The Emerging of Adaptive Contrast: Evidence and Lack thereof from Dutch Voiceless Sibilants

Xinyu Zhang

Under the Supervision of

Paul Boersma

A Thesis submitted for the degree of Master of Arts
in
General Linguistics

University of Amsterdam

Date

Contents

1 Introduction

1.1 Topic and Goals

Since the 1970s, many scholars (e.g. Liljencrants & Lindblom, 1972; Lindblom et al., 1983; Disner, 1983; Maddieson & Disner, 1984; Flemming, 2013 et seq.; Schwartz et al., 1997 et seq.; Boersma, 1998 et seq.; Boersma & Hamann, 2008; Hauser, 2017; etc.) have looked at the universals trends of phoneme inventories more or less within the framework of dispersion theory. Most of the previous work have been in the realm of vowel inventories with the exception of e.g. Schwartz et al. (2012) and Hauser (2017) on stop consonants, and Boersma and Hamann (2008) 's computational simulations on sibilants. Among them, most are concerned with synchronic distributions except for Boersma and Hamann (2008), although their model did not include language-specific articulatory learning, and the model has not yet been tested on real world data in any specific languages, to my best knowledge. The current study makes an attempt at investigating diachronic changes in the acoustic and auditory dispersion of voiceless sibilants in Dutch. Dutch sibilants were chosen as the subject of investigation because unlike in other languages such as German and English that also have two voiceless sibilants, the two voiceless sibilants in Dutch seem to be very much articulatorily and perceptually similar to the extent that arguments about their phonemic status have repeated been raised. Hence, there is a possibility that the two sibilants are either becoming more dispersed diachronically. It is also worth looking into whether the generalizations and predictions made by previous work apply to the Dutch voiceless sibilants.

1.2 Outline

The outline of the current paper is as follows: the first section describes the topic and goals of the current study and lays out a map of the paper. The second section provides some relevant background information about the acoustics and articulation as well as the inventories of voiceless sibilants. The third section serves as a general overview of sound change and auditory dispersion. Section Four tries to define a phonetic space for the Dutch voiceless sibilants. Section Five describes the experiment. The sixth and final section discusses the implications and limitations of the results, and speculates on possible future research.

2 Some Phonetics and Phonology of Voiceless Sibilants

In this section I briefly sketch out some relevant background information in some aspects of the phonetics and phonology of sibilants, especially voiceless sibilants, that are relevant to the current study.

2.1 The Articulation and Acoustics of Voiceless Sibilants

Sibilants are a subset of fricatives. Articulatorily, fricatives are produced by close approximation of two articulators so that the airstream is partially obstructed and turbulent airflow is produced (Ladefoged & Johnson, 2011, p.14). According to Ladefoged and Johnson (2011), there are two ways to produce said "turbulent airflow": it may be the result of the air passing through a narrow gap, as in the formation of [f], or it may be because the airstream is first speeded up by being forced through a narrow gap and then is directed over a sharp edge, such as the teeth, as in the production of [s]. Conventionally, the latter kind is categorized as sibilants, described by Ladefoged and Maddieson (1996) as "produced by the high-velocity jet of air formed at a narrow constriction going on to strike the edge of some obstruction such as the teeth".

Acoustically, fricatives have random energy distributed over a wide range of frequencies, and sibilants have more acoustic energy at a higher pitch than the other fricatives (Ladefoged & Johnson, 2011). In general, [ʃ] will have a lower pitch than [s] due to both the lower velocity of the airstream and the lengthening of the vocal tract by added lip-rounding in [ʃ]. Fricatives and sibilants could of course also be sub-divided by voicing but since the current paper only studies the voiceless fricatives, the focus will not be put on voicing or voiced fricatives. The same goes for fricatives whose place of articulation is not alveolar or palatal-alveolar or alveolo-palatal. According to Hughes and Halle (1956, p.308-309), the most useful measurements found to distinguish [s] and [ʃ] in English was "the energy in dB in the band from 4200 cps to 10 kc subtracted from the energy in dB in the band from 720 to 10kc". In other words, whether a fricative has more energy in the range of above 4.2kHz. Similarly, Strevens (1960) investigated isolated and lengthened voiceless fricatives that would usually only occur in paralinguistic communication of English speakers, and described [s] as having its lowest frequency "almost always above 3500 cps" whereas for [ʃ] the lowest frequency "varies between 1600 and 2500 cps".

Olive, Greenwood, and Coleman (1993) observed that in American English, /s/ shows the greatest concentration of energy above 3700Hz and /ʃ/ has its highest energy concentration between 1700Hz and 4500Hz, and that since the palatal-alveolar /ʃ/ is articulated close to the velum, a velar pinch may be expected for some vowels. In addition, they also noted that the vowel that follows the sibilant has some effect on the acoustics of the fricative. From their descriptions of the spectrograms, in both /s/ and /ʃ/ the lower edge of the frication frequency is dependent on the F2 of the following vowel, and since palato-alveolars are the most constrained in their distribution of formant values (indicating that the tongue has less freedom to prepare for the following sound), the fricative region of the palato-alveolars does not extend as far into the lower frequency as it did for the alveolars. But Olive et al. (1993) did not provide specific values. F2 transitions are included as one of the factors in Flemming (2018)'s prediction of markedness in sibilant inventories.

In a less language-specific study, Boersma and Hamann (2008) stated that sibilants in a language can often be ordered along a continuum of the spectral center of gravity or the spectral mean which, articulatorily, correlates to frontness of the tongue and with frontness of the place of articulation. But they did also mention that auditory dispersion by means other than Center of Gravity is possible for sibilants, although without exploring said possibilities further. This is indeed confirmed in e.g. Kochetov (2017) where he found that the anterior [s] can be palatalized to [sʲ] with only minimal reduction in Center of Gravity especially at the midpoint and offset of the frication and in female speakers.

2.2 Voiceless Sibilants in Dutch

The literature on Dutch[1] phonology is not in agreement on the phonemic status of the palatal sibilant /ɕ/, which is sometimes also transcribed as /ʃ/[2].

Mees and Collins (1982) described that the sequence <sj> is realized as an alveolar-palatal [ɕ] in Standard Dutch (*Algemeen Beschaafd Nederlands*, or ABN for short), and that it differs from the /ʃ/ in English, French, or Germa in that there is no labialization in the Dutch [ɕ]. According to Collins and Mees, the occurrence of the <sj> sequence is restricted only to loanwords and forms resulting in assimilation, hence did not merit a phonemic status in their analysis. They did acknowledge that there are arguments for regarding /sj/ as

[1]The Dutch language discussed here is limited to the Dutch spoken in the Netherlands.

[2]This non-/s/ voiceless Dutch sibilant will be transcribed as /ɕ/ instead of /ʃ/ throughout this text due to its palatalized nature and the frontness in its place of articulation.

Feature	/s/	/ʃ/
[high]	-	+
[mid]	+	-
[ant]	+	-
[dist]	-	+

Table 1: Feature matrix for Dutch consonants /s/ and /ʃ/, adapted from Schatz (1986)

an additional phoneme /ɕ/, however, the example they gave was an English pronunciation guide for Dutch speakers.

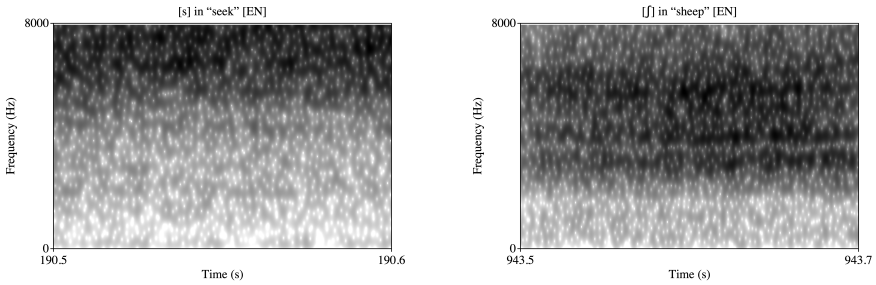
In a description of Dutch phonology, Booij (1999) listed /s/ as the sole voiceless sibilant in the Dutch consonant inventory and analyzes [ʃ, ʒ, c, ɲ] as /s,z,t,n/ palatalized before /j/, and the postalveolar fricatives that occur in loan-words such as *chique* [ʃik] and *jury* [ʒyːri] as "phonologically, combinations of /s, z/ and /j/" due to the reason that their realization is predictable.

