot be categorized as either onsets or codas (e.g., was ook 'was also' [wɑsoːk]), was excluded from analysis.

As reviewed above, labialization of adjacent context affects fricative spectra. To test whether the measures extracted from telephone speech are sensitive to contextual labialization, preceding and following context was furthermore labelled as labial or non-labial. Rounded vowels /u, ɔ, o, ø, y, ʏ/, (partially) rounded diphthongs /œy, ɑu/ (cf. temporal patterns of lip-rounding: Bell-Berti & Harris, 1982), and bilabial consonants /p, b, m/, were considered to be labial. Labiodental consonants /f, v, ʋ/ were not coded as labial because the teeth-to-lip movement in these sounds does not involve lip-rounding or closure, but rather eliminates the anterior cavity and can therefore not be assumed to

[1] It is unclear from the metadata from the Spoken Dutch Corpus how the label 'male' was assigned to speakers. It is assumed here that 'male' refers to biological sex.

have the same lowering effect on the spectrum. Speakers with fewer than 25 tokens per fricative sound were excluded, which excluded 23 speakers and left a total of 43 speakers with a sufficient number of tokens for both /s/ and /x/. The resulting numbers of tokens per factor level are presented in Table 2.2.

Table 2.2: *Totals, and means, standard deviations, and ranges for numbers of /s/ and /x/ tokens by speaker (N = 43) and by linguistic context factor level.*

			Syllabic Position		Left Context		Right context	
		Total	Onset	Coda	Non-labial	Labial	Non-labial	Labial
/s/	Total	2,346	1,066	1,280	1,846	500	1,903	443
	M	55	25	30	43	12	44	10
	SD	19	11	11	16	5	15	7
	range	25-108	9-63	15-78	20-88	3-22	24-88	1-35
/x/	Total	2,820	1,460	1,360	2,336	484	2,250	570
	M	66	34	32	54	11	52	13
	SD	26	13	15	23	6	22	7
	range	27-124	11-67	9-73	20-106	3-29	23-100	3-31

2.2.2 Acoustic analysis

The telephone dialogues available in the Spoken Dutch Corpus have a sampling frequency of 8 kHz with an 8-bit resolution and were originally filtered at a bandwidth of 340 – 3,400 Hz. There are separate channels

for the two speakers in each telephone conversation. A low-frequency cut-off of 500 Hz was used to reduce the influence of background noise and (partial) voicing. For each fricative token, seven measures were taken in Praat version 6.0.46 (Boersma & Weenink, 2020). First, duration (DUR; in milliseconds, ms) was computed from fricative onset to fricative offset as characterized by the presence of aperiodic fricative noise, which was then used to establish the middle 50% of each fricative over which the static spectral measures were taken. The static spectral measures consisted of two spectral moments – spectral center of gravity (CoG) and standard deviation (SD) – and spectral tilt. After filtering the fricative tokens to the 0.5 - 3.4 kHz band (band pass Hann filter, smoothing = 100 Hz), the center of gravity and the standard deviation (CoG and SD; in Hertz, Hz) were computed from the spectrum determined over the mid-50% of the fricative, using power spectrum weighting. Although the formant-like structure of spectral energy for Dutch /x/ (see Figure 2.1b) might be captured better by more complex measures such as discrete cosine transforms (DCT), the relatively simple measure CoG has been shown to capture between-speaker variation such as regional variation (Harst et al., 2007)[2].

Spectral tilt (TILT) was measured to reflect vocal effort as an alternative to absolute amplitudinal measures, and computed from the long-term average spectrum determined over the mid-50% of the fricative (bin = 1 Hz) on a logarithmic frequency scale (dB/decade), using a least-squares fit. A decade is a step on the frequency scale with the power of 10, i.e., 1 Hz, 10 Hz, 100 Hz, etc. Mean amplitude (AMP; in dB) was measured over the full fricative's duration and normalized by speaker through Z-transformation.

