

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

Xinyu Zhang

Under the Supervision of

Paul Boersma

Second Reader:

Silke Hamann

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

University of Amsterdam

August, 2020
Student number: 12255394

Acknowledgments

Gratitude should never be summarized. But still, I would like to thank my supervisor Paul Boersma, for his teaching, mentoring, guidance, for being my Dutch informant in this small project, but mostly, for bearing with me. Some of the most delightful and most intellectually stimulating conversations in my life so far has happened in Paul's office during our meetings. Words won't do justice to the great fun I have been having studying from Paul.

I would also like to thank my second reader Silke Hamann, not only for expanding my interest from phonetics to phonology, but also for the teaching and mentoring since my very first Phonetics & Phonology class. Countless things wouldn't have been possible without Silke. This thesis included.

I can never thank Silke and Paul enough, but I will most definitely try from time to time.

Lastly, I thank Maggie for her mere existence.

To Maggie

Abstract

Dispersion theoretic analyses on sound change predict that small contrasts are likely to either merge into one category or become more dispersed diachronically. Previous research on adaptive dispersion is mostly done on vowel inventories. The current study makes an attempt at examining the change in the acoustic and auditory dispersion of Dutch voiceless sibilants, which are acoustically and perceptually similar to each other. Acoustic data was collected from two age groups of native speakers. Mixed-effect linear regression, spectral principal component analysis, as well as decision tree and random forest modeling were adopted. Results show that the two age groups investigated in the current study do differ how they contrast the two sibilants according to some metrics but not according to others.

Contents

1	Introduction	1
1.1	Topic and Goals	1
1.2	Outline	1
2	Some Phonetics and Phonology of Voiceless Sibilants	2
2.1	The Articulation and Acoustics of Voiceless Sibilants	2
2.2	Voiceless Sibilants in Dutch	4
2.3	Sibilant Inventories	6
3	Diachronic Change in Inventories	9
3.1	Sound Change	9
3.2	Adaptive Dispersion	10
4	A Phonetic Space for Voiceless Sibilants	11
4.1	Possible Dimensions	11
4.2	Two-Dimensional Mapping	12
5	Data Collection	13
5.1	Participants	13
5.2	Material and Design	14
5.3	Acoustic Analysis	15
5.3.1	Linear Mixed-Effects Models	18
5.3.2	Spectral Principal Component Analysis	20
5.3.3	Random Forests	28
5.4	Auditory Estimations	31
6	Discussion	32
	References	34
	Appendix A Participant Statistics	39
	Appendix B List of Stimuli	40
	Appendix C Post Test Questionnaire	44
	Appendix D Praat Script for Acoustic Measurements	47
	Appendix E R Script for Linear Models	51
	Appendix F R Script for Classification Trees and Random Forests	72
	Appendix G Full Results of Acoustic Measurements	95

1 Introduction

1.1 Topic and Goals

Since the 1970s, and as early as [Passy \(1891\)](#) and [Roudet \(1910\)](#), many scholars (e.g. [Liljencrants & Lindblom, 1972](#); [Lindblom, MacNeilage, & Studdert-Kennedy, 1983](#); [Disner, 1983](#); [Maddieson & Disner, 1984](#); [Vallée, 1994](#); [Flemming, 1995](#) et seq.; [Schwartz, Boë, Vallée, & Abry, 1997](#) et seq.; [Boersma, 1998](#) et seq.; [Boersma & Hamann, 2008](#); [Hauser, 2017](#); etc.) have looked at the universal trends of phoneme inventories more or less within the framework of dispersion theory. Most of the previous work has been in the realm of vowel inventories with the exception of e.g. [Schwartz, Boë, Badin, and Sawallis \(2012\)](#) 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 learning algorithm in [Boersma and Hamann \(2008\)](#)’s simulation is to some extent supervised in such a way that it was provided in the algorithm that the input is to be classified into two categories, while in reality newborns do not receive such instruction when acquiring phonemic inventories. Additionally, the model has not yet been tested on real-world data in any specific language, 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 repeatedly been raised. Hence, there is a possibility that the two sibilants are either becoming more dispersed or gradually merging into one category 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 thesis 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 adaptive dispersion. Section Four tries to define a phonetic space for the Dutch voiceless sibilants. The penultimate section describes the experiment and the analyses. 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 described/defined by (Ladefoged & Johnson, 2011, p.14) as two articulators approximating closely and producing a partially obstructed airstream which ends up being turbulent. According to Ladefoged and Johnson (2011), there are two ways to produce said “turbulent airflow”: it could be because the air is passing through a narrow constriction such as that between the upper teeth and the lower lip in the production of [f], or more intensely; it could be that the air first passes through a narrow gap of the aforementioned kind at the same time being speeded up, and was subsequently directed towards a sharp edge, such as the edge of the teeth, as in the production of the English [s]. Conventionally, the latter kind is categorized as sibilants, described by Ladefoged and Maddieson (1996, p.138) 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, energy in fricatives are randomly distributed in terms of frequency ranges, and sibilants have more acoustic energy at a higher pitch than the other fricatives (Ladefoged & Johnson, 2011). For instance, [ʃ] will have a lower pitch than [s] for two reasons: that the articulatory airstream in [ʃ] has a lower velocity due to the less constriction in the vocal apparatus, and that the added lip-rounding involved in the production of [ʃ] lengthens the vocal tract. Fricatives and sibilants could, of course, also be sub-divided by voicing, but since the current work 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, palato-alveolar, or alveolo-palatal.

According to [Hughes and Halle \(1956, p.308-309\)](#), the most useful measurement 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 10kcps”. Similarly, [Stevens \(1960\)](#) investigated isolated and lengthened voiceless fricatives, including some of which that would usually only occur in para-linguistic 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 palato-alveolar /ʃ/ is articulated close to the velum, a velar pinch may be expected for some vowels where the second and third formants approach each other, as it usually happens immediately before and after velar consonants when the tongue approaches the back of the mouth. Besides the influence of the sibilant on the surrounding vowels, 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 also 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 \(2008\)](#) states 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 to the 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 Dutch¹ phonology is not in agreement on the phonemic status of the palatal sibilant /ç/, which is sometimes also transcribed as /ʃ/².

Mees and Collins (1982, p.6) describes that the sequence /sj/ is realized as an alveolo-palatal fricative [ç] in Standard Dutch (*Algemeen Beschaafd Nederlands*, or ABN for short), and that it differs from the /ʃ/ in English, French, or German in that there is no labialization in the Dutch [ç]. According to Mees and Collins (1982), 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 do acknowledge that there are arguments for regarding /sj/ as an additional phoneme /ç/. However, the example they give of such an argument is a comparison of English and Dutch pronunciations in an English pronunciation guide for Dutch speakers by Gussenhoven and Broeders (1976), which is more of a phonetic comparison between the sounds in English and Dutch than a phonemic description.

In a description of Dutch phonology, Booij (1999, p.7) listed /s/ as the sole voiceless sibilant in the Dutch consonant inventory and analyzes [ʃ, ʒ, ç, ɲ] as /s,z,t,n/ palatalized before /j/, and the postalveolar fricatives that occur in loan-words such as *chique* [ʃik] and *jury* [ʒy:ri] as “phonologically, combinations of /s, z/ and /j/ with the fricatives realized as the postalveolar allophones”.

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 sibilant phonemes 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 Amsterdams”, or Broad Amsterdam Speech, before a word boundary or morpheme boundary, [s] is often palatalized when preceded by the short vowels /ɑ/, /ɛ/, /ɐ/, or /ɪ/, and also when it is at an initial position in a word or a morpheme. However, participant 13 in the current study, who was born and raised in Amsterdam and has lived in Amsterdam all his life, does have a distance of 1358 Hz between the CoG of his two voiceless sibilants, which is even slightly higher than the mean CoG distance of 1347 Hz between /s/ and

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

²This non-/s/ voiceless Dutch sibilant will be transcribed as /ç/ 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)

/ç/ 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 that they examined, and the Dutch sibilants [s] and [ʃ] were treated as allophones with [s] as the “default consonant” (p.351). Their results show that although the boundary values may vary, the same metric is equally efficient at distinguishing the two phones regardless of their phonemic status. As such, whether [s] and [ʃ] are allophones of the same phoneme or two separate phonemes should not affect the topic at hand and therefore, will not be a major concern of the current study.

In terms of comparing the Dutch sibilants to sibilants in other languages, apart from the phonemic status of /ç/ mentioned at 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. In contrast, the Dutch /ç/ 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 [ç], 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 (Received Pronunciation)³ and the [s] and [ç] produced by one of the participants in this study.

³Extracted from *BBC Learning English* (<https://www.youtube.com/watch?v=htmkblboG9Q>)

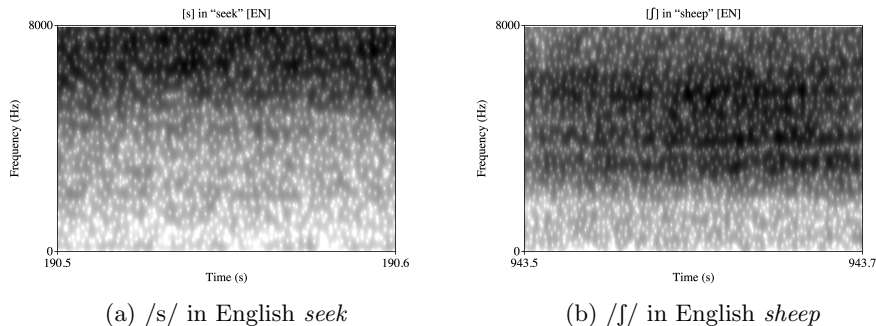


Figure 1: /s/ and /ʃ/ in English

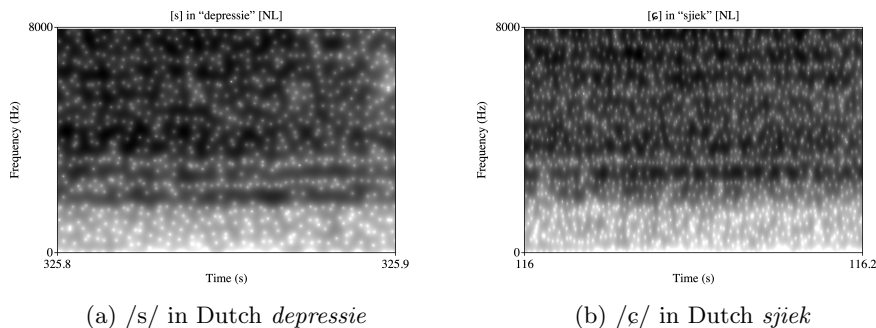


Figure 2: /s/ and /ɕ/ in Dutch

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 to Maddieson and Disner (1984), is the voiceless palato-alveolar sibilant /ʃ/. Schatz (1986, p.77) also mentioned that [s] is “reasonably frequent” 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) showed that /s/ is the least marked sibilant, and that in two-sibilant inventories, [s, ʃ] is maximally distinct, [s, ʃ] minimizes effort, and that [s, ɕ] only occurs when the weighted

ranking	Intensity	Frequency of Occurrence
1	ç	“s”
2	ʃ	ʃ
3	x	f
4	s	x
5	χ	χ
6	f	ϕ
7	θ	θ
8	ϕ	ç

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

F2 transition distance is high (when the assigned weight for the contribution of F2 transitions is larger than 0.87, in a component-weighting scheme similar to [Schwartz et al. \(1997\)](#)). The markedness of [s, ʃ], [s, ʒ] and [s, ç] pairings were not compared in [Flemming \(2018\)](#) since the former two are considered CoG-only contrasts (with zero weighted F2 distance difference) and [s, ç] is already harmonically bounded by [s, ʃ] and [s, ʒ] with the existence of “wF2=0” (i.e. when no weight was assigned to the contribution of F2 transitions). The explanation proposed by [Maddieson and Disner \(1984\)](#) for /s/ being the least marked sibilant is that languages possibly 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 [Stevens \(1960\)](#) to the frequency (i.e., rate of occurrence, not frequency in the sense of vocal fold vibrations per second) rankings of the fricatives (see Table 2). The results did not seem to indicate much correlation between intensity and the frequency of occurrence⁴.

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 [ç] 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

⁴However, note that according to [Maddieson and Disner \(1984, p.50\)](#), intensity readings in this table were “obtained and divided by subglottal air-pressure readings for the same tokens obtained using a nasal catheter inserted into the esophagus. From this procedure, a rank-order of intensity per unit air-pressure was obtained”.

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 Lindblom et al. (1983)’s simulations were “sensory discriminability” and “preference for ‘less extreme’ articulation”; the listener-based constraints used were “perceptual distance” and “perceptual salience” (ibid, p.191).

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) tends 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, by e.g., taking into account that F1 is perceptually more salient than F2 (see, e.g., [Diehl, Lindblom, & Creeger, 2003](#)).

Nonetheless, it is true for vowel inventories that the system is easier to acquire as well as to process when vowels are acoustically well-dispersed, for the reason that they are easier to discriminate. This creates an evolutionary tendency for languages to find themselves with such more dispersed inventories ([Joanisse & Seidenberg, 1998](#), p.335). Additionally, evidence such as the Hyperspace Effect ([Johnson, Flemming, & Wright, 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 is thus 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 variation in production between speakers, and such variation within what is considered by the listener as the same category is acceptable. Hence utterances within the acceptable variation range are perceived as the same sound. Hypo-correction is when enough individuals start producing outliers, and the outliers do not 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. 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.⁵

3.2 Adaptive Dispersion

Adaptive dispersion, according to the definition by Johnson (2000, p.181), refers to “the hypothesis that the distinctive sounds of a language tend to be positioned in phonetic space so as to maximize perceptual contrast”. Some scholars (e.g. Liljencrants & Lindblom, 1972; Boersma, 1998) also consider the interaction between production and perception constraints.⁶ This can be traced back to Martinet (1955, p.62)’s prediction of the general mechanism of the evolution of sound systems “*Les unités distinctives, les phonèmes qui coexistent, tendront à utiliser au mieux les latitudes que leur offrent les organes dits de parole; ils tendront à être aussi distants de leurs voisins qu’il est loisible pour eux de l’être tout en restant facile à articuler et facile à percevoir*”. From the previous sections, it can be predicted that the less optimally dispersed sound inventories (i.e., inventories where the perceptual distances between phonemes are not wide enough to maintain perceptual distinctiveness, or where the articulations of phonemes, e.g., in terms of manner, place, or both, are more extreme than necessary) are less likely to remain stable and more likely to become more optimally dispersed diachronically 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 [ʃ] and [ʒ] (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/ (Boersma, 1998, summarized in 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

⁵See Hamann (2009) for a counter-argument, and see Fruehwald (2017) for more details on how phonologization could happen simultaneously with the onset of phonetic changes.