Nooteboom and Cohen (1984, p.22) listed /ʃ/ as a separate phoneme in Dutch consonants on the basis that there exist minimal pairs distinguishing /s/ from /ʃ/. Similarly, Schatz (1986) also treated both /s/ and /ʃ/ as sibilants in Standard Dutch, and in a feature matrix distinguished the two by various features (see Table 1). However, she did point out that the SPE feature *distributed* [dist] might be redundant for Dutch consonants because laminals and apicals in Dutch have different places of articulation. According to Schatz (1986), in "plat Amsterdam", or Broad Amsterdam Speech, before a word boundary or morpheme boundary, [s] is often palatalized when preceded by the short vowels /ɑ/, /ɛ/, /ʉ/, or /ɪ/, and also when it is at an initial position in a word or a morpheme. However, participant 13 in the current study was born and raised in Amsterdam and lived in Amsterdam all his life, does have a distance of 1358.247 Hz between the CoG of his two voiceless sibilants, which is even slightly higher than the mean CoG distance of 1347.200 Hz between /s/-/ɕ/ among all young participants. This is possibly also affected by sociolinguistic reasons (see e.g. Faddegon (1951) and Schatz (1986) for more details).

Evers, Reetz, and Lahiri (1998) compared acoustic characteristics of sibilants between languages where /s/ and /ʃ/ are separate phonemes and languages in which the [s] and [ʃ] are allophonic. Dutch was included in the languages they examined and the sibilants [s] and [ʃ] were treated as allophones with [s] as the "default consonant" (p.351). The results show that the same predictor is equally

efficient at distinguishing the two phones regardless of their phonemic status. Hence whether [s] and [ʃ] are two separate phonemes or allophones of the same phoneme should not be a major concern of this paper.

In terms of comparing the Dutch sibilants to sibilants in other languages, apart from the phonemic status of /ɕ/ mentioned in the beginning of this section, the /s/, as well as /z/, in Dutch is also "far less articulatorily tense comparing to their counterparts in German, French and English" and produced with more lip protrusion, while the Dutch /ɕ/ is generally produced with no lip-protrusion (Mees & Collins, 1982). The lip-protrusion and the lack of tenseness in [s] lower its CoG while the palatalization and lack of lip-protrusion raise the CoG in [ɕ], making the two sibilants acoustically closer. Figures 1 and 2 show the spectrograms of the [s] and [ʃ] produced by a native speaker of British English (the Received Pronunciation)[3] and the [s] and [ɕ] produced by one of the participants in this study.



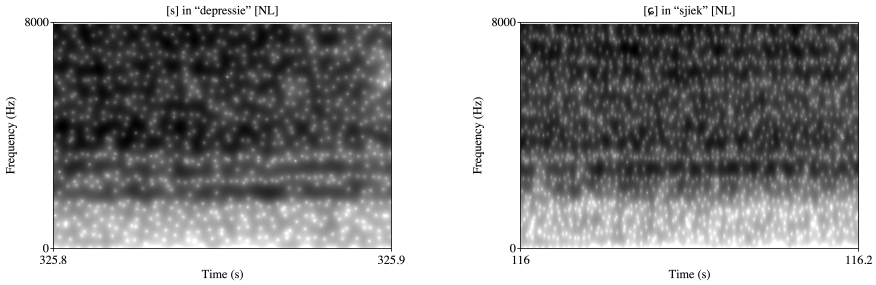
(a) /s/ in English *seek* (b) /ʃ/ in English *sheep*

Figure 1: /s/ and /ʃ/ in English

2.3 Sibilant Inventories

The sibilants [s] and [ʃ] are rather common in consonant inventories. Of the 317 languages that Maddieson and Disner (1984) investigated, about 83% of them have at least one anterior (dental or alveolar) /s/ (Maddieson & Disner, 1984, p.44). They concluded that /*s/ (referring to all types of s-sounds with unspecified dental or alveolar place) is the most common fricative, appearing in 88.5% of the languages that have fricatives, and that /s/ is the most common member of the group /*s/. The next most frequent fricative after /*s/, according

[3] Extracted from *BBC Learning English* (https://www.youtube.com/watch?v=htmkbIboG9Q)



(a) /s/ in Dutch *depressie* (b) /s/ in Dutch *sjiek*

Figure 2: /s/ and /ɕ/ in Dutch

to Maddieson and Disner (1984), is the voiceless palato-alveolar sibilant /ʃ/. Schatz (1986, p.77) also mentioned that [s] occurs frequently in the Dutch speech she collected, at the frequency of 35 times in a five-minute stretch of speech. In a study on markedness of sibilant inventories, Flemming (2018) indicated that /s/ is the least marked sibilant, and that in two-sibilant inventories, [s, s̻] is maximally distinct, [s, ʃ] minimizes effort, and that [s, ɕ] only occurs when the weighted F2 transition distance is high (>0.87). The markedness of [s, s̻], [s, ʃ] and [s, ɕ] pairings were not compared in (Flemming, 2018) since the former two are considered CoG-only contrasts (with zero weighted F2 distance difference) hence [s, ɕ] is already harmonically bounded with "wF2=0" (i.e. when no weight was assigned to F2 transitions). Maddieson and Disner (1984) also proposed the possibility that languages prefer saliency and that sounds that are more frequent in the inventories across languages are the ones with more acoustic energy, entailing good transmission properties. In this regard, Maddieson and Disner (1984, p. 50) compared the intensity rankings of fricatives as measured in Strevens (1960) to the frequency (i.e. rate of occurrence, not frequency in the sense of vocal fold vibration per second) rankings of the fricatives (see table 2). The results did not seem to indicate much correlation between intensity and the frequency of occurrence.

Earlier theories of inventory typology include Quantal Theory, markedness theory, and dispersion theory. Quantal theory has been criticized to have made incorrect predictions such as [ɛ] being unstable, and the existence of universally preferred hot spots (see e.g., Carré, 1996; Disner, 1983; Livijn, 2000). Traditional markedness theory has faced the objection that it merely formalizes the attested facts, rather than explaining them in terms of constraints on human

ranking	Intensity	Frequency of Occurance
1	ɕ	"s"
2	ʃ	ʃ
3	x	f
4	s	x
5	χ	χ
6	f	ɸ
7	θ	θ
8	ɸ	ɕ

Table 2: Ranking of fricatives by intensity and frequency of occurance, adapted from Maddieson (1984)

articulation, perception, and processing (Vaux & Samuels, 2015). Dispersion theory takes a more functional approach and incorporates factors like articulatory effort and perceptual contrast, and proposes principles of e.g. minimizing articulatory effort, maximizing perceptual contrasts, etc..

Building on Liljencrants and Lindblom (1972) on vowel inventories which focused more on perception rather than production, Lindblom and Maddieson (1988) mentioned that different from vowel inventories, more articulatory factors need to be considered in addition to perceptual distinction in consonant inventories, and stated that "Consonant inventories tend to evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory cost". Lindblom et al. (1983) proposed that languages are self-organizing systems and that phoneme inventories are emergent from the interaction of subsystems such as certain phonetic tendencies over time. The constraints used for speakers in their simulations were "sensory discriminability" and "preference for 'less extreme' articulation"; the listener based constraints used were "perceptual distance" and "perceptual salience" (ibid, p.191).