Additional to the static measures, dynamic spectral measures were computed by measuring spectral CoG in non-overlapping 20%-portions of the entire fricative's duration. Coefficients from quadratic

[2] To pilot our data and acoustic measures, all /x/ tokens were auditorily labelled on place of articulation (velar versus uvular) and CoG was shown to predict place of articulation with a cross-validated accuracy of 83.9% in a linear-discriminant analysis (LDA). We therefore expect CoG to adequately capture the linguistic effects and speaker-dependent spectral characteristics in fricative acoustics.

polynomial equations over the five resulting data points per fricative token constituted our dynamic measures for analysis. Both cubic and quadratic models to the data were estimated; likelihood-ratio tests showed no significant difference between these two models (/s/: χ^2 (1) = 0.96, p = .33; /x/: χ^2 (1) = 0.11, p = .74). The simpler quadratic function ($\gamma = \beta_0 + \beta_1 x + \beta_2 x^2$) was chosen as the fewer coefficients reduced the number of predictors in further modelling. The intercept (β_0) was excluded because it correlated highly with the static CoG measure (/s/: r = .95, N = 2,346, p < .001; /x/: r = .96, N = 2,820, p < .001), resulting in only a linear (CoG^{linear}) and quadratic ($CoG^{quadratic}$) coefficient.

2.2.3 Statistical analysis

The statistical analysis consisted of two parts: (1) linear mixed-effect modelling was used to check whether linguistic factors affected /s/ and /x/ acoustics in spontaneous telephone speech, and (2) multinomial logistic regression was used to investigate whether the amount of speaker information in /s/ and /x/ varied as a function of syllabic position and labial co-articulation. Additionally, segmental effects as well as the relative importance of acoustic measures in speaker classification were estimated from the regression model. A more traditional measure for speaker-specificity, called the Speaker-Specificity Index (SSI), was also computed for all acoustic variables to assess its relationship with the regression modelling results. The SSI relates the between-speaker variance to the within-speaker variance (Van den Heuvel, 1996).

2.2.3.1 Linear mixed-effect modelling: Linguistic effects

In the first part of the analysis, the effects of linguistic context factors on acoustic measures were investigated by means of linear mixed-effect modelling (LMM) in R version 3.5.1. (R Core Team, 2018). First, a model with maximal fixed and random structure was built for each dependent variable, i.e., each acoustic measure (CoG, SD, TILT, DUR, and AMP). This maximal model contained six fixed factors: three main factors for Syllabic Position (CODA, ONSET; sum coded), Left Context (NON-LABIAL, LABIAL; dummy coded), and Right Context (NON-

LABIAL, LABIAL; dummy coded) and three one-way interactions between these main factors. One-way interaction terms were included because Right Context for factor level CODA contained only consonants and pauses coded for labialization (see section 2.2.1). Because labial consonants possibly produce attenuated coarticulation effects on neighboring fricatives compared to labial vowels (Munson, 2004), an interaction between the Left and Right Context factors and Syllabic Position might be expected. The random structure of the maximal model contained random intercepts for Word and Speaker, as well as random slopes by Speaker over all three fixed factors. This means that Syllabic Position and Left and Right Context were added to the model as both within-speaker and between-speaker factors.

All fixed and random terms in the maximal model were tested via model comparisons. First, a full model with maximal random structure was built by restricted maximum likelihood (REML) estimation (Barr et al., 2013). Next, stepwise deletion was used to reduce the random structure of the model, given this led to a better-fitting model as estimated by the Bayesian information criterion (Bates et al., 2015). Model fit was assessed through inspection of the residuals and duration was log-transformed (base = 10) for a better model fit. The *p*-values were generated empirically with bootstrapping using function *mixed()* from R package 'afex' (Singmann, 2019). This function derives a mean *p*-value for a fixed effect by comparing the optimal model with a model without the fixed effect in question for a specified number of data simulations (N = 10,000). The significance level (α = .05) of fixed effects was adjusted via Bonferroni correction (α = .05/(5*2)), to account for the fact that the different acoustic measures (N = 5) and fricative sounds (N = 2) were extracted from the same dataset of speakers.