⁶Flemming (2002) stresses that “articulatory representations have no bearing on distinctiveness”.

the acoustic and auditory dispersion of the Dutch voiceless sibilants is in consistency with the predictions made in previous work. A small contrast, such as that between the Dutch [s] and [ç], 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. The current study hypothesizes that the two sounds would become more dispersed rather than 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](#))⁷. 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 /ç/ in the Dutch consonant inventory to increase contrast from, if /s/ and /ç/ were to merge into one category, the condition does not fit with that which would enhance contrast.

4 A Phonetic Space for Voiceless Sibilants

As all dispersion models assume a phonetic space whereby phonetic distances are measured, in this section, I describe the acoustic measurements adopted as an attempt to define a phonetic space for voiceless sibilants, 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

⁷This observation is likely limited to mergers that enhance contrasts, as was stated in [Becker-Kristal \(2010\)](#). [Labov \(1994, p.321\)](#) did mention that mergers which are the results of chain shifting, as well as mergers caused by social stigma or prestige, do not always show such intermediate forms but instead rather have “the same mean value of one of the members of the merger, but with an enlarged class membership” sometimes. However, the potential merger under discussion here is not the result of chain shifting, and the current thesis does not focus on the socio-linguistic factors, though it is possible that they play a role (see §2.2).

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, diffuse-compact (e.g. [f] and [θ] being more diffuse, and [ʃ] more compact), and 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, palatalized fricatives in Russian have a higher intensity than non-palatalized fricatives. 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 their non-palatalized counterparts.

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 established than 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, diffuse-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, Barthmaier, and Sands (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” (p. 29). Diffuse-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 energy spreading out over a wider range of frequencies and consequently a wider peak or even multiple “diffuse” 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 Data Collection

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). See Appendix A for an anonymized list of participant age and gender. 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 had been raised in monolingual households with native Dutch-speaking parents. All participants reported to have no abnormalities in their vision or speech ⁸.

⁸Hearing was not specifically included in the pre-recording screening, since the task only involved reading the displayed sentences aloud (see §5.2). Nonetheless, the hearing of all

5.2 Material and Design

A list of 113 Dutch sentences was constructed (see Appendix B for the full list⁹). The sentences contained 33 tokens of /s/, and 26 tokens of /ç/¹⁰. Fifty-four sentences containing neither of the two target sibilants served as fillers. All but the utterance-final target sibilants were situated in intervocalic positions and stressed syllables. Among the /s/ tokens, 10 were word-initial, 7 were word-medial, and 17 were word-final. Among the /ç/ tokens, 13 were word-initial, 7 were word-medial, and 6 were word-final¹¹.

The interactive interface was written in E-Prime. The aforementioned list of sentences was 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 was then prompted to press a key to start the experiment. The prompting speed after this point 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 that they had finished reading and that the next sentence should be shown. 50 ms of delay was added as a buffer after the pressing of the key. 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.

The recordings were done in a soundproof studio in the Speech Lab at the University of Amsterdam, using a Sennheiser 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 between their mouth and the microphone, and the screen on which the stimuli were displayed was adjusted

participants was at a level to be able to communicate in English (their second language) with the experimenter with no difficulties at all.

⁹I thank Paul Boersma for being my Dutch informant in creating this list.

¹⁰There were instances where the participant pronounced the words *jus* and *jam* as [ʒy] and [ʒɛm] respectively. Such tokens were not used in the analysis.

¹¹Some stimuli contained more than one sibilant in more than one positions, e.g. *cynisch* has both word-initial /s/ and word-final /s/.

to their desired height and distance. 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 C) to gather information about the speaker’s language background, including their age, profession, birthplace, cities and towns that they lived in 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 from participants who might have guessed the targets and hyper-articulated during the recording. None of the participants guessed correctly or even close.

5.3 Acoustic Analysis

Acoustic measurements are done in PRAAT (Boersma & Weenink, 2020). 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 PRAAT script that fills in the TextGrid annotations of the underlyingly identical segments in different recordings (e.g., the /ç/’s in all the *koosjer* tokens in the recordings of different speakers). 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, therefore not annotated or extracted for measurements¹² (see Appendix D 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 transition as possible was included, as is shown in Figure 3, where part of the very beginning of /ç/ 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 that participants varied in speaking rate, the durations of the sibilants were not measured. Each relevant part (i.e., every annotated sibilant segment in every recording) was extracted with a rectangular window shape. Each of the 916

¹²I thank Dirk Jan Vet for the technical support and scripting help.

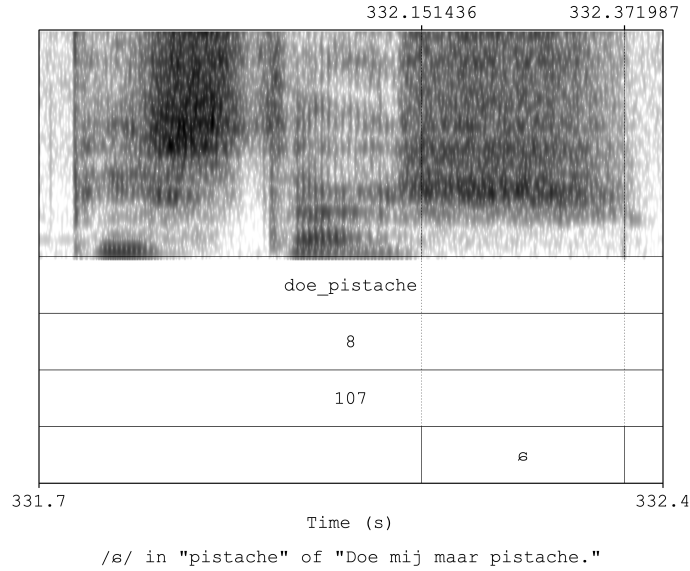


Figure 3: Spectrogram of “pistache” in the stimulus “Doe mij maar pistache” with dashed lines marking the duration segmented for and annotated as /ɛ/

extracted tokens was subjected to spectral analysis and passed through a stop Hann Band filter from the frequency of 0 Hz to 550 Hz¹³, 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 G for the full results of the acoustic measurements.

Figure 5 shows a scatter plot of sibilants produced by the speakers aged between 19 and 27, with /ɛ/ marked by red circles, and /s/ marked by blue plus signs, and one sigma ellipses numbered in the center by participant ID. 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 group aged between 61 and 75. The only overlapping of /ɛ/ and /s/ in the younger group occurs *between* different speakers (i.e. the /ɛ/ of Participant 11 overlaps with both the /s/ produced by Participant 10 and slightly with the /s/ produced by Participant 15), but

¹³Even though all segments under investigation were voiceless, considering that one of the measurements used would be the spectral center of gravity, the filter was still applied as a second guard besides segmentation to avoid the influence of voicing from surrounding vowels.

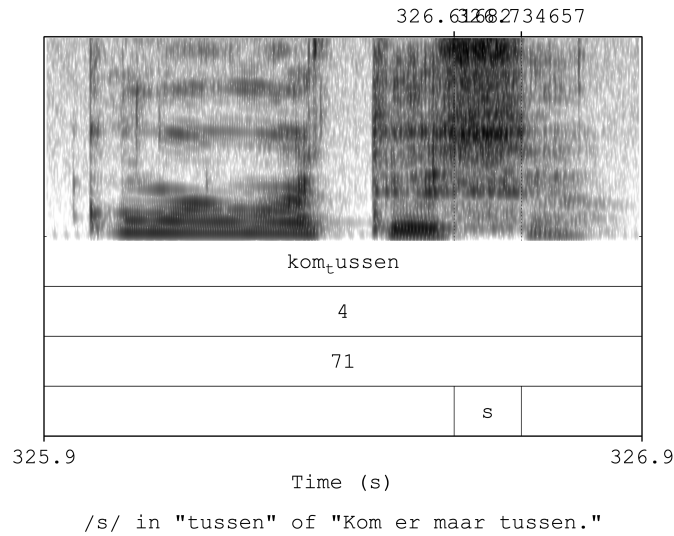


Figure 4: Spectrogram of “tussen” in the stimulus “Kom er maar tussen” with dashed lines marking the duration segmented for and annotated as /s/

never *within* speakers. Additionally, there is more variance in spectral standard deviation in the older group as well as more within-category variation in general in the older group.

In a slightly different scheme, Figures 7 and 8 show the ellipses of the sibilants produced by each participant in the two groups, assigning each participant with an ellipse of a different color, with the sibilants marked in the center of each ellipse. It can be seen from Figure 7 that the ellipses of the same colors never collide in the younger group. Whereas in the older group, the ellipses of the same colors are located much closer, as is shown in Figure 8. One of the colors has visible overlapping, and several others are very much closer together, indicating the overlapping of CoG and spectral standard deviation of the two sibilants produced by the older group. There is also a fair amount of between-speaker overlapping between the two sibilants in Figure 8. For instance, the black /ɛ/ locates completely within the blue /s/, and the CoG range of the green /ɛ/ is entirely within the CoG range of the red /s/. Moreover, in the older group, some /ɛ/’s are even higher in CoG compared to the /s/’s produced by a different speaker of the same group.

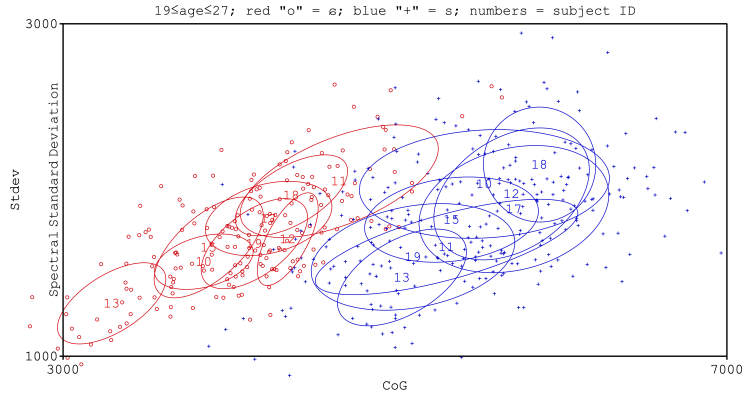


Figure 5: Scatter plot of the sibilants produced by younger speakers

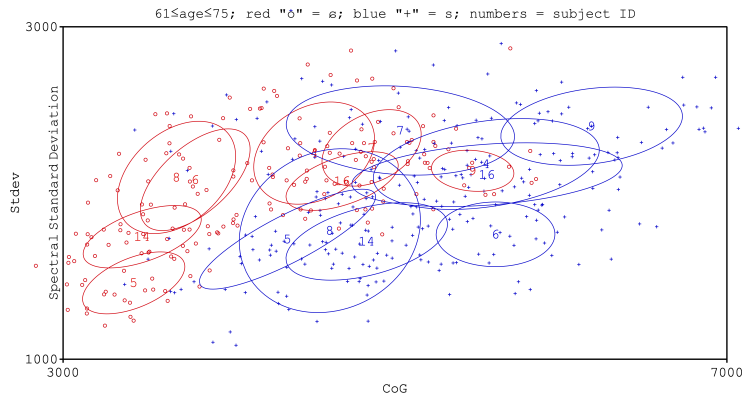


Figure 6: Scatter plot of the sibilants produced by older speakers

5.3.1 Linear Mixed-Effects Models

The data was analyzed in R (R Core Team, 2020) using 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 stimulus token and *Participant ID* were included as random effects. The height of the succeeding vowel was coded as a ternary predictor, with /æ/, /au/, /a/, /ai/, /ã/ coded as “low”, /ɔ/, /ɪ/, /o/, /ə/, /ɚ/, /au/ coded as “mid”, and /i/, /u/, /y/ coded as “high”. Rounding was coded as a binary predictor. Frontness was coded as a ternary predictor, with /æ/, /au/, /a/, /ai/, /ɪ/, /i/, /ã/, /y/ as “front”, /ə/ as mid, and /ɔ/, /o/, /u/, /ɚ/ as “back”. All levels of

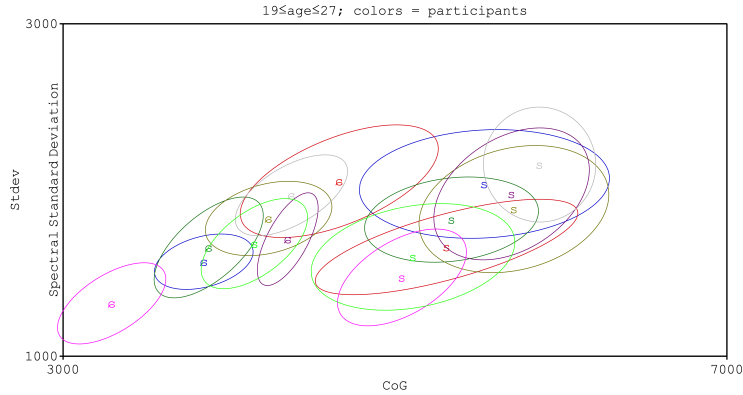


Figure 7: Ellipses of the sibilants produced by younger speakers, with each participant in one color

all contrasts are both orthogonal to each other and orthogonal to the intercept. See Appendix E for the formula and coding of contrasts and predictors.