[Boersma (1998) considered the interaction between production and perception, According to Boersma's (1998) disentangling of Passy's *Principal of Economy*, the sibilant inventory should at the same time minimize both articulatory effort and perceptual confusion. Minimizing articulatory effort usually conflicts with minimizing perceptual confusion, since, per the assumption made in Boersma and Hamann (2008), among others (e.g., Lindblom & Maddieson, 1988), less articulatory effort is associated with more "central" auditory values e.g. in the case of Boersma and Hamann (2008), center of gravity, while minimizing perceptual confusion requires more auditory distance, consequently

resulting in more sounds with peripheral auditory values. Under this conflict of the two constraints, the final result of the inventory is an optimally dispersed one.]

Similarly, Flemming (2017) described a three-way conflict of constraints, namely among "MAXIMIZE CONTRAST", "maximizing distinctiveness", and "effort-minimization". "MAXIMIZE CONTRAST" refers to the preference of a higher number of contrasting sounds in the inventory. "Maximizing distinctiveness", as a distinctive constraint, favors more distinct contrasts, and "effort-minimization" penalizes articulatory effort. Among the three (types of) constraints, MAXIMIZE CONTRAST conflicts with maximizing distinctiveness since the space (hence the possible places and manners of articulation) in the oral cavity is limited, and fitting more contrastive sounds into the same limited space would result in less sharp distinctions between sounds. Additionally, effort-minimization conflicts with both MAXIMIZE CONTRAST and "maximizing distinctiveness" in such a way that the latter two constraints necessitate auditorily and articulatorily peripheral sounds which are difficult to realize without violating some effort-minimization constraints.

However, articulatory effort has been criticized as being difficult to measure (e.g. Stevens, 1980; Ohala, 1993, p.260), and dispersion theory has been undercut by the existence of vowel inventories such as that of Wari' (MacEachern, Kern, & Ladefoged, 1997, p.4-8). Additionally, more recent work (e.g., Schwartz et al., 2012; Hauser, 2017) tend to show that dispersion theory cannot fully explain stop consonant inventories in terms of place of articulation.

Though, arguably, Hauser (2017)'s metrics were only based on acoustic measurements and did not explicitly address the role of auditory perception, e.g. by taking into account e.g. that F1 is perceptually more salient than F2.

Nonetheless, it is true in vowel systems that inventories involving acoustically well-dispersed vowels are easier to both acquire and process because they are easier to discriminate, creating a tendency for languages to recruit such inventories (Joanisse & Seidenberg, 1998, p.335). Additionally, evidence such as the Hyperspace Effect (Johnson et al., 1993; Johnson, 2000), and that infant directed speech tend to have more extreme vowel qualities (Kuhl et al., 1997) provide some tentative support for the notion that a more dispersed system reduces perceptual confusion and thus is more learnable and more likely to remain stable diachronically. Vaux and Samuels (2015)'s model also supports the hypothesis that more dispersed inventories are more easily learnable.

3 Diachronic Change in Inventories

3.1 Sound Change

Ohala (1993, 243-247) described the process of sound change as when (synchronic) variation in production is hypo-corrected (where new categories are created) or hyper-corrected (where one phone is perceived as another existing phone) by the perceiver. In other words, there exist a fair amount of acceptable variations in production between speakers, and such variations within what is considered by the listener as the same category are acceptable, hence utterances within the acceptable variation range are perceived as the same sound. Hypo-correction is when enough number of individuals start producing outliers and the outliers don't get corrected by e.g. puzzlement or amusement from an interlocutor, resulting in the phonetic perturbations getting 'phonologized'. Hyper-correction, on the contrary, is when the listener implements a correction when the phonetic/auditory input was actually what was intended by the speaker (i.e. when no correction was needed). It is worth noting that the mechanism of sound change, according to Ohala (1993), is not teleological. The change occurs not in the message source (the speaker's brain) nor the message destination (the listener's brain) but in the transmission channel between them. This includes the speech production system and the listener's decoding system(ibid., p.262). Besides categorizing it as non-teleological, this account of sound change also ascribes the "locus of control" primarily to the listener's side and locates the mechanism centrally in the phonetic domain.[4]

3.2 Adaptive Dispersion

Adaptive dispersion refers to the hypothesis that the distinctive sounds of a language tend to be positioned in phonetic space so as to maximize perceptual contrast (Johnson, 2000). From the previous sections, it can be predicted that the less optimally dispersed sound inventories (i.e. inventories where the perceptual distance between phonemes are not wide enough to maintain perceptual distinctiveness, or where the articulations of phonemes e.g. in terms of manner and/or place, are more extreme than necessary) are less likely to remain stable and more likely to become more optimally dispersed diachroni-

[4] See S. Hamann (2009) for an alternative account where the learner's phonological knowledge is also involved, and see Fruehwald (2017) for more details on the role phonology could play on phonetic change. <- maybe check again

cally and that languages may apply diverse phonological processes to avoid a perceptually weak contrast. This has been observed in several attested sound changes e.g., in Korak (Bright 1978), where the contrast between the sibilants [s̪] and [s̻] (the former described as "a very far-forward apico-dental sound" and the latter as an "apico-alveolar", and further identified as "a retracted ess") was enhanced in younger speakers by pronouncing the former as an interdental [θ]; the voicing-only contrast between /g/ and /k/ was enhanced in Arabic by fronting and affricating /g/, in Japanese by nasalizing /g/, in low German by spirantizing /g/, and in Czech, Slovak, and Ukrainian by both spirantizing and pharyngealizing /g/ (Li, 2017). However, to quantify the auditory dispersion in consonants, or sibilants to be more precise, in a less impressionistic manner, an auditory (or at least acoustic) space might be needed.

Instead of making only post hoc guesses of causes and mechanisms, the current study makes an attempt to investigate whether the diachronic change in the acoustic and auditory dispersion of the Dutch voiceless sibilants is in consistency with the predictions made in previous work. A small contrast such like that between the Dutch [s] and [ɕ] is likely to become more dispersed after even one generation, according to Boersma and Hamann (2008). If also taking into account the fact that infants are not provided with the number of categories of the input they receive during first language acquisition, in a case like the Dutch voiceless sibilants where the two categories are too close or even overlap to certain extent, the infant acquiring the phoneme inventory might establish only one category instead of two. In this regard, a merger could also occur. I hypothesize that the two sounds would become more dispersed rather than to merge. The reason being that mergers happen as a way to enhance contrast, namely, when a merger happens, which usually locates somewhere in the middle of the auditory range between the two categories (assuming it is a merger of two categories) that are merged, the auditory distance between the merged sound and the remaining categories become larger than the pre-merger state (Becker-Kristal, 2010). Thus, though non-teleological, a merger is more likely to happen if there are other neighboring categories. Given that there are no other voiceless sibilants than /s/ and /ɕ/ in the Dutch consonant inventory to increase contrast with if /s/ and /ɕ/ were to merge into one category, the condition does not fit with that which a merger is likely to happen.

4 A Phonetic Space for Voiceless Sibilants

In this section I describe the acoustic measurements adopted, and give some brief justifications of the choices made.

4.1 Possible Dimensions

The literature has different metrics for differentiating fricatives acoustically. Ladefoged and Johnson (2011) mentioned multiple possibilities to distinguish fricatives such as voicing, articulatory gestures, tongue shape (tongue grooved v.s. tongue flat), and concluded that a better way is to separate them into groups on a purely auditory basis, for instance, according to the loudness in high pitches, which distinguishes the sibilants from non-sibilant fricatives, but they did not go into detail about the acoustic or auditory measurements that would separate one sibilant from another. Hayward (2014) listed frequency of main spectral peak, diffused-compact (e.g. [f] and [θ] being more diffused and [ʃ] more compact), slope of the overall spectrum ([ʃ] rises steeply to its peak and [s] rises more gradually) to distinguish fricatives. She also mentioned that the spectra of English fricatives vary considerably from speaker to speaker, and that at least for English, it seemed appropriate to describe fricative spectra by category in terms of the above perspectives rather than in terms of specific formant frequencies as in vowels. Also for the fricatives in English, Jongman, Wayland, and Wong (2000) found that acoustic properties such as spectral peak location, spectral moments (mean, variance, skewness, kurtosis), normalized amplitude, normalized duration, F2 onset frequency, and relative amplitude, are all relevant and are all robust enough in distinguishing /s/ and /ʃ/.