Lastly, the results were tested in the presence of two prosodic factors that would possibly confound results obtained by previous modelling. Models were rebuilt including factors for Phrasal Position (INITIAL, MEDIAL, FINAL; sum coded) and Word Stress (NON-STRESSED, STRESSED; sum coded) to see if results were maintained. For Word Stress, only tokens from content words (nouns, verbs, adjectives, and adverbs) were labelled for word stress, as function words can have stressed syllables only in special circumstances (Selkirk, 1996). This

resulted in the exclusion of 16% of the data for /s/ and 12% of the data for /x/. Results from these latter models are not presented because these extended models did not change the results obtained by earlier models, although exact statistics were slightly different.

2.2.3.2 Multinomial logistic regression: Speaker classification

Multinomial logistic regression (MLR) was used to test which linguistic context factors and acoustic measures significantly predicted the dependent variable Speaker. Function *buildmultinom()* from R Package 'buildmer' (Voeten, 2020) was used to automatically build and then reduce the maximal MLR model by estimating each predictor with backward stepwise selection using likelihood-ratio tests. Highly correlating predictors ($r > .70$) were excluded, which resulted in the exclusion of TILT because it correlated highly with CoG (/s/: $r = .76$, $N = 2{,}346$, $p < .001$; /x/: $r = .91$, $N = 2{,}820$, $p < .001$). This means the maximal MLR model to predict SPEAKER contained 27 predictors: six acoustic measures (CoG, SD, AMP, DUR, CoG^{linear}, and $CoG^{quadratic}$), three linguistic factors (Syllabic Position, Left Context, and Right Context), and 18 one-way interactions between the acoustic measures and linguistic factors.

In a second step, the optimal model obtained by function *buildmultinom()* was inspected to see which fricative contained more speaker-dependent information and which combinations of acoustic measures and linguistic context factors affect speaker classification predictions. The predicted speaker classification of factor levels was compared, i.e., for Syllabic Position, speaker classification of codas is compared to onsets. This was achieved by splitting the data on factor level and then predicting speaker classification on the resulting two datasets using the best-fitting model acquired in the previous step. This was done for factor levels from all linguistic context factors that were included in the best-fitting models. Secondly, acoustic measures and their interactions with linguistic context factors were excluded from the best-fitting model one at a time to assess the relative importance of each acoustic measure.

2.3 Results

2.3.1 Linguistic effects

2.3.1.1 Labialization

Linear mixed-effect modelling results for /s/ and /x/ are summarized in Table 2.3. For /s/, there were no effects for Left Context. However, /s/ tokens with labial Right Context have lower SD, shorter duration in codas, and – opposite to what we hypothesized – higher CoG. Contrary to results for /s/, there is a clear labialization effect for /x/. When Left Context is labial, /x/ CoG lowers and spectral tilt decreases, i.e., there is less energy at higher frequencies. When Right Context is labial, CoG lowers (although this effect is larger for onsets), spectral tilt decreases, and amplitude decreases. The interaction between Left and Right Context for spectral tilt indicates that spectral lowering is attenuated by 4.7 dB per decade when both Left and Right Context are labial.

2.3.1.2 Syllabic position

Results in Table 2.3 show that /s/ onsets have higher CoG, higher positive tilt, i.e., more high-frequency energy, higher amplitude and shorter duration than codas. In other words, all measures from /s/ except duration show coda reduction. Note also that the spectral tilt intercept in Table 2.3 is a positive value, i.e., there is no energy drop-off but an increase in higher frequencies. This is expected for /s/ because, within the telephone band, all the spectral energy is expected to reside in the higher frequencies.

Fricative /x/ also showed coda reduction; onsets have higher amplitude than codas. Contrasting our data for /s/, tilt for /x/ shows a negative value. This shows that, whereas there is no energy drop-off for high-frequency /s/, there is an average energy drop-off of 7.8 dB per decade for /x/.