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, on average, 1546.18 Hz higher in /s/ than in /ç/ (95% confidence interval = 1272.588 Hz .. 1818.093 Hz; $t = 9.981$) among all the Dutch speakers who participated. This fits the general expectation. The estimated mean for the interaction effect of AgeGroup and Sibilant is 391.03 Hz, meaning the CoG difference between the two sibilants is 391.03 Hz higher in the younger group than the older group on average. The effect 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 /ç/ 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 E for the formula and the

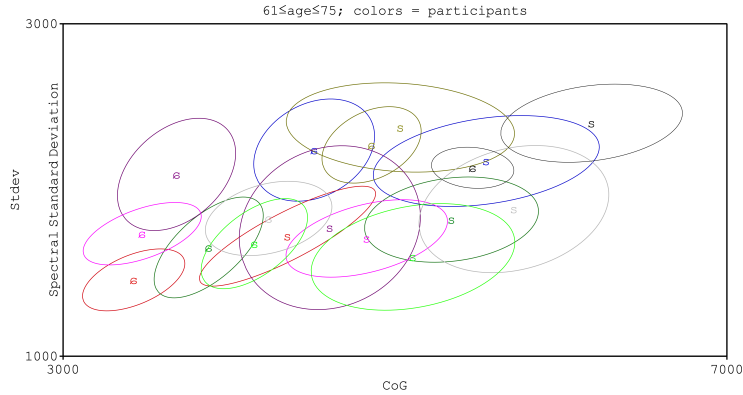


Figure 8: Ellipses of the sibilants produced by older speakers, with each participant in one color

contrast coding scheme). Results show that the difference of spectral standard deviation between the sibilants /s/ and /ʃ/ of the participants in the younger group is 389.80 Hz wider than that of the older group, which is statistically 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, especially when retroflexion is involved, as is pointed out in [Hamann \(2003\)](#) and shown in [Harris \(1954\)](#).

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 purpose of 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 suggested in [Evers et al. \(1998\)](#). 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. Figure 9 shows the eigenvalues of each of the principal components on a scree plot.

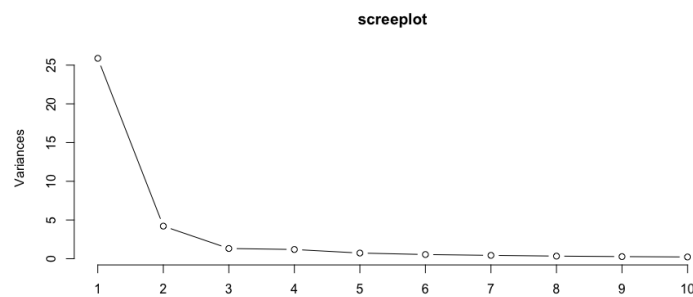


Figure 9: Scree plot of the first 10 components

Figures 10 to 13 show eigenvectors 1 to 4. As was explained above, the elements on the x-axis represent frequency bins with a width of 250 Hz, 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. This distinction is irrelevant for the purpose of investigating spectral shape.

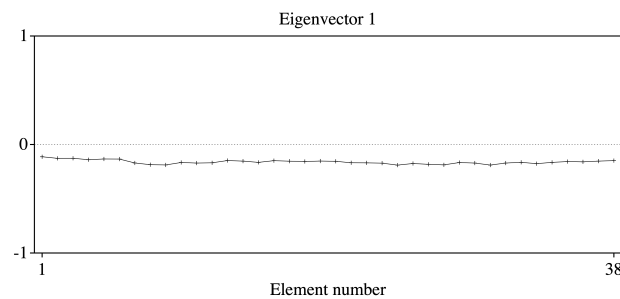


Figure 10: Eigenvector 1 of all speakers and both sibilants

The second eigenvector has one zero-crossing near bin 14, which shows that

eigenvector 2 is an indication of whether the energy level is on average higher in the frequency range below or above $550 \text{ Hz} + 14 * 250 \text{ Hz} = 4050 \text{ Hz}$. This is slightly lower than the threshold of 4.2k Hz in [Hughes and Halle \(1956\)](#), mentioned in §2.1.

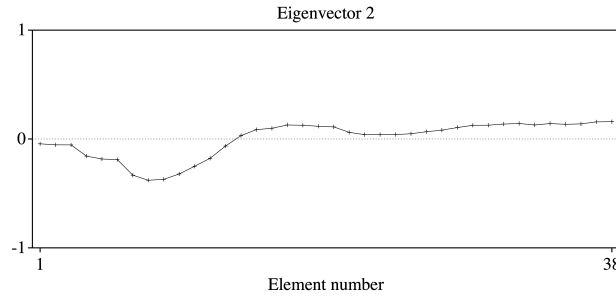


Figure 11: Eigenvector 2 of all speakers and both sibilants

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

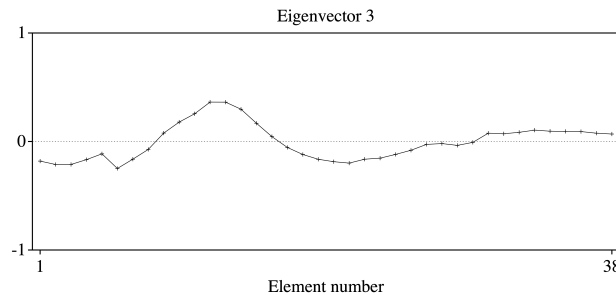


Figure 12: Eigenvector 3 of all speakers and both sibilants

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 14 is the sibilants plotted according to their principal component scores of the second and third principal components. The marks “sy”, “cy”, “so” “co” represents data points of [s]’s and [ç]’s produced by

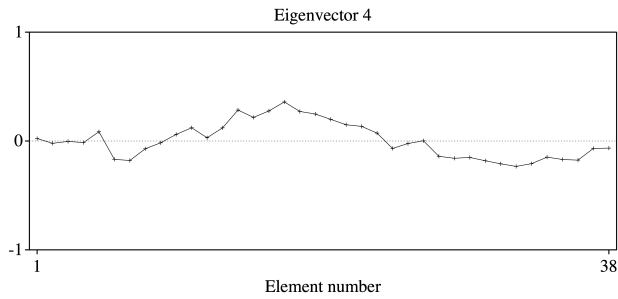


Figure 13: Eigenvector 4 of all speakers and both sibilants

“y”ounger and “o”lder speakers among the participants, respectively. It can be seen from the scatter plot that there is some overlap between the 1 SD ellipses of “ ɛo ” and “ so ”, but a wide gap between the edges of the “ ɛy ” and “ sy ” ellipses, denoting that the acoustic distance between the two sibilants is indeed wider in the younger generation.

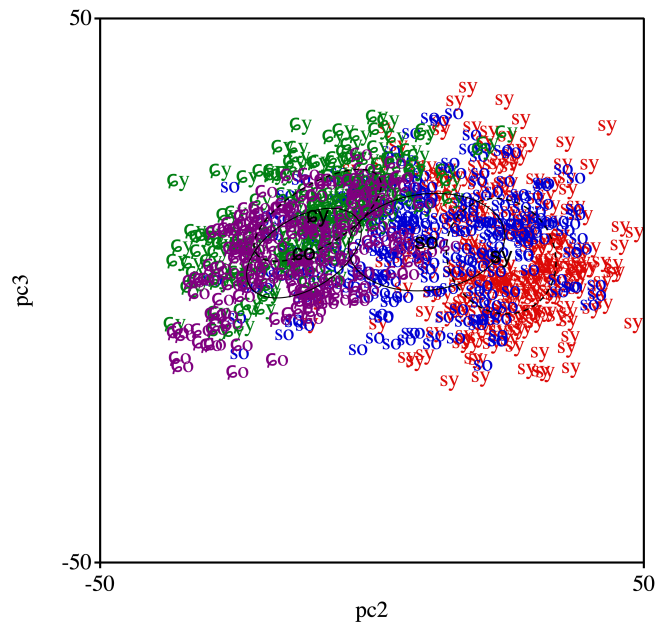


Figure 14: Two sibilants produced by two age groups, plotted according to the principal component scores of the second and third principal components

Age Group Data

To better compare how the two age groups differ in the way they differentiate the two sibilants in production, one separate principal component analysis was run for each age group. Figures 15a and 15b show the scree plots of the younger group and the scree plot of the older group, respectively.

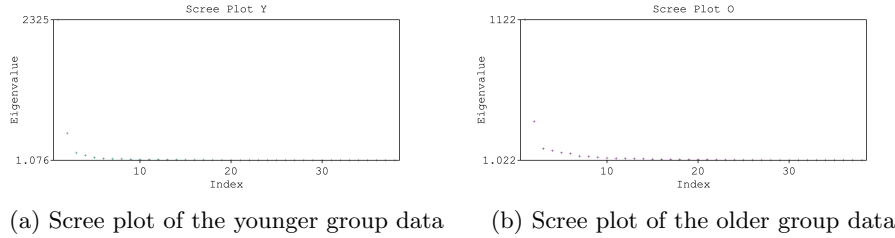


Figure 15: Scree plots for the age group data

Figures 16a and 16b show the first eigenvector from the spectral principal component analysis of the two age groups. Figures 17 and 18 show the second eigenvector of the two age groups, respectively.

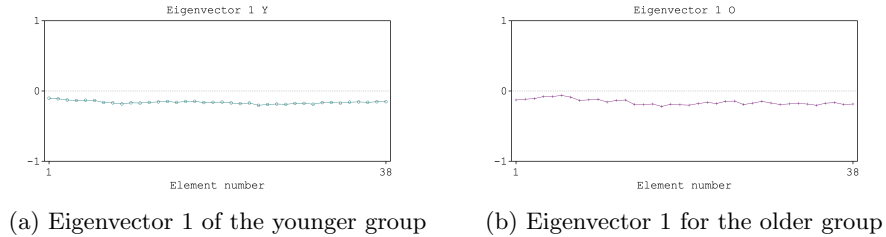


Figure 16: Eigenvector 1 for the age group data

There is no fundamental difference in the first two eigenvectors between age groups, except that there is some distance in bin 13 ($550 \text{ Hz} + 13 * 250 \text{ Hz} = 3800 \text{ Hz}$). Looking at each of the two age groups separately, the younger group has a zero crossing at a frequency range slightly higher than bin 14, and the older group has a zero crossing at bin 13. 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 14 ($550 \text{ Hz} + 14 * 250 \text{ Hz} = 4020 \text{ Hz}$), while for the older group, the threshold is slightly above bin 13 ($550 \text{ Hz} + 13 * 250 \text{ Hz} = 3800 \text{ Hz}$).

It can be seen in the third eigenvectors (shown in Figures 19 and 20) that

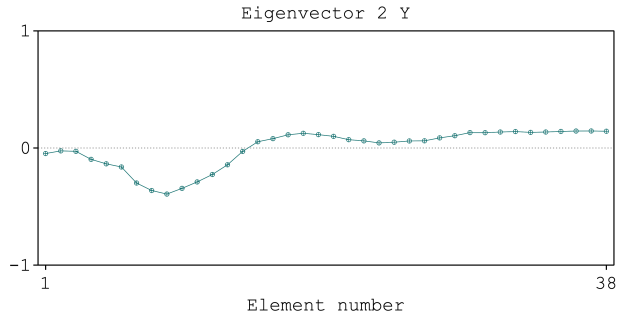


Figure 17: Eigenvector 2 of both sibilants produced by the younger group

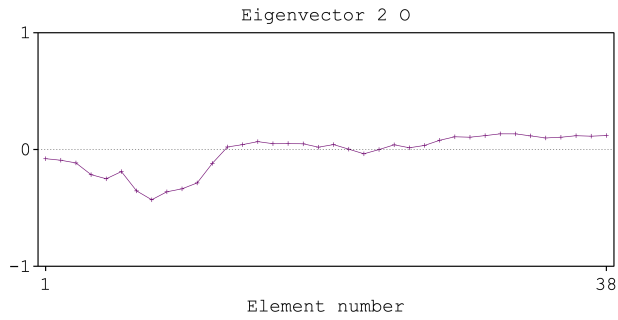


Figure 18: Eigenvector 2 of both sibilants produced by the older group

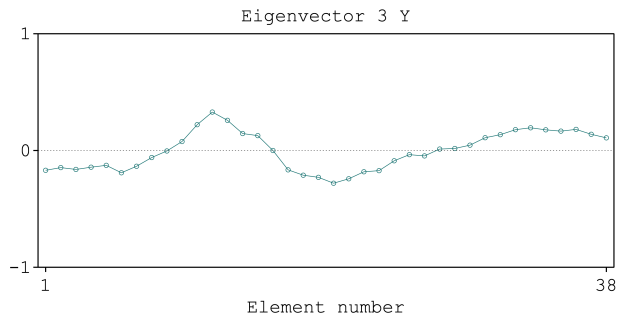


Figure 19: Eigenvector 3 of both sibilants produced by the younger group

the younger group has a prominent peak between bin 9 to bin 16 (i.e. $550 \text{ Hz} + 9 \cdot 250 \text{ Hz} = 2800 \text{ Hz}$ to $550 \text{ Hz} + 16 \cdot 250 \text{ Hz} = 4550 \text{ Hz}$) marked by two zero crossings, and a valley between bin 16 and bin 27 (i.e. $550 \text{ Hz} + 16 \cdot 250 \text{ Hz} = 4550 \text{ Hz}$ to $550 \text{ Hz} + 27 \cdot 250 \text{ Hz} = 7300 \text{ Hz}$), while the peak in the older

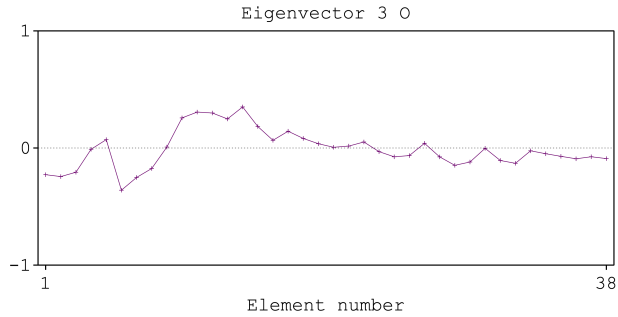


Figure 20: Eigenvector 3 of both sibilants produced by the older group

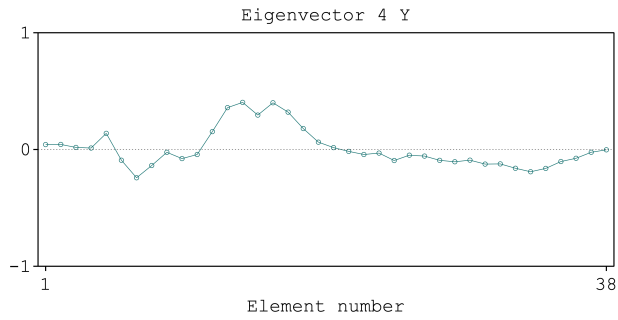


Figure 21: Eigenvector 4 of both sibilants produced by the younger group

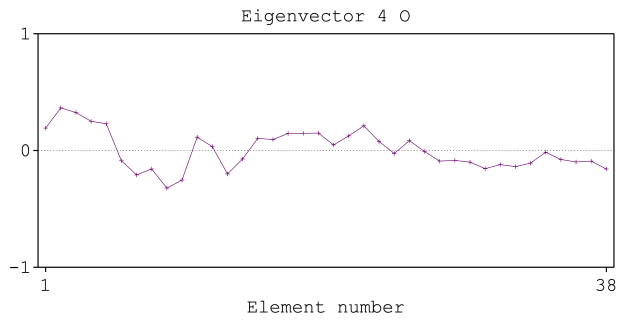


Figure 22: Eigenvector 4 of both sibilants produced by the older group

group is less sharp and more prolonged, between bin 9 to bin 23 (i.e. $550 \text{ Hz} + 9 \cdot 250 \text{ Hz} = 2800 \text{ Hz}$ to $550 \text{ Hz} + 23 \cdot 250 \text{ Hz} = 6300 \text{ Hz}$), in addition to a valley between bin 5 and bin 9 (i.e. $550 \text{ Hz} + 5 \cdot 250 \text{ Hz} = 1800 \text{ Hz}$ to $550 \text{ Hz} + 9 \cdot 250 \text{ Hz} = 2800 \text{ Hz}$). That is, the younger group differentiates the two sibilants by

the energy difference between the frequency ranges of 550 to 2800 Hz, 2800 to 4550 Hz, 4500 to 7300 Hz, and above 7300 Hz; the older group differentiates the two sibilants by the energy differences between the ranges of 1800 to 2800 Hz, 2800 Hz to 6300 Hz, and above 6300 Hz.