In Bolla and Varga (1981)'s observation, fricatives with a higher place of articulation (or palatalized, in the case of Russian fricatives, which were the subject of their research) have higher intensity than fricatives with a lower place of articulation (or non-palatalized). But Bolla and Varga (1981)'s results were only based on one (male) speaker. In a similar study on Russian fricatives, Kochetov and Radišić (2009) did not find intensity useful in distinguishing palatalized fricatives from non-palatalized fricatives.

In a more recent Optimality Theoretic typological study, Kokkelmans (2019) showed that "distributedness" is one possible dimension to implement auditory dispersion in sibilant inventories.

Other factors such as lexical frequency can also influence dispersion (see e.g.,

Lindblom, 1996; Van Son, Beinum, & Pols, 1998; Bybee, 2003).

In sum, a phonetic space (as dispersion theory models usually adopt) for fricatives is far less as established as the vowel space. Since the present study focuses on voiceless sibilants, I will opt for measurements that are more relevant for sibilants or fricatives in general.

4.2 Two-Dimensional Mapping

As was mentioned in Section 4.1, previous studies categorized fricatives by different metrics such as spectral peak location, frequency of main spectral peak, spectral center of gravity, diffused-compact-ness, slope of the overall spectrum, spectral moments, F2 onset frequency, intensity, and duration, etc.

Among the above, spectral center of gravity contains information including spectral peak location and the frontness of the place of articulation, and auditorily correlates to the listener's averaging of frequency and intensity components of a speech-like signal (Fagelson & Thibodeau, 1994). Additionally, according to Gordon et al. (2002)'s cross-linguistic study of the acoustics of voiceless fricatives in seven languages, "gravity center frequencies robustly differentiated many of the fricatives in the examined languages" (pg. 29). Diffused-compact and distributedness could roughly translate to the width of the spectral peak acoustically, since more distributed sounds have more filtering in the vocal tract thus leading to engergy spreading out over a wider range of frequencies and hence a wider peak or even multiple "diffused" peaks (Johnson, 2011; Stevens, 2000).

Considering that factors such as the slope of the overall spectrum, spectral peak location, and frequency of the main spectral peak are partially represented by the spectral center of gravity and that spectral center of gravity alone is often robust enough in differentiating different fricatives (Gordon et al., 2002), the present study uses a two-dimensional space of spectral center of gravity and width of the spectral peak as an acoustic space for the sibilants in question.

5 The Experiment

5.1 Participants

Native speakers of Dutch of two age groups were recruited mainly from the University of Amsterdam and were paid for their participation. One group aged between 19 and 27 (5 female, 3 male, mean age = 23.38, standard deviation=

2.20). All participants in this age group were students at the University of Amsterdam and none of them studied linguistics. The other group aged between 61 to 75 (8 female, 2 male, mean age = 67.60, standard deviation = 4.32), mostly consisting of professors and staff members from the University of Amsterdam, none of whom specialized in phonetics or phonology. All participants were raised in monolingual households with native Dutch-speaking parents. All participants reported to have no abnormalities in their vision or speech.

5.2 Material and Design

A list of 113 Dutch sentences were constructed (see Appendix [reflabel] for the full list). The sentences contained 33 tokens of /s/, and 26 tokens of /ɕ/[5]. 54 sentences containing neither of the two target sibilants served as fillers. All the target sibilants were situated in intervocalic positions and in stressed syllables. Among the /s/ tokens, 10 were word-initial, 7 were word-medial, and 17 were word-final. Among the /ɕ/tokens, 13 were word-initial, 7 were word-medial, and 6 were word-final[6]. All but the word-final sibilants were in inter-vocalic environments.

The list of sentences were randomized for each participant. A trial sentence with multiple sibilants built in ("this is a sentence that I am reading aloud to help set up the recording devices") was shown as an example to familiarize the participant with the procedure, as well as to help the experimenter adjust the gain constant of the microphone while the participant was reading the trial sentence aloud. The participant is then prompted to press a key to start the experiment. The prompting speed was controlled by each participant by pressing a key after they finish reading each item out loud. The recording started each time the participant pressed a key to show a sentence, and stopped when the participant pressed the key to indicate the reading is finished and that the next sentence should be shown. 50ms of delay was added after the pressing of the key as a buffer. Each chunk of recording was labeled automatically with the index of the sentence from the stimuli list and then concatenated by the sequence of the list so that the sequence of the sentences was identical in each final product without being influenced by the randomization of the stimuli at the time of the recording.

[5]There were instances where the participant pronounced the words *jus* and *jam* as [ʒy] and [ʒɛm] respectively. Such tokens were not used in the analysis.

[6]Some stimuli comtained more than one sibilant in more than one positions, e.g. *cynisch* has both word-initial /s/ and word-final /s/.

The recordings were done in a soundproof studio in the Speech Lab at the University of Amsterdam, using a Senheiser MKH105T microphone and a pre-amplifier designed and built by the lab technicians at the Speech Lab. The subjects were recorded one by one in a seated position. They were each instructed to keep a constant distance of 20 cm from the microphone. To reduce the effect of read speech and to preserve naturalness to some extent, participants were instructed to first look at the sentence and then read it out loud as if the sentence was part of a conversation. Prior to the recording, participants were informed in the consent form that the purpose of the recording was to collect natural speech samples of Dutch phrases. Each participant filled out a post-test questionnaire after the recording (see Appendix [reflabel]) to gather information about the speaker's language background including their age, profession, birth place, cities and towns that they lived in the Netherlands and abroad for over 6 months, whether they were raised in a monolingual household and their second language(s). The post-test questionnaire also asked each participant for their speculations about the topic of the study in order to exclude the results of participants who might have guessed the targets and hyperarticulated during the recording. None of the participants guessed correctly or even close.

5.3 Acoustic Analysis

All the relevant sections (i.e. the target sibilants and their surrounding vowels) in the recordings were segmented by hand. Annotation was automated by a script that fills in the annotation of the underlyingly identical segments in different recordings (e.g. the /ɕ/'s in all the *koosjer* tokes). The script ignores the instances where the pronunciation does not match the intended sibilant (e.g. when *jus* was pronounced as [ʒy]) which were not segmented in the TextGrids to begin with hence not annotated or extracted for measurements (see Appendix [reflabel] for the script).

For more precise calculation of the spectral center of gravity and the spectral standard deviation of the sibilants, the sibilants were segmented in such a way that as little formant transitions as possible were included, as is shown in Figure 3[7], where part of the very beginning of /ɕ/ in *pistache* was intentionally left out to avoid the influence of voicing, and in Figure 4, where the transition into and out of /s/ in *tussen* was left out. For this reason, in addition to the fact

[7]I am not sure if I should use the format in Fig.3 or Fig. 4. The information in Fig. 4 seems to be enough, but without the drawing the selection lines from the TextGrid over the spectrogram again, the dashed lines are barely visible (as in Fig. 4).