Table 2.3. *Summary of fixed effects from linear mixed-effect modelling for /s/ (N = 2,346) and /x/ (N = 2,820) with Kenward-Roger degrees of freedom approximation. Reference values are CODA for Syllabic Position and NON-LABIAL for Left and Right Context. Empty cells indicate that the factor was not included in the best-fitting model. The p-values for fixed effects were obtained empirically by bootstrapping (N simulations = 10,000). Non-significant effects are in italic.*

		/s/			/x/		
	Fixed effects	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*
CoG	(intercept)	2541	37	68.2	1648	34	48.6
	SyllPos: ONSET				-5	13	*-0.4*
	Left Context: LABIAL	-15	28	*-0.5*	-192	25	-7.8
	Right Context: LABIAL	86	19	4.6	-281	39	-7.3
	SyllPos × Right Context				-103	30	-3.5
SD	(intercept)	603	18	32.7	599	14	43.0
	SyllPos: ONSET				-6	4	*-1.6*
	Left Context: LABIAL				27	9	3.0
	Right Context: LABIAL	-42	9	-4.7	-7	20	*-0.4*
	SyllPos × Left Context				-54	9	-6.1
	SyllPos × Right Context				-57	9	-6.3
TILT	(intercept)	17.3	1.5	11.8	-7.8	1.3	-6.2
	SyllPos: ONSET				-0.6	0.4	*-1.4*
	Left Context: LABIAL				-7.4	1.1	-6.6
	Right Context: LABIAL	2.1	0.5	4.6	-8.6	1.3	-6.7
	SyllPos × Right Context				-3.9	1.0	-4.0
	Left × Right Context				4.7	1.8	2.5
AMP	(intercept)	0.04	0.03	*1.5*	0.01	0.03	*0.3*
	SyllPos: ONSET	0.15	0.03	5.5	0.24	0.03	8.1
	Right Context: LABIAL				-0.26	0.07	-3.5
DUR	(intercept)	1.95	0.01	235	1.92	0.01	212.8
	SyllPos: ONSET	-0.03	0.01	-5.0	0.01	0.01	*1.1*
	Right Context: LABIAL	-0.07	0.01	-6.2	-0.02	0.01	*-1.2*
	SyllPos × Right Context	0.06	0.01	5.2	0.09	0.01	6.5

2.3.1.3 Intermediate discussion

Linear mixed-effect modelling has indicated that both /s/ and /x/ are affected by our fixed factors for several measures, but not in the same way. Figure 2.2 illustrates the differences in effects of context labialization on CoG between the two fricatives under study. Whereas /x/ CoG lowers when context is labial, this is clearly not the case for /s/. As hypothesized, this may be due to the telephone bandwidth. If the speaker-specificity is sensitive to linguistic context factors, the acoustic results would predict stronger effects for /x/ than for /s/ in the speaker-classification analysis, since /x/ shows more context-dependent acoustic variation than /s/.

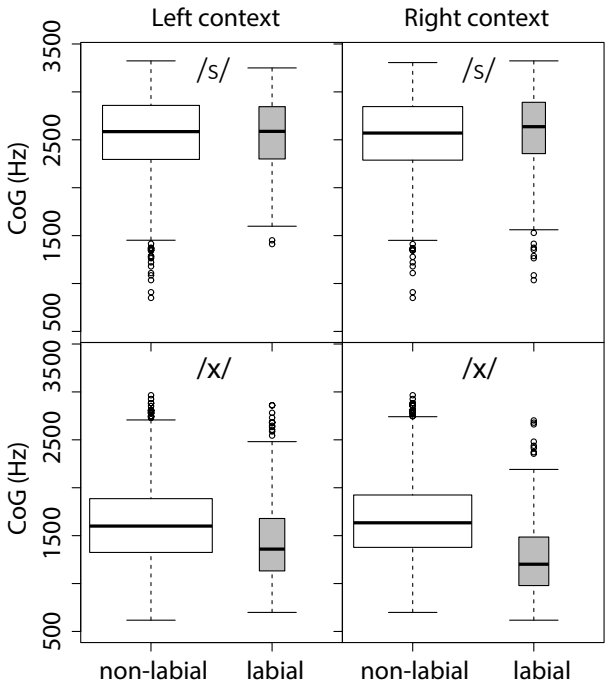

Figure 2.2: *Boxplots for CoG by fricative sound, Syllabic Position, and Left and Right Context labialization. The width of the box represents the number of cases included in the MLE and MLR analyses (see Table 2.2 for exact numbers).*