Additionally, the most prominent contrast in energy between the two sibilants is between the points of bin 6 and bin 14 (i.e., energy at 2050 Hz and energy at 4050 Hz) for the older group, while for the younger group the contrast is between the points of 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 are different between the two age groups.

As principal components 2 and 3 are the two components that should differentiate the two sibilants the best besides principal component 1, which is irrelevant for spectral shape, I now scatter plot each token of /s/ and /ʃ/ in each age group by their eigenvalues in the second and third principal components in Figures 23 and 24.

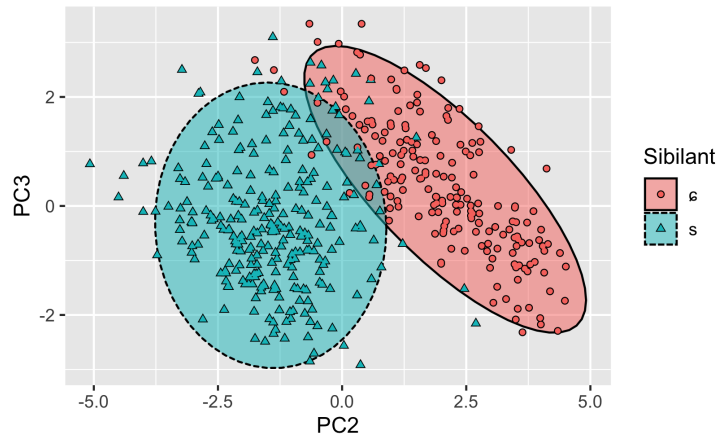


Figure 23: Both sibilants produced by the younger group plotted by their PC scores in principal components 2 and 3

The second and third principal components are the two principal components with the highest eigenvalues that are of interest in this study for both age groups. It can be seen from the scatter plots that the second principal component of the younger group can mostly separate the two sibilants. However, for the

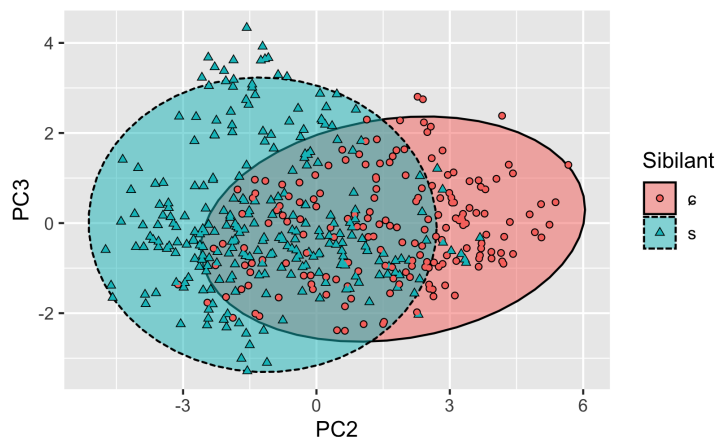


Figure 24: Both sibilants produced by the older group plotted by their PC scores in principal components 2 and 3

older participants, even by looking at the two group-specifically most important principal components, neither of them classify the two sibilants very well.

5.3.3 Random Forests

Since principal component analyses have the drawback of being subject to over-fitting, and its result becomes difficult to interpret once the eigenvectors fluctuate, modeling by decision trees and random forests were done on the spectra. Decision tree learning is a supervised learning approach that has the benefit of, e.g., being able to handle collinearity in the data, having built-in feature selection, as well as producing more easily interpretable outcomes. Random decision forests correct the single decision trees' over-fitting to the training set by training a multitude of trees and using bootstrap aggregation in selecting the training set for each tree and then using a random subset of features when training each tree, before averaging across all trees that are trained to get a final model. In other words, each tree in the forest is trained on randomly re-sampled data, with a random subset of all the features, therefore producing a model that is less susceptible to over-fitting.

The spectra of each group were randomly divided into two parts: one training set (80%) and one test set (20%). The same long-term average spectra and the same frequency bins used for the principal component analyses were used for the decision tree training. The energy in each frequency bin was used as input.

The frequency bins were treated as parameters, and the sibilants corresponding to the spectra were used as labels. A total of 500 trees were trained for each age group, and a random forest was grown for each age group by feature-bagging to reduce over-fitting. See Appendix F for the R code used for training the trees and forests).

A random forest model was fitted for each age group, Figures 25a and 25b show the difference in the ranking of frequency bins between two age groups. Tables 3 and 4 are the confusion matrices of the forests for each age group respectively.

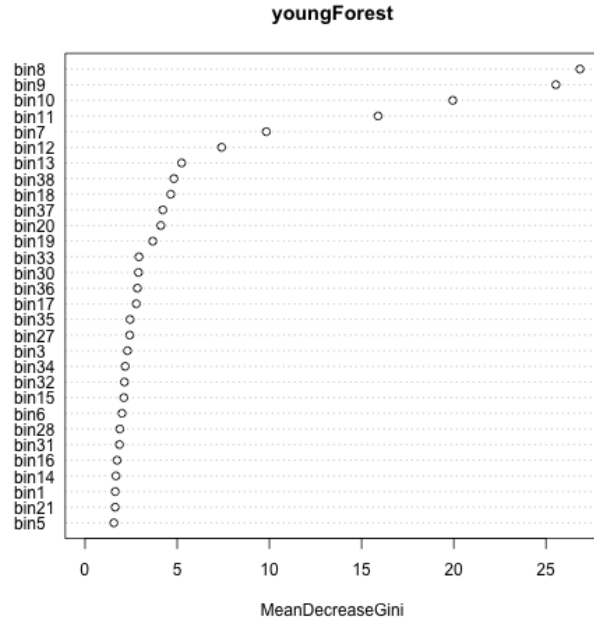
	Reference	
Prediction	ç	s
ç	40	1
s	0	53

Table 3: Confusion matrix for the younger participants

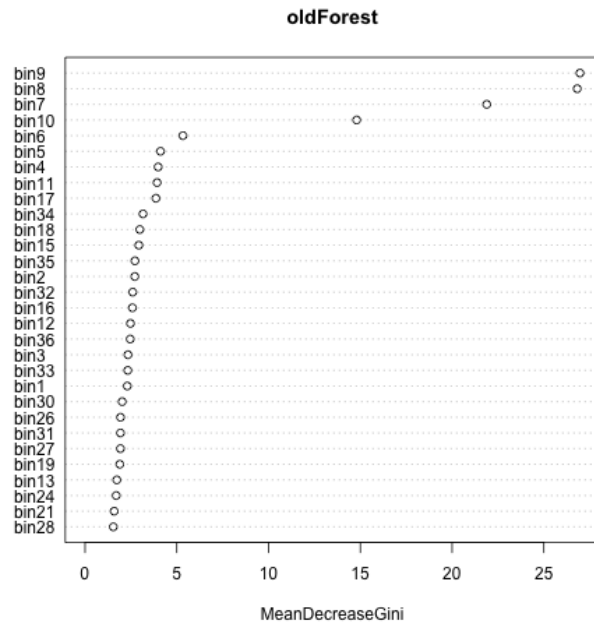
	Reference	
Prediction	ç	s
ç	41	1
s	2	45

Table 4: Confusion matrix for the older participants

The rankings of parameter importance show that the two age groups indeed used different rankings of frequency bins to distinguish the two sibilants. Additionally, the out-of-bag error rate for the older group is always higher than that of the younger group (the test accuracy of the random forests models for the older and younger groups are 0.987 and 0.989, respectively, and the Area Under Curve evaluated with the test sets of the older and younger groups in the random forest models are 0.985 and 0.998, respectively), indicating that the sibilants produced by the older age group are more difficult to classify, providing tentative support for the hypothesis that the two sibilants are more merged for the older speakers.



(a) Importance of frequency bins for younger speakers



(b) Importance of frequency bins for older speakers

Figure 25: Importance ranking of frequency bins used by two age groups to distinguish /s/ and /ʃ/

5.4 Auditory Estimations

In consideration of the potential role that perception plays in the dispersion-theoretic non-teleological diachronic changes (e.g. [Flemming, 1995](#) and [Boersma, 1998](#) both explicitly point out that the dispersion concerns *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 chose to convert the acoustic measurements into a psychoacoustically relatively 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, 6500 Hz]” in humans ([Greenwood, 1990](#), p.2601, in [Smith & Abel, 1999](#), p.705; see, also, [Moore, 1986](#), p.146, for the function, and [Moore, 1986](#), Fig. 6, for the frequency restriction). 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 formula¹⁴ below illustrates the process:

$$StdevErb = 0.5 * [hertzToErb(CoG + CoG_SD) - hertzToErb(CoG - CoG_SD)]$$

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 E for the R script and full results in detail).

Results show that the CoG difference (on the ERB scale) between the two

¹⁴In the format of the Praat scripting language.

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

Caveat

The residuals in the linear models are not normally distributed in the linear regression model for the width of the peak both on the Hertz scale and the ERB scale (termed as StdevHz, and StdevErb in the analyses above), neither were they Gaussian in the linear regression model for the CoG on ERB scale (CoGErb). Only the residuals of CoGHz are normally distributed. Hence the robustness of the relevant models might be affected. Due to the shape of the data, step-wise model comparison was also not implementable.

6 Discussion

The results from linear regression models in the current study, robustness aside, are inconclusive. Among the four dependent variables, only one showed a significant difference, namely the width of the spectral peak of the sibilants between age groups.

It is plausible that any generalization could be violated within a small subgroup and still holds true at population level, and since the participants from the current study come from rather limited socio-economic classes (i.e., university students and professors), more participants from more diverse backgrounds are needed for a more valid conclusion.

Despite the inconclusive results from the linear regression models, spectral analyses lend some support for a difference in the way that the two age groups differentiate the two sibilants in production. Random forest modeling confirms the difference.

Nevertheless, there are still some variables that were not well-controlled in the current study. For instance, there might be different degrees of reduction happening in different stimulus items, depending on the word frequency (Bybee, 2003), and neighborhood density (e.g., Goldrick, Vaughn, & Murphy, 2013; Fox, Reilly, & Blumstein, 2015), and sound change is very often first observed to

take place in high-frequency words of a language (Phillips, 1984; Morley, 2019). This can be improved by assigning a Word Frequency Score to each of the relevant words in each target stimulus and controlling for the effect of word frequency in the linear mixed-effect models. Additionally, the within-category variance in each sibilant was not well-incorporated in the current study. It might be worth investigating whether there are changes in within-category variance across generations, by e.g. the Jeffreys-Matsushita distance.

Counter-arguments can be made that such changes found in the current study might not necessarily be explained by adaptive dispersion, but rather by factors such as increased exposure to second languages such as English (where the acoustic/auditory distance between the two sibilants are further) in the younger generation. This is admittedly possible, and previous longitudinal research has shown phonetic change (VOT, F0) occurring in the first language as early as two weeks into a second language class (Chang, 2010). However, the same increased exposure is arguably also happening in the older generation at the same time instead of only affecting one of the two sub-populations. It might be worth further investigation to look into the difference of phonemic status of the two sibilants across generations, by, e.g., wug-testing and examining loanword adaptation involving the two sibilants across generations. One other aspect that may be of interest is the acoustic and auditory distance for different generations to consider two sibilants as different.

More importantly, in order to shed light on the (a)symmetry of production and perception, it is interesting to find out whether younger and older listeners also use different auditory cues to distinguish the two sibilants, in the same way as they do in production (see §5.3.2), as recent studies (e.g. Luthra, Correia, Kleinschmidt, Mesite, & Myers, 2020) even claim that acoustic information does not play any role at all in the perception of the /s/-/ʃ/ contrast.