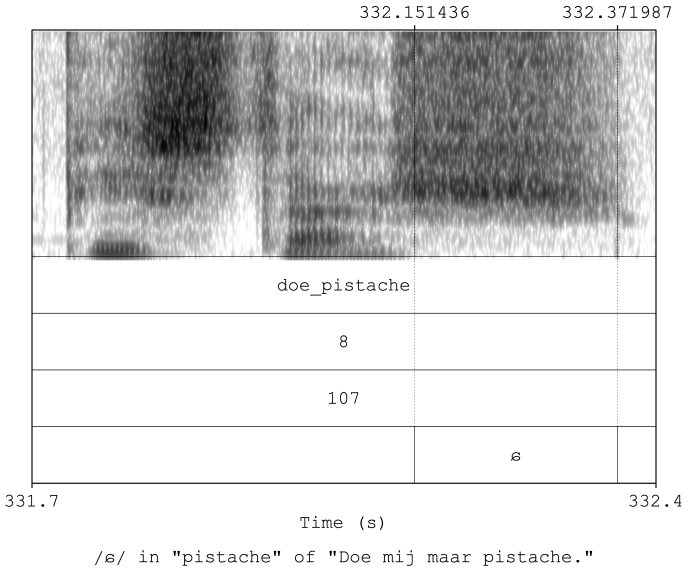


Figure 3: Spectrogram of "pistache" in the stimulus "Doe mij maar pistache" with dashed lines marking the duration segmented for and annotated as /ɕ/

that participants varied in speaking rate, durations of the sibilants were not measured[8]. Each relevant part (i.e. every annotated sibilant segment in every recording) was extracted with a rectangular window shape, relative width 1.0. Each of the 916 extracted tokens was subjected to a spectral analysis, and passed through a stop Hann Band filter from the frequency of 0 Hz to 550 Hz, with 50 Hz smoothing. The spectral center of gravity (CoG) was measured from each of the spectra. The width of the spectral peak was measured as the spectral standard deviation (power = 2). See Appendix [reflabel] for the full results of the acoustic measurements.

Figure 5 is the scatter plot of sibilants produced by the speakers aged between 19 and 27, with one sigma ellipses numbered by participant ID, /ɕ/ in red circles, and /s/ in blue plus signs. The scatter plot of the speakers aged between 61 and 75 is shown in Figure 6. Without much statistical analysis, one can already see that there is more overlapping between the two sibilants in the groups aged between 61 and 75. The only overlapping of /ɕ/ and /s/ in the younger group occurrs between different speakers (i.e. the /ɕ/ of Participant 11

[8]This might be improved by using a different segmenting scheme.

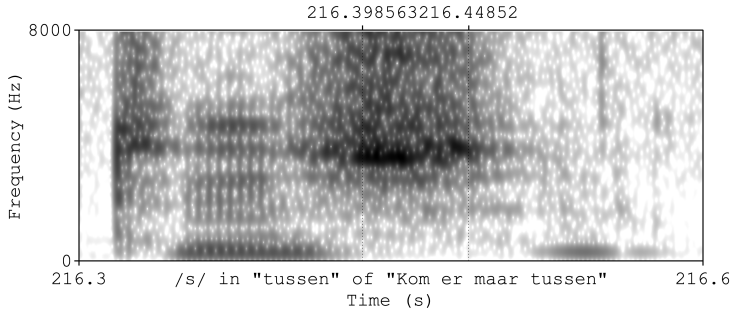


Figure 4: Spectrogram of "tussen" in the stimulus "Kom er maar tussen" with dashed lines marking the duration segmented for and annotated as /s/

overlaps with both the /s/ produced by Participant 10 and slightly with the /s/ produced by Participant 15), but never within speakers.

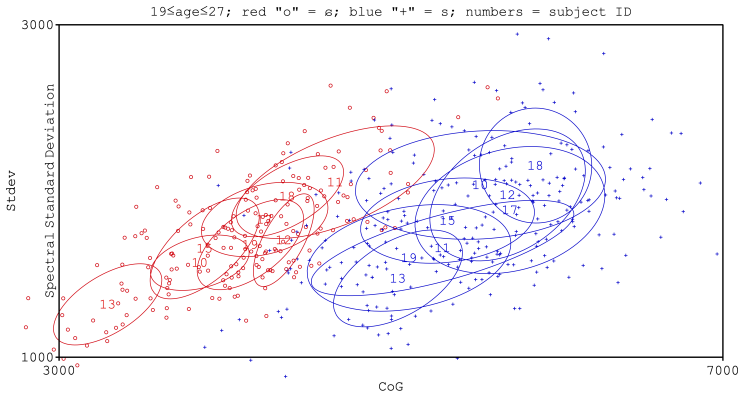


Figure 5: Scatter plot of the sibilants produced by younger speakers

In a slightly different scheme, Figures 7 and 8 show the ellipses of the sibilants produced by each participant in the two groups[9], assigning each participant in a different color of ellipses, with the sibilants marked in the center of each ellipsis. It can be seen from Figure 7 and Figure 8 that the ellipses of the same colors never merge in the younger group. However, the ellipses of the same colors are located much closer in the older group. One of colors have visible overlapping, and several others are very much closer together, making the two sibilants almost a continuum.

[9]maybe this is not necessary?

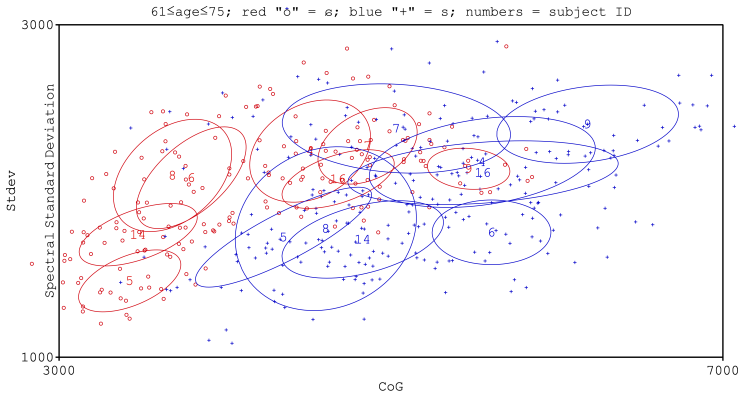


Figure 6: Scatter plot of the sibilants produced by older speakers

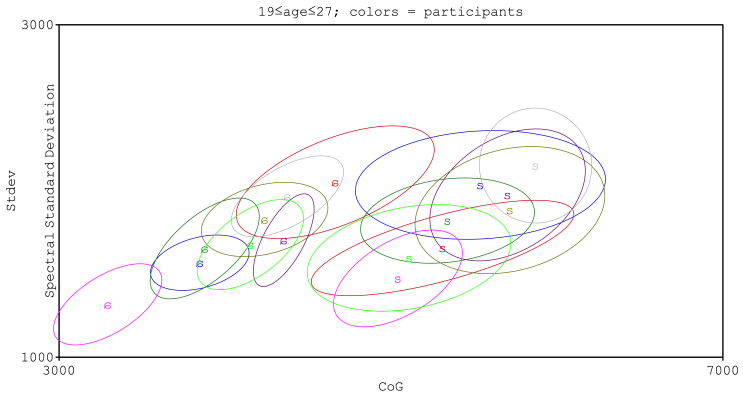


Figure 7: Ellipses of the sibilants produced by younger speakers, with each participant in one color

5.3.1 Linear Mixed-Effects Models

The data was analyzed in R (R Core Team, 2020) using the linear mixed-effects models. *Age Group* and *Sibilant*, as well as the *height*, *rounding*, and *frontness* of the succeeding vowel were the fixed effects that were modeled. The *Index* of the token and *Participant ID* were included as random effects. The height of the succeeding vowel was coded as a tertiary contrast, with /œ/, /au/, /a/, /ai/, /ã/ coded as "low", /ɔ/, /ɪ/, /o/, /ə/, /ɔ̃/,/au/ coded as "mid", and /i/, /u/, /y/ coded as "high". Rounding was coded as a binary contrast. Frontness was coded as a tertiary contrast, with /œ/, /au/, /a/, /ai/, /ɪ/, /i/, /ã/, /y/ as "front", /ə/ as mid, and /ɔ/, /o/, /u/, /ɔ̃/ as "back". All levels of all contrasts

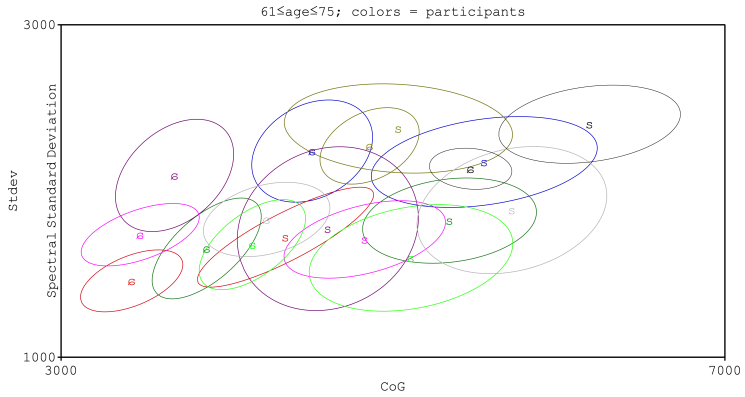


Figure 8: Ellipses of the sibilants produced by older speakers, with each participant in one color

are orthogonal both to each other and to the intercept. See appendix [ref] for the formula and coding of contrasts.