2.3.2 Speaker classification

2.3.2.1 Segmental effects

For both /s/ and /x/, the best-fitting model to predict Speaker (/s/: $N = 43$, $n = 2{,}346$; /x/: $N = 43$, $n = 2{,}820$) included all acoustic measures and all linguistic context factors as significant predictors. The interactions that were included as predictors are indicated in Table 2.4. The model for /s/ had a speaker-classification accuracy of 19.5% against a chance level of 2.3%. The model for /x/ had a speaker-classification accuracy of 18.4% (chance = 2.3%). This means that, despite the limited telephone band, speaker classification from fricative /s/ acoustics was better than from fricative /x/ acoustics.

Table 2.4: *Included one-way interactions in the optimal MLR models for /s/ and /x/.*

Predictor	Syllabic position		Left context		Right context	
	/s/	/x/	/s/	/x/	/s/	/x/
CoG	✓	✓	✓	✓	✓	✓
SD		✓	✓	✓		✓
CoG^{linear}	✓	✓			✓	✓
$CoG^{quadratic}$	✓				✓	✓
AMP	✓	✓			✓	✓
DUR	✓	✓	✓		✓	✓

2.3.2.2 Contribution of individual acoustic measures

The decreases in speaker-classification accuracy when a single acoustic measure and its interactions with linguistic context factors were dropped from the model are presented in Table 2.5. For example, excluding CoG and the interactions between CoG and linguistic context factors from the best-fitting model for /s/ resulted in a decrease in speaker-classification accuracy from 19.5% (for the optimal model) to 13.9%, which makes a decrease of 5.6%. As can be seen in Table 2.5, CoG and SD were relatively important measures for speaker classification. Moreover, measures contributed to speaker classification in comparable ways across fricatives. The contribution of acoustic measures to the speaker classification from the MLR model is accompanied by the more traditional SSI measure; these more or less mirror the relative ranking from the MLR model.

Table 2.5: *Speaker-classification accuracy decreases (in %) per acoustic measure relative to the full models' speaker-classification accuracy of 19.5% for /s/ and 18.4% for /x/ and speaker-specificity index (SSI) per acoustic measure for /s/ and /x/.*

	/s/		/x/	
Excluded measure	*Δacc*	*SSI*	*Δacc*	*SSI*
CoG	5.6	0.56	4.5	0.26
SD	4.5	0.63	3.4	0.31
DUR	1.9	0.07	2.1	0.10
CoG^{linear}	0.9	0.07	1.6	0.06
$CoG^{quadratic}$	1.3	0.08	1.2	0.07
AMP	1.1	0.14	0.7	0.06

2.3.2.3 Linguistic effects

Per linguistic context, speaker-classification accuracies were similar (see Table 2.6), but there seems to be a small, yet systematic, advantage for articulatory weak locations, i.e., codas and tokens with labial co-articulation.

Table 2.6: *Speaker classification accuracies (in %) per fricative sound and per linguistic context factor level (chance level = 2.3%).*

	Linguistic context	/s/	/x/
Syllabic Position	Total	19.5	18.4
	Onset	19.5	18.2
	Coda	19.5	18.6
Left Context	Non-labial	18.3	18.5
	Labial	24.2	18.8
Right Context	Non-labial	18.5	17.6
	Labial	18.8	21.4

The small advantage in speaker classification for articulatory weak locations was examined to see whether it was due to an increase in between-speaker variation. The between- and within-speaker variances per linguistic context factors are presented for the most-contributing measure in speaker-classification for /x/, i.e., CoG (see Figure 2.3). Consistent with the SSIs reported in Table 2.5, Figure 2.3 shows that the within-speaker variance is consistently higher than the between-speaker variance across all linguistic contexts. Additionally, as hypothesized, the between-speaker variance seems to be increased in articulatory weak locations compared to strong locations. Against expectation, the within-speaker variation seems to be decreased in articulatory weak locations.