References

- Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus* (PhD dissertation). University of California, Los Angeles.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Den Haag, Holland Academic Graphics/IFOTT.
- Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, 25(2), 217–270.
- Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer*. Retrieved from <http://www.praat.org>
- 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).
- Chang, C. B. (2010). *First language phonetic drift during second language acquisition* (PhD dissertation). University of California, Berkeley.
- Diehl, R. L., Lindblom, B., & Creeger, C. P. (2003). Increasing realism of auditory representations yields further insights into vowel phonetics. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 2, pp. 1381–1384).
- 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. (1995). *Auditory representations in phonology* (PhD dissertation). University of California, Los Angeles.
- Flemming, E. (2002). *Auditory representations in phonology*. Routledge.
- Flemming, E. (2017). Dispersion theory and phonology. In *Oxford Research Encyclopedia of Linguistics*.
- Flemming, E. (2018, Oct. 6). Systematic markedness in sibilant inventories. In *Annual meeting on phonology 2018*. San Diego, California. Poster. Retrieved from <http://phonology.ucsd.edu/program/sunday/posters-2/>
- Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*, 97–117.
- Fruehwald, J. (2017). The role of phonology in phonetic change. *Annual Review of Linguistics*, *3*, 25–42.
- Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America*, *134*(2), EL172–EL177.
- Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 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.
- Gussenhoven, C., & Broeders, A. (1976). *The pronunciation of English: A course for Dutch learners*. Wolters-Noordhoff-Longman.
- Hamann, S. (2003). *The phonetics and phonology of retroflexes*. Netherlands Graduate School of Linguistics.
- Hamann, S. (2009). The learner of a perception grammar as a source of sound change. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 111–149). Berlin: Mouton de Gruyter.
- Harris, K. S. (1954). Cues for the identification of the fricatives of American English. *The Journal of the Acoustical Society of America*, *26*(5), 952–952.
- 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 *The 27th Manchester Phonology Meeting*. Poster. Retrieved from <http://www.lel.ed.ac.uk/mfm/27mfm-program.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.
- Labov, W. (1994). *Principles of linguistic change: Volume 1: Internal factors*. Blackwell.
- Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Wadsworth, Cengage.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.
- Li, M. (2017). *Sibilant contrast: Perception, production, and sound change* (PhD 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. In L. M. Hyman, V. Fromkin, & C. N. Li (Eds.), *Language, speech, and mind: Studies in honour of victoria a. fromkin* (pp. 62–78). Taylor & Francis.
- 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.
- Luthra, S., Correia, J. M., Kleinschmidt, D. F., Mesite, L., & Myers, E. B. (2020). Lexical information guides retuning of neural patterns in perceptual learning for speech. *Journal of Cognitive Neuroscience*, 1–12.
- MacEachern, M., Kern, B., & Ladefoged, P. (1997). Wari' phonetic structures. *Journal of Amazonian Languages*, 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.
- Martinet, A. (1955). *Economie des changements phonétiques*. Francke, Bern.
- 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.
- Moore, B. (1986). Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. *Scand. Audiol. Suppl.* 25, 139-152., 25.
- Morley, R. (2019). *Sound structure and sound change: A modeling approach*. Language Science Press. Retrieved from <https://books.google.nl/books?id=Uci5DwAAQBAJ>
- 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. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.
- Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of American English speech: a dynamic approach*. Springer Science & Business Media.

- Passy, P. E. (1891). *Étude sur les changements phonétiques et leurs caractères généraux*. Firmin-Didot.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 320–342.
- R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Roudet, L. (1910). *Éléments de phonétique générale*. H. Welter.
- 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.
- Smith, J. O., & Abel, J. S. (1999). Bark and erb bilinear transforms. *IEEE Transactions on speech and Audio Processing*, 7(6), 697–708.
- 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.
- Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions* (PdD dissertation). Université Stendhal (Grenoble).
- 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.

A Participant Statistics

Participant ID	Age	Group	Gender
1	66	o	f
2	66	o	f
4	68	o	f
5	75	o	m
6	61	o	f
7	73	o	f
8	71	o	f
9	67	o	f
10	19	y	m
11	23	y	f
12	24	y	f
13	24	y	m
14	66	o	m
15	23	y	f
17	23	y	f
17	63	o	f
18	24	y	f
19	27	y	m

Table 5: Age and gender of the participants

B List of Stimuli

Sentence	ReclIndex
ik voel me somber	1
ik voel me blij	2
dat doen we samen	3
dat doen we apart	4
ik zoek de suiker	5
ik zoek de melk	6
je vindt me saai	7
je vindt me leuk	8
dat is mijn saldo	9
dit is mijn rondje	10
ik wil de saus	11
ik wil de mayonaise	12
ik wil de jus	13
ik wil de boter	14
doe mij maar sinas	15
doe mij maar cola	16
wat doe je cynisch	17
wat doe je raar	18
daar komt de sint	19
daar komt een klant	20
dat gaan we sussen	21
dat gaat ons lukken	22
ik voel me sjofel	23
ik vind je jofel	24
wat ben je sjiek	25
wat ben je lief	26
help me maar sjouwen	27
help me maar bouwen	28
een mooie sjaal	29
een mooie trui	30
doe mij maar Chinees	31
doe mij maar friet	32
dan gaan we sjoelen	33
dan gaan we dammen	34
ze zat te sjansen	35
ze zat te kijken	36
je loopt te sjokken	37
je loopt te hinken	38
doe mij maar jam	39
doe mij maar honing	40
ik zie de machine	41

ik zie de fabriek	42
ik lees de brochure	43
ik lees de folder	44
daar loopt je meisje	45
daar loopt je vriend	46
ik wil een flesje	47
ik wil een beker	48
da's niet zo koosjer	49
da's niet zo handig	50
knibbel knabbel knuisje	51
pief paf poef	52
ik pak een glaasje	53
ik pak een kopje	54
je bent het haasje	55
jij bent af	56
de gele hesjes	57
het groene hart	58
ik ga me douchen	59
ik ga me scheren	60
ik zal je missen	61
ik zal je bellen	62
in een depressie	63
in een deuk	64
wij vrouwen eisen	65
wij willen brood	66
ga lekker vissen	67
ga lekker tukken	68
ga je wassen	69
ga maar lopen	70
kom er maar tussen	71
kom er maar bij	72
dat gaan we sussen	73
ik wil de saus	74
ik ga onder de douche	75
ik ga in bad	76
ik loop naar de crèche	77
loop naar de maan	78
ze komen voor de hasj	79
ze komen voor de bollen	80
je maakt me boos	81
je maakt me blij	82
ik sta op het ijs	83
ik sta op de brug	84

je bent niet goed wijs	85
je bent niet goed snik	86
ik ga naar huis	87
ik ga naar bed	88
ik hang voor de buis	89
ik zit in de keuken	90
ik loop in het bos	91
ik loop door de duinen	92
een mooie poes	93
een mooie hond	94
doe mij maar kaas	95
doe mij maar ham	96
een rotte kies	97
een gat in mijn tand	98
ik lig in het gras	99
ik lig op de grond	100
een mooie bloes	101
een mooi hemd	102
kom naar de les	103
kom naar de fuif	104
het is hier niet pluis	105
ik vind het hier eng	106
doe mij maar pistache	107
doe mij maar vanille	108
ze plakt aan het pluche	109
ze zit in de kamer	110
doe mij maar sinas	111
wat doe je cynisch	112
doe mij maar Chinees	113

C Post Test Questionnaire

Post-Test Questionnaire

Thank you for participating in our experiment. In order to interpret the results as best as we can, we would like you to answer the questions below. Your data will be kept confidential, and we will make sure that no information will be revealed in ways that it could be traced back to you.

Personal Information

First Name:

Last Name:

Age:

Gender:

Profession:

Geographical Background

Birth Place:

In which cities in the Netherlands have you lived and for how long:

City 1, length of time:

City 2, length of time:

City 3, length of time:

City 4, length of time:

How long you have lived (if longer than 6 months) *outside of* the Netherlands, and where:

Language Background

Were you brought up in a monolingual home?

Yes No

If "no", please indicate your native languages other than Dutch:

Have you learned any second languages? For example in school, university, from friends etc.? For each language, please indicate the name and level of proficiency: (see next page)

Language 1:

Proficiency:

- Near-Native
- Fluent
- Intermediate
- Beginner

Language 2:

Proficiency:

- Near-Native
- Fluent
- Intermediate
- Beginner

Language 3:

Proficiency:

- Near-Native
- Fluent
- Intermediate
- Beginner

The Experiment

Please indicate below what you think this experiment was about:

Thank you for your participation!

D Praat Script for Acoustic Measurements

```

path$ = "Result"

binNames$ = ""
for ibin from 1 to 38
    binNames$ += " bin" + string$ (ibin)
endfor
result = Create Table with column names: "Result", 0, "Subject Age Index IPA Sibilant
CoG Stdev Position Vp Vf CoGErb StdevErb" + binNames$
resultPtr = 0

dl = Create Strings as directory list: "directoryList", path$
dl_nos = Get number of strings
for dir from 1 to dl_nos
    selectObject: dl
    dirName$ = Get string: dir
    @AnalyzeFolder (dirName$)
endfor

procedure AnalyzeFolder (.dirName$)

.r = Read from file: path$ + "/" + .dirName$ + "/" + .dirName$ + ".wav"
.tg = Read from file: path$ + "/" + .dirName$ + "/" + .dirName$ + ".TextGrid"

.dt = Down to Table: "no", 6, "yes", "no"
Rename: "Full"
.exName = Extract rows where column (text): "tier", "is equal to", "Name"
.exName_nor = Get number of rows
Rename: "Name"

for .i to .exName_nor
    selectObject: .exName
    .tmin = Get value: .i, "tmin"
    .tmax = Get value: .i, "tmax"

    selectObject: .r
    .snd2 = Extract part: .tmin, .tmax, "rectangular", 1, "no"
    .duration = Get total duration
    .midTime = .duration/2
    selectObject: .tg
    .tg2 = Extract part: .tmin, .tmax, "no"

    .interval = Get interval at time: 1, .midTime
    .name$ = Get label of interval: 1, .interval
    .interval = Get interval at time: 2, .midTime
    .subject$ = Get label of interval: 2, .interval
    .interval = Get interval at time: 3, .midTime
    .index = Get label of interval: 3, .interval

    .dt2 = Down to Table: "no", 6, "yes", "no"
Rename: "Sentence"

```

```

"IPA"
.ipa = nowarn Extract rows where column (text): "tier", "is equal to",
.ipa_nor = Get number of rows
for .row to .ipa_nor
  selectObject: .ipa
  .tmin2 = Get value: .row, "tmin"
  .tmax2 = Get value: .row, "tmax"
  .ipa$ = Get value: .row, "text"

  selectObject: .snd2
  .snd3 = Extract part: .tmin2, .tmax2, "rectangular", 1, "no"
  Rename: "IPA"
  .spec = To Spectrum: "yes"
  Filter (stop Hann band): 0, 550, 50
  .cog = Get centre of gravity: 2
  .stdev = Get standard deviation: 2
  .ltas = To Ltas: 250

  selectObject: result
  if index (.ipa$, "s")
    sibilant$ = "s"
  elsif index (.ipa$, "ɛ" )
    sibilant$ = "ɛ"
  else
    exitScript: "Found neither an s nor a ɛ in text ", .ipa$,
    "in interval", .interval, "of TextGrid", .tg2, "."
  endif

  v1$ = left$ (.ipa$, index (.ipa$, sibilant$) - 1)
  v2$ = right$(.ipa$, length(.ipa$) - index (.ipa$, sibilant$))
  if v2$ = ""
    position$ = "final"
  elsif index (v2$, "ə") or .ipa$ = "ɛsi"
    position$ = "medial"
  else
    position$ = "initial"
  endif

  selectObject: result
  Append row
  resultPtr += 1
  Set string value: resultPtr, "Subject", .subject$
  Set numeric value: resultPtr, "Index", .index
  Set string value: resultPtr, "CoG", fixed$ (.cog, 0)
  Set string value: resultPtr, "Stdev", fixed$ (.stdev, 0)
  Set string value: resultPtr, "IPA", .ipa$
  Set string value: resultPtr, "Vp", v1$
  Set string value: resultPtr, "Sibilant", sibilant$
  Set string value: resultPtr, "Vf", v2$
  Set string value: resultPtr, "Position", position$

```

```

        Set string value: resultPtr, "CoGErb", fixed$(hertzToErb(.cog), 0)
        Set string value: resultPtr, "StdevErb",
fixed$(0.5*(hertzToErb(.cog + .stdev)- hertzToErb(.cog - .stdev)), 0)
        for ibin to 38
            Set numeric value: resultPtr, "bin" + string$ (ibin),
object [.ltas, ibin + 2]
            endfor
            selectObject: .snd3, .spec, .ltas
            Remove
        endfor
        selectObject: .snd2, .tg2, .dt2, .ipa
        Remove
    endfor
    selectObject: result
    Formula: "Age", ~ if self ["Subject"] <= 9 or self ["Subject"] = 14 or self
["Subject"] = 16 then 0 else 1 fi

        selectObject: .dt, .exName
        Remove
endproc

```

E R Script for Linear Models

(Some identical R warnings omitted.)