The same sets of fixed effects and random effects were used in the two models, one with CoG as the dependent variable, and one with the width of the spectral peak as the dependent variable.

CoG

Results show that without taking Age into consideration, the spectral center of gravity is 1546.18 Hz higher in /s/ than in /ɕ/ on average (95% confidence interval = 1272.588 Hz .. 1818.093 Hz; t = 9.981) among the Dutch speakers who participated, which fits the general expectation. The estimated mean for the interaction effect of AgeGroup and Sibilants is 391.03 Hz, which is not significant (95% confidence interval = -44.039 Hz .. 829.662 Hz; t= 1.723). In other words, from the data collected in this study alone, we cannot conclude that the CoG difference (on the Hertz scale) between the two sibilants /s/ and /ɕ/ is significantly different between the two age groups.

Width of Spectral Peak

The same fixed and random effects as in the CoG analysis were modeled as a function of the width of the spectral peak, with the same contrasts coding used in the linear regression model for CoG (see Appendix [ref] for the formula and the contrast coding scheme). Results show that the difference of spectral standard deviation between the sibilants /s/ and /ɕ/ of the participants in the younger

group is 389.80Hz wider than that of the older group, which is significant (95% confidence interval = 75.491 Hz .. 705.519 Hz; t = 2.331).

5.3.2 Spectral Principal Component Analysis

Although considered important by many (e.g. Flemming, 2018; Olive et al., 1993), formant transitions are not considered in the present study, due to the difficulty to control for different vowels that surround the sibilants in the stimuli. Additionally, even though formant transitions can be a prominent cue in perception, formant trajectories are not easily detectable in fricative signals, and therefore might not be as useful as spectral information for classifying fricatives, as is pointed out in (S. R. Hamann, 2003).

For a more in-detail comparison and description of the between-sibilants acoustic difference between the two age groups, a spectral principal component analysis was conducted. The reason for adopting spectral principal analysis is to take into account the spectral shape as a reflection of the characteristic energy difference between frequencies in the two sibilants, as was pointed out by (Evers et al., 1998) [10]. The pre-processing of the acoustic signal for the spectral component analysis is described as follows. A long-term average spectrum (LTAS) analysis was performed on each of the relevant segments. Each LTAS was computed with a bin width of 250 Hz and a frequency range of 550-10000 Hz. The energy in each of the 38 250-Hz-bins of each LTAS of each of the 916 relevant tokens was calculated.

Pooled Data

A principal component analysis was run on both sibilants produced by both age groups. Figures 9 to 12 show eigenvectors 1 to 4. As was explained above, the elements on the x-axis represent frequency bins with the width of 250Hz, and the y-axis indicates energy in the corresponding bins. The first eigenvector has no zero crossings, indicating that it differentiates the sounds by loudness only, which is irrelevant for the purpose of investigating spectral shape.

The second eigenvector has one zero crossing near bin14 which shows that eigenvector 2 is an indication of whether the energy level is on average higher in the frequency range below or above 550Hz + 14 ∗ 250Hz = 4050 Hz. This

[10] Evers et al. (1998) deemed it obvious for acoustic analyses of fricatives to consider spectral shape difference. They suggested the use of the spectral slopes besides absolute energy levels, but due to human errors in the current study, the option of spectral slopes is unfortunately not viable

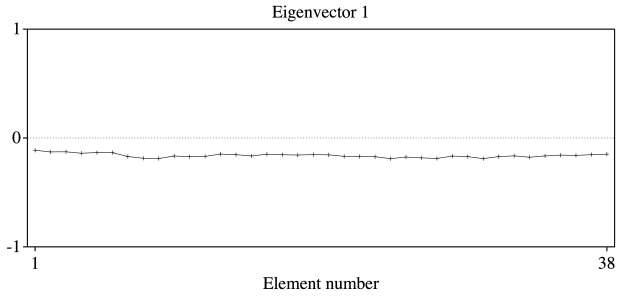


Figure 9: Eigenvector 1 of all speakers and both sibilants

is slightly lower than the threshold of 4.2k Hz in Hughes and Halle (1956), mentioned in §2.1.

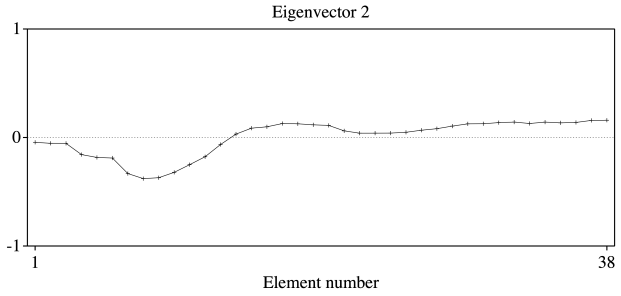


Figure 10: Eigenvector 2 of all speakers and both sibilants

Eigenvector 3 has three zero crossings, indicating that it differentiates the spectra between energy in three frequency ranges, namely bin8 to bin16 (i.e. $550\text{Hz} + 8 * 250\text{Hz} = 2550$ Hz to $550\text{Hz} + 16 * 250\text{Hz} = 4550$ Hz), bin 17- bin 29 (i.e. $550\text{Hz} + 17 * 250\text{Hz} = 4800$ Hz to $550\text{Hz} + 29 * 250\text{Hz} = 7800$ Hz), as well as above bin30 (i.e. $550\text{Hz} + 30 * 250\text{Hz} = 8050$ Hz).

The fourth eigenvector reflects the variation of energy in more specific parts of the spectra.

Thus, the second and third principal components together account for the main differences. Figure 13 is the sibilants plotted according to their eigenvalues in the second and third principal components. The marks "sy", "ɕy", "so" "ɕo" represents data points of [s]'s and [ɕ]'s produced by "y"ounger and "o"lder speakers among the participants, respectively. It can be seen from the scatter plot that there is some over lap between the 1SD ellipses of "ɕo" and "so",

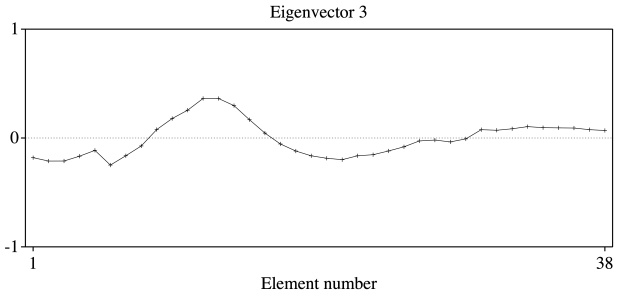


Figure 11: Eigenvector 3 of all speakers and both sibilants

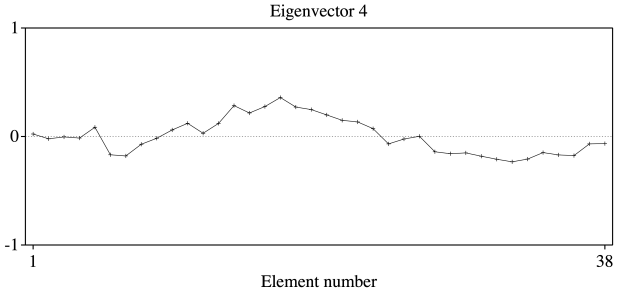


Figure 12: Eigenvector 4 of all speakers and both sibilants

but a wide gap between the edges of the "ɕy" and "sy" ellipses, denoting that the acoustic distance between the two sibilants is indeed wider in the younger generation.