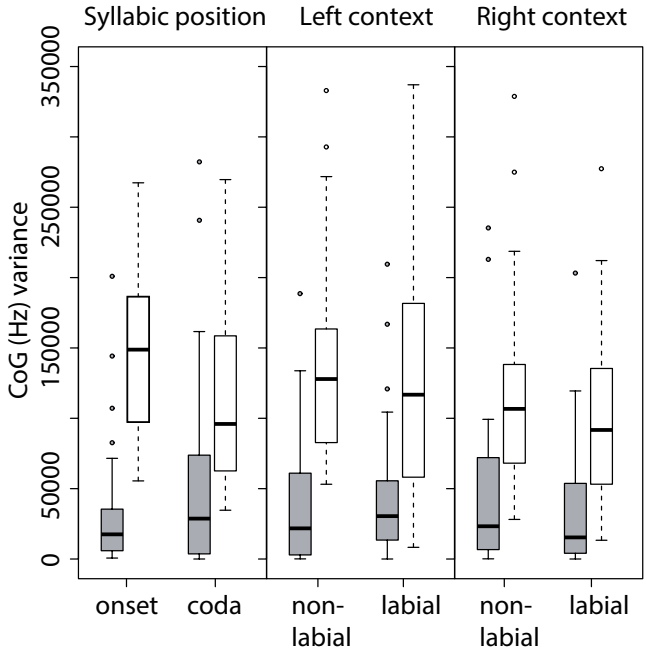

Figure. 2.3: *Boxplots of between- (grey bars) and within-speaker (white bars) variances per linguistic context factor level for /x/ CoG.*

2.4 Discussion

Previous work on read speech has shown that linguistic effects such as labial co-articulation and syllabic position have effects on fricative acoustics, and that some segments, such as /s/, are more speaker-specific than other segments. The present study wished to further investigate (1) whether linguistic effects on fricative spectra are present in speech materials that were not recorded in highly-controlled circumstances (in

this case, telephone dialogues), and (2) whether there is an interaction between segments' speaker-specificity and their linguistic context.

2.4.1 Linguistic effects

Regarding the first aim, linguistic effects were present in /x/, but were less prominent in /s/. The effect of syllabic position was present in both fricative sounds. Onsets showed higher intensity for both fricatives, which is consistent with results reported by Solé (2003) for American English fricatives. However, only for /s/ was there any indication for coda reduction in spectral measures, namely a higher center of gravity in /s/ onsets compared to codas.

As for labialization, the results confirmed the expected linguistic effects in /x/ acoustics; both left and right labial neighbors lower the resonance frequencies in /x/ by around 200 Hz and 300 Hz respectively. This is consistent with work on /s/ from read speech where anticipatory labialization lowered spectral energy by around 300 ~ 400 Hz (Koenig et al., 2013). Two significant interaction effects for center of gravity and spectral tilt furthermore indicated that spectral lowering is attenuated when both left and right context are labial and that the effect of anticipatory labialization is slightly larger in onsets. Regarding the first interaction, spectral lowering in these cases might be attenuated to not undershoot the articulatory target for /x/. The second interaction could be explained by more resistance to co-articulation across word boundaries; all onsets in this dataset were word-initial and all codas were word-final. This means that right context for onsets was part of the same syllable, whereas left context for onsets was part of the previous word. Previous work, however, found only minor effects of prosodic boundaries on co-articulation of consonant cluster [kl], and then predominantly when articulation rate was slow (Hardcastle, 1985). Regarding fricatives, work on s#CV versus #sCV clusters has shown no effects of word boundary on /s/ duration (Cho et al., 2014; Dumay et al., 1999). These findings suggest that word boundary effects may not explain why anticipatory labialization is larger for onsets than codas. Alternatively, this