Statistical Analysis

Xinyu Zhang

Participants and Stimuli stats:

participants:

```
#filtering out younger speakers  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
participants <- read.csv("participants.csv")  
young <- participants%>%  
  #count(Group == "y")  
  filter(Group == "y")  
young
```

```
##   ID Age Group Gender  
## 1 10 19     y      m  
## 2 11 23     y      f  
## 3 12 24     y      f  
## 4 13 24     y      m  
## 5 15 23     y      f  
## 6 17 23     y      f  
## 7 18 24     y      f  
## 8 19 27     y      m
```

```
#arranging Age by ascending order for min and max  
young%>%  
  arrange(Age)
```

```
##   ID Age Group Gender  
## 1 10 19     y      m  
## 2 11 23     y      f  
## 3 15 23     y      f  
## 4 17 23     y      f  
## 5 12 24     y      f  
## 6 13 24     y      m  
## 7 18 24     y      f  
## 8 19 27     y      m
```

```
#mu & sigma  
mean(young$Age)
```

```
## [1] 23.375
```

```
sd(young$Age)
```

```
## [1] 2.199838
```

```
#counting female  
young%>%  
  filter(Gender == "f")
```

```
##   ID Age Group Gender  
## 1 11 23   y     f  
## 2 12 24   y     f  
## 3 15 23   y     f  
## 4 17 23   y     f  
## 5 18 24   y     f
```

```
#counting male  
young%>%  
  filter(Gender == "m")
```

```
##   ID Age Group Gender  
## 1 10 19   y     m  
## 2 13 24   y     m  
## 3 19 27   y     m
```

```
#filtering out older speakers & count total number  
participants <- read.csv("participants.csv")  
old <- participants%>%  
  filter(Group == "o")  
old
```

```
##   ID Age Group Gender  
## 1  1 66   o     f  
## 2  2 66   o     f  
## 3  4 68   o     f  
## 4  5 75   o     m  
## 5  6 61   o     f  
## 6  7 73   o     f  
## 7  8 71   o     f  
## 8  9 67   o     f  
## 9 14 66   o     m  
## 10 17 63  o     f
```

```
#arranging age by ascending order for min and max
old%>%
  arrange(Age)
```

```
##      ID Age Group Gender
## 1     6  61     o      f
## 2    17  63     o      f
## 3     1  66     o      f
## 4     2  66     o      f
## 5    14  66     o      m
## 6     9  67     o      f
## 7     4  68     o      f
## 8     8  71     o      f
## 9     7  73     o      f
## 10    5  75     o      m
```

```
#mu and sigma
mean(old$Age)
```

```
## [1] 67.6
```

```
sd(old$Age)
```

```
## [1] 4.325634
```

```
#counting females
old%>%
  filter(Gender == "f")
```

```
##      ID Age Group Gender
## 1     1  66     o      f
## 2     2  66     o      f
## 3     4  68     o      f
## 4     6  61     o      f
## 5     7  73     o      f
## 6     8  71     o      f
## 7     9  67     o      f
## 8    17  63     o      f
```

```
#counting males
old%>%
  filter(Gender == "m")
```

```
##      ID Age Group Gender
## 1     5  75     o      m
## 2    14  66     o      m
```

Stimuli Stats:


```
library(dplyr)
fulltable <- read.csv("ResultswSlope.csv")
#selecting subject 4 for description of stimulus because it is used as the model to automate all the rest
subject4 <-fulltable%>%
  filter(Subject == 4)
subject4%>%
  count(Sibilant == "s", Position == "initial")
```

```
##   Sibilant == "s" Position == "initial"  n
## 1          FALSE                FALSE 12
## 2          FALSE                TRUE  13
## 3           TRUE                FALSE 25
## 4           TRUE                TRUE  10
```

```
subject4%>%
  count(Sibilant == "s", Position == "medial")
```

```
##   Sibilant == "s" Position == "medial"  n
## 1          FALSE                FALSE 18
## 2          FALSE                TRUE   7
## 3           TRUE                FALSE 28
## 4           TRUE                TRUE   7
```

```
subject4%>%
  count(Sibilant == "s", Position == "final")
```

```
##   Sibilant == "s" Position == "final"  n
## 1          FALSE                FALSE 20
## 2          FALSE                TRUE   5
## 3           TRUE                FALSE 17
## 4           TRUE                TRUE  18
```

```
subject4%>%
  count(Sibilant == "ç", Position == "initial")
```

```
##   Sibilant == "ç" Position == "initial"  n
## 1          FALSE                FALSE 25
## 2          FALSE                TRUE  10
## 3           TRUE                FALSE 12
## 4           TRUE                TRUE  13
```

```
subject4%>%
  count(Sibilant == "ç", Position == "medial")
```

```
##   Sibilant == "ç" Position == "medial"  n
## 1          FALSE                FALSE 28
## 2          FALSE                TRUE   7
## 3           TRUE                FALSE 18
## 4           TRUE                TRUE   7
```

```
subject4%>%
  count(Sibilant == "ç", Position == "final")
```

```
##   Sibilant == "ç" Position == "final"  n
## 1          FALSE          FALSE  17
## 2          FALSE          TRUE   18
## 3           TRUE          FALSE  20
## 4           TRUE          TRUE    5
```

Acoustic Analysis:

```
fulltable <- read.csv("ResultswSlope.csv")
library(dplyr)
nobins<-fulltable %>%
  select(Subject, Age, Index, IPA, Sibilant, CoG, Stdev, Position, Vp, Vf, CoGErb, StdevErb)

nobins$Sibilant <- as.factor(nobins$Sibilant)

nobins$Age <- as.factor(nobins$Age)
#saving the csv:
write.csv(nobins, file = "nobins.csv")
```

coding for features in vowels, and contrasts:

```

library(dplyr)

#high-low only contrast
erbs <- read.csv("ResultswSlope.csv")
hrxf <- erbs %>%mutate(
  heightP = case_when(
    Vp == "ɛin" ~ "h",
    Vp == "y" ~ "h",
    Vp == "i" ~ "h",
    Vp == "u" ~ "h",
    Vp == "ɛi" ~ "h",
    Vp == "æy" ~ "h",
    Vp == "au" ~ "h", #"a" w/ a hat but not distinguished because of font in R
    Vp == "ei" ~ "l",
    Vp == "ou" ~ "m",
    Vp == "ə" ~ "m",
    Vp == "ɔ" ~ "m",
    Vp == "ɛ" ~ "m",
    Vp == "aɪ" ~ "l", #"a" w/ a hat but not distinguished because of font in R
    Vp == "a" ~ "l" #"a" w/ a hat but not distinguished because of font in R
  )
)
hrxf$heightP <- as.factor(hrxf$heightP)

hrxf <- hrxf %>%mutate(
  heightF = case_when(
    Vf == "æy" ~ "l",
    Vf == "ɔ" ~ "m",
    Vf == "a" ~ "l", #"a" w/ a hat but not distinguished because of font in R
    Vf == "ai" ~ "l", #"a" w/ a hat but not distinguished because of font in R
    Vf == "ɪ" ~ "m",
    Vf == "o" ~ "m",
    Vf == "i" ~ "h",
    Vf == "u" ~ "h",
    Vf == "ā" ~ "l", #"a" w/ a hat but not distinguished because of font in R
    Vf == "y" ~ "h",
    Vf == "ə" ~ "m",
    Vf == "ɚ" ~ "m",
    Vf == "au" ~ "m" #"a" w/ a hat but not distinguished because of font in R
  )
)
hrxf$heightF <- as.factor(hrxf$heightF)

#contrast for preceding vowel
contrast1 <- cbind(c(0.5, -0.5, 0), c(-1/3, -1/3, +2/3))#h,l,m
colnames(contrast1) <- c("+h-l", "-hl+m")
contrasts(hrxf$heightP) <- contrast1

#contrast for the following vowel
contrast2 <- cbind(c(0.5, -0.5, 0), c(-1/3, -1/3, +2/3))#h,l,m
colnames(contrast2) <- c("+h-l", "-hl+m")
contrasts(hrxf$heightF) <- contrast2

```

```

hrxf <- hrxf %>%mutate(
roundingP = case_when(Vp == "ɔ" ~ "r",
                      Vp == "ou" ~ "r",
                      Vp == "æy" ~ "r",
                      Vp == "u" ~ "r",
                      Vp == "y" ~ "r",
                      Vp == "au" ~ "r", #"a" w/ a hat but not distinguished because of
font in R
                      TRUE ~ "nr")
)

hrxf$roundingP <- as.factor(hrxf$roundingP)

hrxf <- hrxf %>%mutate(
roundingF = case_when(Vf == "ɔ" ~ "r",
                      Vf == "æy" ~ "r",
                      Vf == "o" ~ "r",
                      Vf == "u" ~ "r",
                      Vf == "y" ~ "r",
                      TRUE ~ "nr")
)

hrxf$roundingF <- as.factor(hrxf$roundingF)

hrxf <- hrxf %>%mutate(
rhoticityP = case_when(
  Vp == "aɪ" ~ "x", #"a" w/ a hat but not distinguished because of font in R
  TRUE ~ "nx"
)
)

hrxf$rhoticityP <- as.factor(hrxf$rhoticityP)

hrxf <- hrxf %>%mutate(
rhoticityF = case_when(
  Vf == "ə" ~ "x",
  TRUE ~ "nx"
)
)

hrxf$rhoticityF <- as.factor(hrxf$rhoticityF)

#frontness of preceeding vowel
hrxf <- hrxf %>%mutate(
frtP = case_when(
  Vf == "ɔ" ~ "b",
  Vf == "ou" ~ "b",
  Vf == "u" ~ "b",
  Vf == "au" ~ "b",
  Vf == "ə" ~ "m",
  TRUE ~ "f"
)
)

hrxf$frtnsP <- as.factor(hrxf$frtP)

```

```
#frontness of following vowel
hrxf <- hrxf %>%mutate(
  frtF = case_when(
    Vf == "ə" ~ "m",
    Vf == "ɔ" ~ "b",
    Vf == "o" ~ "b",
    Vf == "u" ~ "b",
    Vf == "ʊ" ~ "b",
    TRUE ~ "f"
  )
)

hrxf$frtnsF <- as.factor(hrxf$frtF)

library(lme4)
```

```
## Loading required package: Matrix
```

```

#contrast coding for the rounding of the preceeding vowel
contrast3 <- cbind (c(+0.5, -0.5))
colnames (contrast3) <- c("+nr-r")
contrasts(hrxf$roundingP) <- contrast3

#contrast coding for the rounding of the following vowel
contrast4 <- cbind (c(+0.5, -0.5))
colnames (contrast4) <- c("+nr-r")
contrasts(hrxf$roundingF) <- contrast4

#contrast coding for the rhoticity of the preceeding vowel
contrast5 <- cbind (c(+0.5, -0.5))
colnames (contrast5) <- c("+nx-x")
contrasts(hrxf$rhoticityP) <- contrast5

#contrast coding for the rhoticity of the following vowel
contrast6 <- cbind (c(+0.5, -0.5))
colnames (contrast6) <- c("+nx-x")
contrasts(hrxf$rhoticityF) <- contrast6

#contrast coding for the frontness of the preceeding vowel, using the schwa as the ref
erence level
contrast7 <- cbind (c(-0.5, +0.5, 0), c(-1/3, -1/3, +2/3))#b,f,m
colnames (contrast7) <- c("-b+f", "-bf+m")
contrasts(hrxf$frtnsP) <- contrast7

#contrast coding for the frontness of the following vowel, using the schwa as the ref
erence level
contrast8 <- cbind (c(-0.5, +0.5, 0), c(-1/3, -1/3, +2/3))#b,f,m
colnames (contrast8) <- c("-b+f", "-bf+m")
contrasts(hrxf$frtnsF) <- contrast8

#contrast coding for position, using the medials as the reference level
hrxf$Position <- as.factor(hrxf$Position)
contrast9 <- cbind(c(0, -0.5, +0.5),c(2/3, -1/3, -1/3))#f,i,m
colnames (contrast9) <- c("-i+m", "+f-im")
contrasts(hrxf$Position) <- contrast9

#contrast coding for Sibilants
#hrxf$Sibilant <- as.factor(hrxf$Sibilant)
contrast10 <- cbind(c(-0.5, +0.5))
colnames(contrast10) <- c("-ʃ+s")
contrasts(hrxf$Sibilant) <- contrast10

#contrast coding for Age
hrxf$Age <- as.factor(hrxf$Age)
contrast11 <- cbind(c(-0.5, +0.5))#0,1
colnames(contrast11 ) <- c("-0+1")
contrasts(hrxf$Age) <- contrast11

all <- hrxf %>% mutate(
  Gender = case_when(
    Subject == "10" ~ "M",
    Subject == "13" ~ "M",
    Subject == "19" ~ "M",
    Subject == "5" ~ "M",
    Subject == "14" ~ "M",

```

```
TRUE ~ "F")
)
#contrast coding for Gender
all$Gender <- as.factor(all$Gender)
contrast12 <- cbind(c(0.5, -0.5)) #F,M
colnames(contrast12) <- c("+F-M")
contrasts(all$Gender) <- contrast12

#checking whether ζ is treated as alphabetically preceding s
contrasts(hrxf$Sibilant)
```

```
## -ζ+s
## ζ -0.5
## s 0.5
```

```
write.csv(all, file = "forPraat.csv")
```

```
modelCoGERb <- lmer(CoGERb ~ Age*heightF*roundingF*frtnsF*rhoticityF*Sibilant + (Sibi  
lant | Subject) + (Sibilant|Index), data = all, REML = TRUE, na.action = "na.omit")
```

```
## fixed-effect model matrix is rank deficient so dropping 116 columns / coefficients
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.0160721 (tol = 0.002, component 1)
```

```
summary(modelCoGERb)
```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: COGERb ~ Age * heightF * roundingF * frtnsF * rhoticityF * Sibilant +
## (Sibilant | Subject) + (Sibilant | Index)
## Data: all
##
## REML criterion at convergence: 1291.6
##
## Scaled residuals:
## Min 1Q Median 3Q Max
## -3.8071 -0.6009 0.0639 0.6696 3.1470
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## Index (Intercept) 0.01401 0.1184
## Sibilant-Ç+s 0.04357 0.2087 -0.51
## Subject (Intercept) 0.75358 0.8681
## Sibilant-Ç+s 0.47182 0.6869 -0.69
## Residual 0.50463 0.7104
## Number of obs: 551, groups: Index, 35; Subject, 16
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 27.31405 0.36030 75.810
## Age-0+1 -0.51926 0.61370 -0.846
## heightF+h-1 0.05857 0.18139 0.323
## heightF-hl+m 0.19154 0.21068 0.909
## roundingF+nr-r 0.38800 0.25076 1.547
## frtnsF-b+f -0.50929 0.50364 -1.011
## frtnsF-bf+m -1.18799 0.50980 -2.330
## rhoticityF+nx-x 1.45602 0.63451 2.295
## Sibilant-Ç+s 2.44834 0.29015 8.438
## Age-0+1:heightF+h-1 -0.27244 0.29128 -0.935
## Age-0+1:heightF-hl+m -0.33572 0.31510 -1.065
## Age-0+1:roundingF+nr-r 0.32456 0.38795 0.837
## heightF+h-1:roundingF+nr-r -0.08816 0.33947 -0.260
## heightF-hl+m:roundingF+nr-r 0.89695 0.79821 1.124
## Age-0+1:frtnsF-b+f -1.10920 0.76850 -1.443
## Age-0+1:frtnsF-bf+m -1.00448 0.80252 -1.252
## Age-0+1:rhoticityF+nx-x 1.53292 0.95237 1.610
## Age-0+1:Sibilant-Ç+s 0.69369 0.50025 1.387
## heightF+h-1:Sibilant-Ç+s -0.47424 0.27792 -1.706
## heightF-hl+m:Sibilant-Ç+s -0.64965 0.58013 -1.120
## roundingF+nr-r:Sibilant-Ç+s 0.27191 0.34571 0.787
## frtnsF-b+f:Sibilant-Ç+s -0.87703 0.94297 -0.930
## Age-0+1:heightF+h-1:roundingF+nr-r 0.82026 0.58434 1.404
## Age-0+1:heightF-hl+m:roundingF+nr-r 1.77855 1.20050 1.482
## Age-0+1:heightF+h-1:Sibilant-Ç+s 0.11514 0.41972 0.274
## Age-0+1:heightF-hl+m:Sibilant-Ç+s -0.86785 0.88861 -0.977
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s 0.28064 0.54333 0.517
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s -1.52921 1.46547 -1.043
##
##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

```


##	2.5 %	97.5 %
## .sig01	0.00000000	0.17068075
## .sig02	-1.00000000	1.00000000
## .sig03	0.00000000	0.34136151
## .sig04	0.60832298	1.15112948
## .sig05	-0.88271761	-0.27914044
## .sig06	0.45416796	0.92749477
## .sigma	0.65929443	0.74543043
## (Intercept)	26.68419971	27.90466956
## Age-0+1	-1.67847923	0.63513211
## heightF+h-l	-0.26613210	0.33061701
## heightF-hl+m	-0.16476824	0.51994717
## roundingF+nr-r	0.00691897	0.82000538
## frtnsF-b+f	-1.37660281	0.25772032
## frtnsF-bf+m	-2.05939103	-0.40869779
## rhoticityF+nx-x	0.48085538	2.52760047
## Sibilant-Ç+s	1.96092771	2.94739946
## Age-0+1:heightF+h-l	-0.83602244	0.29131773
## Age-0+1:heightF-hl+m	-0.94383913	0.27568367
## Age-0+1:roundingF+nr-r	-0.42439316	1.07684097
## heightF+h-l:roundingF+nr-r	-0.59954026	0.53233664
## heightF-hl+m:roundingF+nr-r	-0.32480989	2.25627647
## Age-0+1:frtnsF-b+f	-2.59834227	0.37584107
## Age-0+1:frtnsF-bf+m	-2.55930105	0.54622976
## Age-0+1:rhoticityF+nx-x	-0.30671872	3.37897912
## Age-0+1:Sibilant-Ç+s	-0.24618551	1.64000452
## heightF+h-l:Sibilant-Ç+s	-0.92345696	-0.02085844
## heightF-hl+m:Sibilant-Ç+s	-1.63087852	0.24573043
## roundingF+nr-r:Sibilant-Ç+s	-0.25163796	0.89089903
## frtnsF-b+f:Sibilant-Ç+s	-2.50319119	0.56889183
## Age-0+1:heightF+h-l:roundingF+nr-r	-0.31456668	1.94706674
## Age-0+1:heightF-hl+m:roundingF+nr-r	-0.54087431	4.10502848
## Age-0+1:heightF+h-l:Sibilant-Ç+s	-0.70122537	0.92313484
## Age-0+1:heightF-hl+m:Sibilant-Ç+s	-2.58994201	0.84864633
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s	-0.77213916	1.33042939
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s	-4.36714047	1.30389431