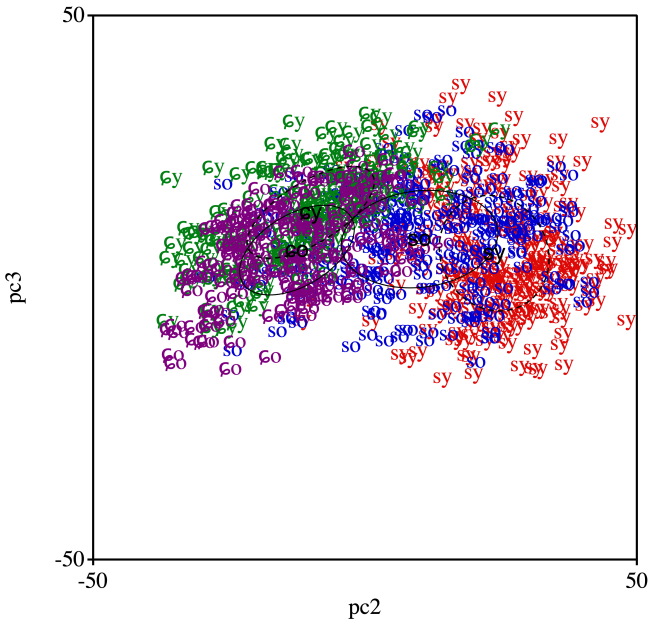


Figure 13: two sibilants produced by two age groups

Age Group Data

To better compare how the two age groups differ in the way they differ the two sibilants in production, one separate principal analysis was run for each age group. Figures 14 to 17 show the first four eigenvectors of the spectral principal component analysis, with green circles for the younger group and purple crosses for the older group.

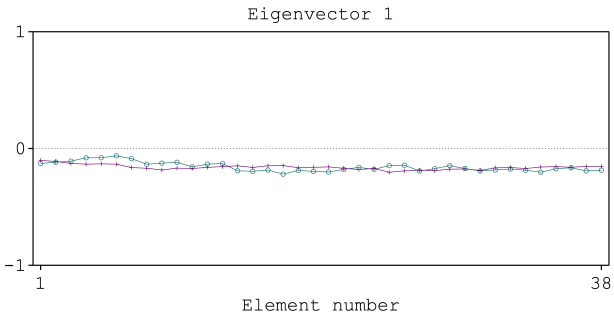


Figure 14: Eigenvector 1 of both sibilants produced by each age group

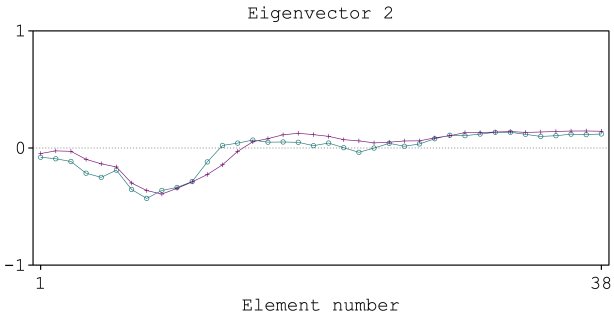


Figure 15: Eigenvector 2 of both sibilants produced by each age group

There is no fundamental difference in the first two Eigenvectors between age groups, except that there is some distance in bin 13 (550Hz + 13 $*$ 250Hz = 3800Hz). Looking at each of the two age groups separately, the younger group has a zero crossing at bin13, and the older group has a zero crossing at a frequency range slightly higher than bin14. In other words, the younger group distinguish the two sibilants by whether the energy level is on average higher in the frequency range below or above bin 13 (550Hz + 13*250Hz = 3800Hz), while for the older group the threshold is slightly above bin14 (550Hz + 14*250Hz = 4020Hz).

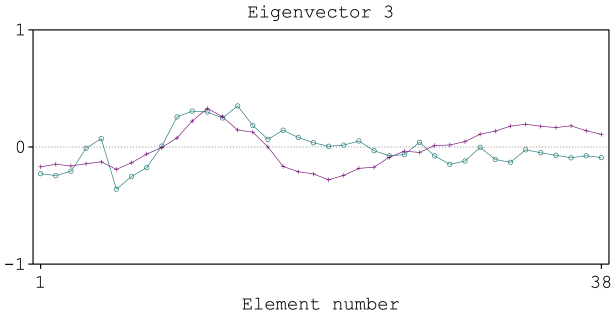


Figure 16: Eigenvector 3 of both sibilants produced by each age group

It can be seen in the third Eigenvector that the older group has a prominent peak in energy between bin 9 to bin 16 (i.e. 550Hz + 9*250Hz = 2800Hz to 550Hz + 16*250Hz = 4550Hz) marked by two zero crossings, while the peak of energy in the younger group is less sharp and more prolonged, between bin 9 to bin 23 (i.e. 550Hz + 9*250Hz = 2800Hz to 550Hz + 23*250Hz = 6300Hz).

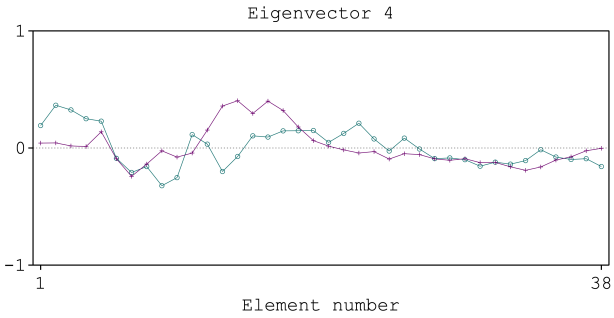


Figure 17: Eigenvector 4 of both sibilants produced by each age group

That is, the older group differentiates the two sibilants by the energy difference between 2800 Hz to 4550 Hz, and the younger group differentiates the two sibilants between 2800 Hz to 6300 Hz.

Additionally, the biggest contrast in energy between the two sibilants is between bin 6 and bin 14 (i.e. energy at 2050Hz and energy at 4050Hz), while for the older group the contrast is between energy in bin 12 and bin 20 (i.e. energy at 3550 Hz and energy at 5550 Hz) for the two sibilants.

From the principal component analyses above, it is clear that in production, the two age groups differentiate the two sibilants in different ways. Namely, the frequency range(s) where the main difference resides differ between two age groups.

5.4 Auditory Estimations

In consideration of the potential role that perception plays in the dispersion-theoretic non-teleological diachronic changes (e.g. Boersma (1998) explicitly points out that the dispersion is about *auditory* distance as opposed to *acoustic* distance), perception studies are also needed to fully understand a phonemic system.

Due to the limit of time and scope of the current project, I convert the acoustic measurements into a more psychoacoustically appropriate estimation in order to indirectly examine the dispersion auditorily. To do so, I convert the measurement of Center of Gravity from the Hertz scale to the ERB scale, since the ERB scale corresponds to a good agreement to the direct physical audio filter bandwidths defined in terms of *place* along the basilar membrane, in the frequency interval [400 Hz, 6.5 kHz] in humans (Greenwood, 1990, p.2601). The

spectral standard deviation is not directly convertible into the ERB scale due to the non-linear nature of the ERB scale and the linearity of the Hertz scale, as well as the fact that the spectral standard deviation is a distance measure rather than a point value. For an estimation, I take the center of gravity, which is the mean frequency weighted by spectral power, convert the value from Hz to ERB (CoG_Erb). I then convert the value of one standard deviation of the CoG from Hertz to ERB (SD_Erb). Next, I add one standard deviation in Erb to the mean in Erb to get the upper bound, and subtract one standard deviation in Erb to the mean in Erb to get the lower bound. Lastly, I divide the difference between the upper bound and the lower bound by 2. The formula[11] below illustrates the process:

$$StdevErb = 0.5x[hertzToErb(CoG+.CoG\_SD)-hertzToErb(CoG-CoG\_SD]$$

Linear Mixed-Effects Models

The same fixed and random effects as in the acoustic analyses were modeled as a function of CoG in ERB and the spectral standard deviation in ERB, respectively. Contrast coding also remained the same as the acoustic analyses (See Appendix [ref] for the R script and full results in detail).