interaction may reflect a qualitative difference in the type of lip-rounding; whereas right labial context for onsets consisted of rounded vowels, right labial context for codas consisted exclusively of bilabial consonants /b, p, m/ (because codas followed by vowels were labelled as ambisyllabic). Given that Munson (2004) has shown that the labialization effect in /s/ before /p/ was smaller than before /u/, the present result that anticipatory labialization lowers /x/ spectra more in onsets is therefore likely to stem from the specific labial segments that followed /x/ in onset versus coda position.

Contrary to /x/, the /s/ acoustics did not show the expected spectral lowering in labial contexts; in fact, when right context was labial, center of gravity showed a small but significant increase. The lack of spectral lowering in /s/ acoustics is likely a result of the speech channel used here, as much of the spectral energy for /s/ falls above the upper limit of the telephone bandwidth. In other words, given that the effect of labial co-articulation is well-attested for /s/, it is likely that labial co-articulation effects are not captured in these data. From the literature as well as the current results on /x/, the lowering due to labialization would be on the order of 300 Hz, which – relative to 4.8 kHz for a Dutch /s/ center of gravity – falls outside of the telephone band. This is supported by the mean CoG values; the mixed model's CoG intercept of 1.6 kHz for /x/ (CoG mean from the data was 1,586 Hz, *SD* = 421 Hz) was very similar to previously reported resonance frequencies for Dutch /x/ in broadband speech (Van der Harst et al., 2007). However, for /s/, the mixed model's CoG intercept of 2.5 kHz (*M* = 2,548 Hz, SD = 387 Hz) was around 2 kHz lower than what previous broadband studies have reported (Ditewig et al., 2019). In other words, we assume that the actual spectral peaks for /s/ were far over the upper limit of the landline telephone bandwidth used here, resulting in much lower CoG values in the present analysis with a lack of linguistic effects as a result.

2.4.2 Speaker classification

Regarding the dependence of speaker information on linguistic context in spontaneous telephone speech, the speaker-dependency of fricatives /s/ and /x/ seems to be distributed across linguistic contexts in a systematic way, but differences in speaker-classification accuracies were very small. In the current results, articulatory weak locations, i.e., codas and fricatives with labial neighbors, had slightly better speaker-classification scores than articulatory strong locations, i.e., onsets and fricatives with non-labial neighbors, for both /s/ and /x/. It seems that our data provides further evidence for the hypothesis proposed by He et al. (2017; 2019) that speech locations that may be less constrained by articulatory targets have more between-speaker variation. Moreover, the present study showed that these locations are more speaker-specific. Further examination of the between- and within-speaker variances showed that, for /x/ center of gravity, both between-speaker variance was increased and within-speaker variation was decreased in articulatory weak locations relative to articulatory strong locations.

Interestingly, speech features sampled from articulatory weak locations seemed to have more between-speaker variation even in the absence of clear acoustic differences. Fricative /x/ acoustics were altered by linguistic context within the telephone band and simultaneously showed differences in speaker-classification per linguistic context. However, /s/ also showed higher speaker-classification accuracies in articulatory weak locations, even though the expected acoustic effects for /s/ were minimal. The relative differences in speaker classification per linguistic context were very similar, but small, for both /s/ and /x/. Therefore, there is a possibility that these results are dependent on the specific sampling of the current dataset, which we assume to reflect distributional patterns of conversational Dutch; there are many more /s/ and /x/ tokens with non-labial context than with labial context (see Table 2.2). We cannot exclude that the lower number of labial contexts may have resulted in an under-estimation of speaker variance in that particular context. Given the minor differences between linguistic contexts, however, the results are expected to have no major implications for either listeners' perception of speaker information or for forensic speaker comparisons.