```
modelStdevErb <- lmer(StdevErb ~ Age*heightF*roundingF*frtnsF*rhoticityF*Sibilant +
(Sibilant | Subject) + (Sibilant|Index), data = all, REML = TRUE)
```

```
## fixed-effect model matrix is rank deficient so dropping 116 columns / coefficients
```

```
## Warning in checkConv(attr("opt", "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0214093 (tol = 0.002, component 1)
```

```
summary(modelStdevErb)
```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: StdevErb ~ Age * heightF * roundingF * frtnsF * rhoticityF *
##       Sibilant + (Sibilant | Subject) + (Sibilant | Index)
## Data: all
##
## REML criterion at convergence: 967.9
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -3.4418 -0.6159 -0.0852  0.6341  3.5529
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   Index    (Intercept)          0.00843  0.09182
##           Sibilant-Ç+s        0.21554  0.46427 -0.95
##   Subject  (Intercept)          0.04515  0.21249
##           Sibilant-Ç+s        0.50487  0.71054 -0.66
## Residual                    0.27151  0.52107
## Number of obs: 551, groups:  Index, 35; Subject, 16
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      3.889609   0.338173  11.502
## Age-0+1          -0.378568   0.335423  -1.129
## heightF+h-1      0.006518   0.192221   0.034
## heightF-hl+m    -0.058036   0.250735  -0.231
## roundingF+nr-r  -0.479657   0.281104  -1.706
## frtnsF-b+f      -0.132032   0.578145  -0.228
## frtnsF-bf+m     0.108689   0.553248   0.196
## rhoticityF+nx-x -0.817053   0.743353  -1.099
## Sibilant-Ç+s    -1.109329   0.317944  -3.489
## Age-0+1:heightF+h-1 0.256755   0.213664   1.202
## Age-0+1:heightF-hl+m 0.167635   0.231163   0.725
## Age-0+1:roundingF+nr-r 0.059524   0.284334   0.209
## heightF+h-1:roundingF+nr-r 0.011528   0.306441   0.038
## heightF-hl+m:roundingF+nr-r 0.050227   0.932427   0.054
## Age-0+1:frtnsF-b+f -0.257409   0.563487  -0.457
## Age-0+1:frtnsF-bf+m -0.306065   0.588476  -0.520
## Age-0+1:rhoticityF+nx-x -0.200029   0.698151  -0.287
## Age-0+1:Sibilant-Ç+s 0.653116   0.444364   1.470
## heightF+h-1:Sibilant-Ç+s 0.329742   0.326793   1.009
## heightF-hl+m:Sibilant-Ç+s -0.078569   0.659438  -0.119
## roundingF+nr-r:Sibilant-Ç+s 0.528454   0.381269   1.386
## frtnsF-b+f:Sibilant-Ç+s -0.309544   1.051600  -0.294
## Age-0+1:heightF+h-1:roundingF+nr-r -0.312639   0.428620  -0.729
## Age-0+1:heightF-hl+m:roundingF+nr-r -0.263497   0.880585  -0.299
## Age-0+1:heightF+h-1:Sibilant-Ç+s -0.076150   0.307931  -0.247
## Age-0+1:heightF-hl+m:Sibilant-Ç+s -0.449609   0.651971  -0.690
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s -0.155608   0.398662  -0.390
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s -0.972293   1.075209  -0.904

```

```

##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
##       vcov(x)           if you need it

```

```
## Warning in confint.thpr(pp, level = level, zeta = zeta): bad spline fit
## for .sig03: falling back to linear interpolation
```

```
##           2.5 %       97.5 %
## .sig01      0.01223826  0.26139626
## .sig02     -1.00000000  0.28504655
## .sig03      0.02408719  0.49838851
## .sig04      0.12689935  0.31246575
## .sig05     -0.90113467 -0.20065906
## .sig06      0.47097988  1.00654482
## .sigma      0.48373156  0.54879122
## (Intercept) 3.35713137  4.46662723
## Age-0+1    -1.02665844  0.26760343
## heightF+h-l -0.28973442  0.35476785
## heightF-hl+m -0.45715913  0.37171487
## roundingF+nr-r -0.96841993 -0.04272955
## frtnsF-b+f   -1.03330313  0.87979694
## frtnsF-bf+m  -0.75292453  1.05822067
## rhoticityF+nx-x -2.09440292  0.35144218
## Sibilant-Ç+s -1.65784367 -0.57556537
## Age-0+1:heightF+h-l -0.15748311  0.67122704
## Age-0+1:heightF-hl+m -0.28131092  0.61528234
## Age-0+1:roundingF+nr-r -0.49382231  0.61020694
## heightF+h-l:roundingF+nr-r -0.54907473  0.50450010
## heightF-hl+m:roundingF+nr-r -1.57546140  1.51006401
## Age-0+1:frtnsF-b+f -1.34980665  0.83568524
## Age-0+1:frtnsF-bf+m -1.45028293  0.83367627
## Age-0+1:rhoticityF+nx-x -1.55568188  1.15395890
## Age-0+1:Sibilant-Ç+s -0.20698480  1.50655547
## heightF+h-l:Sibilant-Ç+s -0.20600328  0.86382762
## heightF-hl+m:Sibilant-Ç+s -1.11134819  1.05424718
## roundingF+nr-r:Sibilant-Ç+s -0.17666393  1.11944491
## frtnsF-b+f:Sibilant-Ç+s -1.94871244  1.53279306
## Age-0+1:heightF+h-l:roundingF+nr-r -1.14153487  0.52096127
## Age-0+1:heightF-hl+m:roundingF+nr-r -1.96757143  1.44853682
## Age-0+1:heightF+h-l:Sibilant-Ç+s -0.67121171  0.52314950
## Age-0+1:heightF-hl+m:Sibilant-Ç+s -1.71941801  0.81011290
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s -0.92585908  0.62047228
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s -3.06513635  1.10609679
```

```
modelCoGHz <- lmer(CoG ~ Age*heightF*roundingF*rhoticityF*Sibilant + (Sibilant | Subject) + (Sibilant|Index), data = all, REML = TRUE)
```

```
## fixed-effect model matrix is rank deficient so dropping 24 columns / coefficients
```

```
## Warning in checkConv(attr("opt", "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00285879 (tol = 0.002, component 1)
```

```
summary(modelCoGHz)
```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: CoG ~ Age * heightF * roundingF * rhoticityF * Sibilant + (Sibilant |
##   Subject) + (Sibilant | Index)
##   Data: all
##
## REML criterion at convergence: 8002.3
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -3.3049 -0.5497 -0.0395  0.5762  3.4880
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   Index    (Intercept)           6890     83.01
##           Sibilant- $\zeta$ +s 42632    206.48  0.85
##   Subject  (Intercept)          270077   519.69
##           Sibilant- $\zeta$ +s 110648    332.64 -0.36
## Residual                168076    409.97
## Number of obs: 551, groups:  Index, 35; Subject, 16
##
## Fixed effects:
##
##                                     Estimate Std. Error t value
## (Intercept)                        4873.76    156.13  31.216
## Age-0+1                             -60.13    291.99  -0.206
## heightF+h-l                          173.95    176.50   0.986
## heightF-hl+m                          78.27    126.25   0.620
## roundingF+nr-r                        10.13    119.97   0.084
## rhoticityF+nx-x                       107.41    124.43   0.863
## Sibilant- $\zeta$ +s                    1546.18    154.91   9.981
## Age-0+1:heightF+h-l                    141.00    204.34   0.690
## Age-0+1:heightF-hl+m                     94.29    141.16   0.668
## Age-0+1:roundingF+nr-r                   14.16    145.33   0.097
## heightF+h-l:roundingF+nr-r              -451.49    353.11  -1.279
## heightF-hl+m:roundingF+nr-r             -378.33    275.86  -1.371
## Age-0+1:rhoticityF+nx-x                  168.42    219.43   0.768
## Age-0+1:Sibilant- $\zeta$ +s               391.03    226.94   1.723
## heightF+h-l:Sibilant- $\zeta$ +s           -452.65    334.47  -1.353
## heightF-hl+m:Sibilant- $\zeta$ +s          -74.54    129.86  -0.574
## roundingF+nr-r:Sibilant- $\zeta$ +s       -189.23    240.07  -0.788
## Age-0+1:heightF+h-l:roundingF+nr-r      -50.57    409.02  -0.124
## Age-0+1:heightF-hl+m:roundingF+nr-r     -153.58    281.96  -0.545
## Age-0+1:heightF+h-l:Sibilant- $\zeta$ +s    -428.07    409.47  -1.045
## Age-0+1:heightF-hl+m:Sibilant- $\zeta$ +s  -104.50    162.15  -0.644
## Age-0+1:roundingF+nr-r:Sibilant- $\zeta$ +s -213.13    280.03  -0.761
## heightF+h-l:roundingF+nr-r:Sibilant- $\zeta$ +s  368.38    668.23   0.551
## Age-0+1:heightF+h-l:roundingF+nr-r:Sibilant- $\zeta$ +s  948.48    816.45   1.162

```

```

##
## Correlation matrix not shown by default, as p = 24 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)           if you need it

```

```
## fit warnings:  
## fixed-effect model matrix is rank deficient so dropping 24 columns / coefficients  
## convergence code: 0  
## Model failed to converge with max|grad| = 0.00285879 (tol = 0.002, component 1)
```

```
modelStdevHz <- lmer(Stdev ~ Age*heightF*frtnsF*roundingF*rhoticityF*Sibilant + (Sibilant | Subject) + (Sibilant|Index), data = all, REML = TRUE, na.action = "na.omit")
```

```
## fixed-effect model matrix is rank deficient so dropping 116 columns / coefficients
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.00404318 (tol = 0.002, component 1)
```

```
summary(modelStdevHz)
```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: Stdev ~ Age * heightF * frtnsF * roundingF * rhoticityF * Sibilant +
## (Sibilant | Subject) + (Sibilant | Index)
## Data: all
##
## REML criterion at convergence: 7357.5
##
## Scaled residuals:
## Min 1Q Median 3Q Max
## -2.9558 -0.5757 -0.0277 0.6168 4.6154
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## Index (Intercept) 3016 54.92
## Sibilant-Ç+s 4314 65.68 -0.71
## Subject (Intercept) 43174 207.78
## Sibilant-Ç+s 54112 232.62 -0.11
## Residual 54926 234.36
## Number of obs: 551, groups: Index, 35; Subject, 16
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 2046.37 88.97 23.000
## Age-0+1 -245.58 141.14 -1.740
## heightF+h-1 300.54 268.06 1.121
## heightF-hl+m 132.94 112.49 1.182
## frtnsF-b+f 40.60 150.26 0.270
## frtnsF-bf+m -125.58 61.32 -2.048
## roundingF+nr-r -215.50 62.82 -3.430
## rhoticityF+nx-x -215.26 165.96 -1.297
## Sibilant-Ç+s 148.77 102.09 1.457
## Age-0+1:heightF+h-1 299.16 359.67 0.832
## Age-0+1:heightF-hl+m 132.18 162.83 0.812
## Age-0+1:frtnsF-b+f -76.81 204.30 -0.376
## Age-0+1:frtnsF-bf+m -207.38 101.69 -2.039
## heightF+h-1:frtnsF-b+f -564.20 580.99 -0.971
## Age-0+1:roundingF+nr-r -76.25 89.92 -0.848
## heightF+h-1:roundingF+nr-r -41.51 113.90 -0.364
## Age-0+1:rhoticityF+nx-x 166.80 209.20 0.797
## Age-0+1:Sibilant-Ç+s 389.80 167.23 2.331
## heightF+h-1:Sibilant-Ç+s 29.28 101.26 0.289
## heightF-hl+m:Sibilant-Ç+s -196.43 208.78 -0.941
## frtnsF-b+f:Sibilant-Ç+s -417.24 337.63 -1.236
## roundingF+nr-r:Sibilant-Ç+s 178.15 123.95 1.437
## Age-0+1:heightF+h-1:frtnsF-b+f -394.91 792.34 -0.498
## Age-0+1:heightF+h-1:roundingF+nr-r -29.57 192.79 -0.153
## Age-0+1:heightF+h-1:Sibilant-Ç+s 116.51 138.49 0.841
## Age-0+1:heightF-hl+m:Sibilant-Ç+s -421.41 293.33 -1.437
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s -662.53 483.69 -1.370
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s -64.56 179.30 -0.360
##
##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