Results show that the CoG difference (on the ERB scale) between the two sibilants is 0.69Erb larger in the older group than in the younger group of speakers who participated in the recording, which is not significant (95% confidence interval = -0.246Erb .. 1.640Erb; t = 1.387). The difference of the width of spectral peak (on the ERB scale) between the two sibilants /s/ and /ɕ/ is 0.65 Erb wider in the older group than in the younger group, which is not significant (95% confidence interval = 0.653 .. 0.444; t = 1.470).

6 Discussion

(factors?) cannot be isolated

perception affected by other factors such as neighborhood density, word co-occurence statistics (McDonal & Shillcock 2003 in Heutting & Janse paper on working memory and visual world)

Hearing loss cause feedback (in older group, because they cannot hear themselves producing anything over 8000Hz) to differ in age groups, hence older group

[11] In the format of the Praat scripting language

modify/adapt/adjust their fricative production according to the frequency range in the feedback from their own perception, by lowering the frequencies. And fricatives, esp. sibilants are mainly consisted of high frequencies.

Could incorporate e.g. the Jefferies-Matusita distance (check spelling) to quantify within-category variation

Possible counter argument: influence from English, but whether L2 input influence L1 phonetic realization is still an open question to some extent. Chang dissertation: vowel & F0 shift, Chang 2010b(=dissertation) in chapter 36, VOT change

Add to "motivation of research": in B&H 2008 because the learning algorithm in the simulation is to some extent supervised (e.g., that it was provided to the learning algorithm that there are two categories) while in reality newborns are not supervised when they acquire phonemic categories.

Might merge as well but need further xxx on what merits as a phenemic category and when are small contrasts considered within category variation.

to varify phonemic status, maybe look into how speakers deal with loanwords containing these two sibilants.

within category variation (to maintain distance, require precision?)

may disperse on other dimensions such as retroflexion and palatalization (check the Phon and Phon of Retroflexes for metrics used to measure retroflexion)

Revamped the scheme for the PCA: separated two age groups, compared sibilants within age group, and then compared the difference (between sibilants within group) of difference (between groups).

dispersion and lack thereof in consonant inventories (chrome page)

future: whether younger and older listener also use different auditory cues to distinguish the two sibilants, in the same way as they do articulatorily as was mentioned in §5.3.2.

References

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and dispersion theory: Insights from a large cross-linguistic corpus.* University of California, Los Angeles.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* Den HaagHolland Academic Graphics/IFOTT.

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(2), 217–270.

Bolla, K., & Varga, L. (1981). *A conspectus of russian speech sounds* (Vol. 32). Akadémiai Kiadó.

Booij, G. (1999). *The phonology of dutch.* Oxford University Press.

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge University Press.

Carré, R. (1996). Prediction of vowel systems using a deductive approach. In *Proceeding of fourth international conference on spoken language processing. icslp'96* (Vol. 3, pp. 1593–1596).

Disner, S. F. (1983). *Vowel quality: The relation between universal and language specific factors* (Vol. 58). Phonetics Laboratory, Department of Linguistics, UCLA.

Evers, V., Reetz, H., & Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of phonetics*, *26*(4), 345–370.

Faddegon, B. (1951). Analyse van een amsterdamse klankwet. *Album Dr. Louise Kaiser*, 26–30.

Fagelson, M., & Thibodeau, L. M. (1994). The spectral center of gravity effect and auditory filter bandwidth. *The Journal of the Acoustical Society of America*, *96*(5), 3284–3284.

Flemming, E. (2013). *Auditory representations in phonology* (Unpublished doctoral dissertation). UCLA.

Flemming, E. (2017). Dispersion theory and phonology. In *Oxford research encyclopedia of linguistics.*

Flemming, E. (2018, Oct. 6). Sytematic markiedness in sibilant inventories. In *Annual meeting on phonology 2018.* San Diego, California. Retrieved from http://phonology.ucsd.edu/program/sunday/posters-3/

Fruehwald, J. (2017). The role of phonology in phonetic change. *Annual Review of Linguistics*, *3*, 25–42.

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Hamann, S. (2009). The learner of a perception grammar as a source of sound change. *Phonology in Perception. Berlin: Mouton de Gruyter*, 111–149.

Hamann, S. R. (2003). *The phonetics and phonology of retroflexes* (Unpublished doctoral dissertation).

Hauser, I. (2017). A revised metric for calculating acoustic dispersion applied to stop inventories. *The Journal of the Acoustical Society of America*, *142*(5), EL500–EL506.

Hayward, K. (2014). *Experimental phonetics: An introduction*. Routledge.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *The journal of the acoustical society of America*, *28*(2), 303–310.

Joanisse, M. F., & Seidenberg, M. S. (1998). Functional bases of phonological universals: A connectionist approach. In *Annual meeting of the berkeley linguistics society* (Vol. 24, pp. 335–345).

Johnson, K. (2000). Adaptive dispersion in vowel perception. *Phonetica*, *57*(2-4), 181–188.

Johnson, K. (2011). *Acoustic and auditory phonetics*. John Wiley & Sons.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 505–528.

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

Kochetov, A. (2017). Acoustics of russian voiceless sibilant fricatives. *Journal of the International Phonetic Association*, *47*(3), 321–348.

Kochetov, A., & Radišić, M. (2009). Latent consonant harmony in russian: Experimental evidence for agreement by correspondence. In *Proceedings of fasl* (Vol. 17, pp. 111–130).

Kokkelmans, J. (2019). A typological model of sibilant inventories and the principles which shape them. In *poster at the 27th manchester phonology meeting*. Retrieved from http://www.lel.ed.ac.uk/mfm/27mfm-prog.pdf

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Wadsworth, Cengage.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the worlds languages*. Blackwell.

Li, M. (2017). *Sibilant contrast: Perception, production, and sound change* (Unpublished doctoral dissertation). University of Kansas.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–862.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *The Journal of the acoustical society of America*, *99*(3), 1683–1692.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. *Linguistics*, *21*(1), 181–204.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. *Language, speech and mind*, 62–78.

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories–hot spots or adaptive dispersion. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *11*.

MacEachern, M., Kern, B., & Ladefoged, P. (1997). Wari' phonetic structures. *J. Amazonian Lang.*, *1*, 3–28. Retrieved from http://etnolinguistica.wdfiles.com/local--files/artigo%3Amaceachern-1997/maceachern_et_al_1997_wari.pdf

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge university press.

Mees, I., & Collins, B. (1982). A phonetic description of the consonant system of standard dutch (abn). *Journal of the International Phonetic Association*, *12*(1), 2–12.

Nooteboom, S. G., & Cohen, A. (1984). *Spreken en verstaan: een nieuwe inleiding tot de experimentele fonetiek*. Van Gorcum.

Ohala, J. J. (1993). The phonetics of sound change. *Historical linguistics: Problems and perspectives*, 237–278.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of american english speech: a dynamic approach*. Springer Science & Business Media.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Schatz, H. F. (1986). *Plat amsterdams in its social context: a sociolinguistic study of the dialect of amsterdam* (Vol. 6). PJ Meertens-Instituut voor Dialectologie, Volkskunde en Naamkunde.

Schwartz, J.-L., Boë, L.-J., Badin, P., & Sawallis, T. R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. *Journal of Phonetics*, *40*(1), 20–36.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics*, *25*(3), 255–286.

Stevens, K. (1980). Discussion. In *Proceedings of the 9th international congress of phonetic sciences* (Vol. 3, pp. 181–194).

Stevens, K. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and speech*, *3*(1), 32–49.

Van Son, R. J., Beinum, F. J. K.-v., & Pols, L. C. (1998). Efficiency as an organizing principle of natural speech. In *Fifth international conference on spoken language processing.*

Vaux, B., & Samuels, B. (2015). Explaining vowel systems: dispersion theory vs natural selection. *The Linguistic Review*, *32*(3), 573–599.