Comparing the contribution of the different acoustic measures to the speaker-classification accuracy of the multinomial logistic regression model, our results are similar to those reported by Kavanagh (2012) for English /s/ from read speech. Namely, spectral center of gravity and standard deviation are speaker-specific acoustic measures compared to temporal and amplitudinal measures. This might be because, whereas spectral measures reflect the size and shape of resonance cavities in the production of fricatives, this is not the case for temporal and amplitudinal measures. The same can be said for the lack of contribution of dynamic spectral measures; whereas static spectral measures reflect the shape and size of the resonance cavity, the dynamic measures reflect temporal patterns of articulation. Given the relatively static nature of fricatives, the lack of contribution of dynamic measures is not surprising. In addition, the short duration of fricatives in spontaneous telephone speech in combination with the large variation in phonetic context might also contribute to the lack of contribution for dynamic measures. Notably, the relative contributions of acoustic measures to speaker-specificity were very similar for the two fricative sounds examined here.

Interestingly, when using the same set of measures, fricative /s/ seems to be slightly more speaker-specific than /x/ even though the spectral peak of /s/ is not captured by the telephone bandwidth. In other words, /s/ retains some speaker-specificity even in limited bandwidths. Moreover, another highly frequent fricative in Dutch, /x/, contains comparable amounts of speaker-specificity in telephone speech. The correlation coefficient between the mean CoG values per speaker for /s/ and /x/ ($r = .46, N = 43, p < .01$) furthermore shows that the two fricative sounds carry partly complementary speaker information.

2.4.3 Limitations

It has to be noted that the current results only apply to male speakers and that it is possible that female speakers would display different behavior. Moreover, although studies have shown that sexual orientation and gender identity affect spectral measures such as CoG for /s/, the Spoken

Dutch Corpus only reports the speakers' sex (Oostdijk, 2000). Furthermore, the telephone dialogues from the Spoken Dutch Corpus were recorded almost two decades ago, which means that these results may not fully generalize to contemporary populations. With regards to Dutch fricatives, it has been shown that there is a general trend of devoicing, whereby /s/-/z/ and /f/-/v/ are merging (Gussenhoven, 1999; Pinget, Van de Velde, & Kager, 2014). In fricative realizations, this progressing merger may result in more variation. This means that it is possible that a contemporary population of speakers of Standard Dutch might show more between-speaker variation for /s/ than the set of male speakers in this study.

The use of the rather simple measures spectral CoG and SD might also be a possible limitation. These measures have been used often in previous work on fricatives, mostly with the goal of distinguishing the different fricative phonemes (e.g., Jongman et al., 2000). Much of this work focused on /s/ especially, which seems to be captured quite well by these measures. However, dorsal fricative /x/ seems to display a formant-like structure for most realizations, i.e., containing multiple spectral peaks. Although CoG seems to capture linguistic effects in /x/, such as contextual labialization, in the expected way, it is possible that some between-speaker variation is captured better by more complex measures such as discrete cosine transforms (DCT: Jannedy & Weirich, 2017). The spectral moments used in this study might thus underestimate the speaker-specificity for fricative /x/.

2.5 Conclusion

The present study investigated the distribution of speaker information in fricatives /s/ and /x/ as a function of syllabic position and labial co-articulation. Results have firstly shown that linguistic contexts affect fricative acoustics; whereas the linguistic-context effects reported in previous studies working with studio-recorded read speech can be replicated for dorsal fricative /x/ in spontaneous telephone speech, this is

less so the case for alveolar fricative /s/. We argue that the lack of effects for labial co-articulation for /s/ is a result of the telephone bandwidth used here. Secondly, for both /s/ and /x/, results showed somewhat more speaker-specificity for codas and for tokens with labial context. However, differences in speaker-specificity per linguistic context were small. These results support the hypothesis that the role of the speaker in speech is more explicit in parts of the speech signal where speakers may have more articulatory freedom, in this case, fricatives occurring in labial context and in coda positions.