```

##	2.5 %	97.5 %
## .sig01	0.0000000	65.9510529
## .sig02	-1.0000000	1.0000000
## .sig03	0.0000000	131.9020758
## .sig04	139.6176769	290.3354402
## .sig05	-0.5523394	0.3874256
## .sig06	148.9611129	332.3548121
## .sigma	217.5298606	246.0760567
## (Intercept)	1894.9471161	2196.1945465
## Age-0+1	-511.4539258	19.9143918
## heightF+h-l	-144.3038269	719.5988157
## heightF-hl+m	-54.8226516	307.6330711
## frtnsF-b+f	-208.5586949	277.1887727
## frtnsF-bf+m	-220.2300549	-23.4770969
## roundingF+nr-r	-317.7577047	-113.7702679
## rhoticityF+nx-x	-478.1404893	59.8139685
## Sibilant-Ç+s	-25.8143634	318.8241547
## Age-0+1:heightF+h-l	-394.9209271	997.4679767
## Age-0+1:heightF-hl+m	-181.7801086	448.5612286
## Age-0+1:frtnsF-b+f	-470.7945317	320.1173321
## Age-0+1:frtnsF-bf+m	-405.4814781	-11.8249032
## heightF+h-l:frtnsF-b+f	-1461.1966603	410.7601344
## Age-0+1:roundingF+nr-r	-250.5042324	97.5936362
## heightF+h-l:roundingF+nr-r	-241.3220814	131.8037116
## Age-0+1:rhoticityF+nx-x	-238.7685533	571.1353774
## Age-0+1:Sibilant-Ç+s	75.4910200	705.5187786
## heightF+h-l:Sibilant-Ç+s	-134.4253388	192.9964789
## heightF-hl+m:Sibilant-Ç+s	-519.2342870	152.2730036
## frtnsF-b+f:Sibilant-Ç+s	-937.7313494	154.7954548
## roundingF+nr-r:Sibilant-Ç+s	-38.0232422	369.0190778
## Age-0+1:heightF+h-l:frtnsF-b+f	-1933.0655823	1134.3301029
## Age-0+1:heightF+h-l:roundingF+nr-r	-403.0003662	343.4134050
## Age-0+1:heightF+h-l:Sibilant-Ç+s	-151.5504892	384.6049086
## Age-0+1:heightF-hl+m:Sibilant-Ç+s	-991.3621139	144.1249306
## Age-0+1:frtnsF-b+f:Sibilant-Ç+s	-1601.9233725	270.5412084
## Age-0+1:roundingF+nr-r:Sibilant-Ç+s	-410.9762384	283.1054229

```
shapiro.test(residuals(modelCoGHZ))
```

```
##
## Shapiro-Wilk normality test
##
## data: residuals(modelCoGHZ)
## W = 0.9852, p-value = 2.186e-05
```

```
shapiro.test(residuals(modelStdevHz))
```

```
##
## Shapiro-Wilk normality test
##
## data: residuals(modelStdevHz)
## W = 0.98925, p-value = 0.000459
```

```
shapiro.test(residuals(modelCoGErb))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(modelCoGErb)  
## W = 0.99539, p-value = 0.1023
```

```
shapiro.test(residuals(modelStdevErb))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(modelStdevErb)  
## W = 0.99025, p-value = 0.001034
```

```
#calculating the mean CoG of /s/ in speaker 8  
library(dplyr)  
CoG8s <- fulltable%>%  
  filter(Subject == "8", Sibilant == "s")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoG8s$CoG)
```

```
## [1] 4607.294
```

```
#calculating the mean CoG of /ç/ in speaker 8  
library(dplyr)  
CoG8ç <- fulltable%>%  
  filter(Subject == "8", Sibilant == "ç")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoG8ç$CoG)
```

```
## [1] 3683.696
```

```
#calculating the mean CoG of /s/ in speaker 13  
library(dplyr)  
CoG13s <- fulltable%>%  
  filter(Subject == "13", Sibilant == "s")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoG13s$CoG)
```

```
## [1] 5041.943
```



```
#calculating the mean CoG of /ʒ/ in speaker 13  
library(dplyr)  
CoG13ʒ <- fulltable%>%  
  filter(Subject == "13", Sibilant == "ʒ")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoG13ʒ$CoG)
```

```
## [1] 3292.652
```

```
#calculating the mean of CoG of /s/ in younger speakers  
library(dplyr)  
CoGYS<- fulltable %>%  
  filter(Age == "1", Sibilant == "s")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoGYS$CoG)
```

```
## [1] 5454.341
```

```
#calculating the mean of CoG of /ʒ/ in younger speakers  
library(dplyr)  
CoGYʒ<- fulltable %>%  
  filter(Age == "1", Sibilant == "ʒ")%>%  
  select(CoG)%>%  
  mutate(CoG = as.numeric(CoG))  
mean(CoGYʒ$CoG)
```

```
## [1] 4107.141
```

F R Script for Classification Trees and Random Forests

TreesAndForests

Xinyu Zhang

6/29/2020

```
fulltable <- read.csv("NewResult.csv")
#head(fulltable)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
young <- fulltable %>%
  filter(Age == 1) %>%
  dplyr::select(-Subject, -Age, -Index, -IPA, -CoG, -Stdev, -Position, -Vp, -Vf, -CoG
Erb, -StdevErb)
head(young)
```

```

## Sibilant bin1 bin2 bin3 bin4 bin5 bin6
## 1 s -9.052738 -8.556997 -4.870987 -4.218492 -4.850850 -6.4336164
## 2 s -2.948580 -4.879446 -2.060097 -2.011352 2.026905 -1.1402887
## 3 s -6.117940 -7.691914 -3.942539 -2.621272 -0.913079 6.7347466
## 4 s -7.470004 -6.927923 -2.576884 -8.418656 -8.956720 -4.4691433
## 5 s -3.647721 -7.621700 -4.805937 -9.255555 -1.093329 0.9828912
## 6 s -5.793275 -5.878609 -1.390356 -4.743696 -5.864162 10.2314562
## bin7 bin8 bin9 bin10 bin11 bin12 bin13
## 1 -9.370378 -7.752136 -8.965112 0.4495370 3.8068596 11.1640723 16.0745454
## 2 -4.632750 -5.243153 -1.184720 1.5086332 2.6536686 3.3904228 7.8810884
## 3 -3.460427 -5.391177 -0.934101 4.3435204 0.9056229 3.5931128 10.1739540
## 4 -7.916696 -6.002341 -8.196552 -0.8540085 -1.5642317 -5.8005777 0.6155142
## 5 -6.337333 -7.819851 -5.640433 -1.6501971 -2.7620005 0.1089989 4.4720045
## 6 1.489044 -1.819993 -4.039301 6.2573727 5.6231383 6.8641713 9.4173490
## bin14 bin15 bin16 bin17 bin18 bin19 bin20
## 1 16.547785 14.194455 13.132448 11.870301 11.218071 12.07828 9.152985
## 2 13.128567 15.637285 13.795315 20.076671 16.624866 16.03775 14.326997
## 3 14.250188 17.706188 13.647122 13.380362 11.903631 12.47035 8.167565
## 4 2.981070 8.416862 2.171053 2.837635 6.971331 10.02746 9.496309
## 5 7.862442 10.683707 13.786008 20.300074 23.024090 17.87379 13.381629
## 6 11.038473 14.321497 15.186060 17.908704 22.731755 20.47064 19.450198
## bin21 bin22 bin23 bin24 bin25 bin26 bin27
## 1 12.896152 6.998751 8.270279 6.748858 7.035672 8.845540 11.705263
## 2 14.517203 12.050778 9.508779 9.654701 11.066853 10.855544 10.907322
## 3 8.935988 7.046180 6.754974 7.694643 6.787817 7.959011 5.268743
## 4 10.276865 8.726646 10.692209 12.919726 8.616948 7.775069 9.993443
## 5 16.871098 13.175957 12.240556 12.854929 9.639143 8.937537 9.873585
## 6 20.236146 14.840057 17.694142 16.608715 14.276433 14.588616 13.873548
## bin28 bin29 bin30 bin31 bin32 bin33 bin34
## 1 13.099344 5.840506 6.105147 10.665144 13.831340 19.000859 16.323290
## 2 6.748886 9.402706 10.395039 8.402617 9.725059 13.127745 12.165429
## 3 5.795232 2.518056 6.897152 10.219491 9.305398 11.885174 10.357887
## 4 8.809584 9.509015 6.250082 7.822943 6.502482 5.431262 2.821759
## 5 11.527875 11.912453 11.725899 12.562875 9.106247 12.484106 9.748338
## 6 11.423138 13.023069 12.058883 14.845923 12.886212 13.515536 13.205338
## bin35 bin36 bin37 bin38
## 1 12.727929 13.035199 7.8862157 8.5584810
## 2 9.993567 9.396044 7.8533025 6.9327306
## 3 9.125429 7.752784 7.4645417 3.2338963
## 4 -3.354909 -1.112687 -1.6734236 -0.4542861
## 5 6.150656 5.909042 0.4569055 2.6961351
## 6 11.319922 8.375685 7.4774573 2.8314308

```

```

old <- fulltable %>%
  filter(Age == 0) %>%
  dplyr::select(-Subject, -Age, -Index, -IPA, -CoG, -Stdev, -Position, -Vp, -Vf, -CoG
Erb, -StdevErb)
head(old)

```

```

## Sibilant bin1 bin2 bin3 bin4 bin5 bin6
## 1 s 0.03079822 5.368053 1.3356620 3.8024194 7.638946 8.518428
## 2 s -1.07993498 4.732777 4.1342010 3.8134596 9.258956 9.534831
## 3 s -1.89311451 2.843402 -1.0546411 0.8310224 10.217786 8.706969
## 4 s -2.62480302 3.399465 -3.0490131 -1.1306421 4.636302 5.944460
## 5 s -1.76925924 2.587684 -0.4604427 1.8163580 6.344636 7.024475
## 6 Ć -4.35499190 -4.189306 -2.5512102 3.1702479 9.836333 19.194536
## bin7 bin8 bin9 bin10 bin11 bin12 bin13
## 1 3.9305149 7.1123151 15.230600 17.545210 19.596762 22.25446 22.374452
## 2 2.7117942 6.1003611 8.216539 11.317408 10.496400 21.12455 25.931900
## 3 1.3153554 0.7429933 8.701973 9.501857 18.532883 26.15295 28.422889
## 4 0.6982653 0.2469375 7.000124 8.336731 11.804192 19.63063 25.143991
## 5 3.1415236 4.8106590 7.773033 10.197163 9.174671 16.51992 19.929402
## 6 16.6411375 15.8706811 11.708023 11.532916 7.521156 5.11676 6.078382
## bin14 bin15 bin16 bin17 bin18 bin19 bin20
## 1 16.807175 16.849622 19.824146 15.705370 15.731594 17.366291 15.171438
## 2 23.925333 21.541203 22.652723 23.515866 22.685180 22.588432 17.838364
## 3 19.400775 19.631410 21.793965 17.189007 16.072513 17.223698 12.740024
## 4 23.287909 22.252068 23.965798 22.861844 22.989191 19.770836 17.148074
## 5 19.566929 18.510509 14.033934 19.170929 17.259525 17.562984 12.246112
## 6 1.008853 7.788682 8.780875 9.029114 5.842555 8.395622 9.309483
## bin21 bin22 bin23 bin24 bin25 bin26 bin27 bin28
## 1 22.50508 22.26885 18.184247 24.981560 25.13370 16.813799 20.099184 21.0763254
## 2 21.81495 24.60651 20.602070 22.524704 23.01943 16.198076 16.844628 18.6271211
## 3 20.13430 20.47248 16.252307 16.363761 16.04937 8.279066 11.772129 14.5855535
## 4 18.70550 19.35043 17.030038 16.664481 18.57928 12.715267 12.068423 17.0858940
## 5 15.72888 16.38258 12.786221 11.935313 14.18757 9.113775 11.132812 14.3556308
## 6 17.94750 15.25452 9.562816 9.191713 6.80332 -1.594965 2.154858 0.2669866
## bin29 bin30 bin31 bin32 bin33 bin34 bin35
## 1 17.869202 14.208469 8.541668 10.113432 11.542438 8.343518 9.655772
## 2 17.947814 21.079367 15.129374 16.850892 13.509220 10.642141 10.924920
## 3 11.682313 7.478382 7.020007 5.375419 1.836066 3.119932 5.005802
## 4 15.263576 16.560007 12.458058 10.117296 12.824886 8.218765 8.858684
## 5 12.349997 14.109316 12.141184 6.679127 10.883913 4.958634 3.895202
## 6 -4.606697 -7.917246 -6.766860 -4.830006 -4.852801 -7.000025 -6.344956
## bin36 bin37 bin38
## 1 11.739608 10.6788560 7.6120026
## 2 11.276608 10.1731825 8.1175219
## 3 3.533633 0.9776531 4.5100433
## 4 9.822343 7.9448941 6.3261298
## 5 5.768801 5.0784504 2.3530547
## 6 -1.665697 -3.6907411 -0.1649988

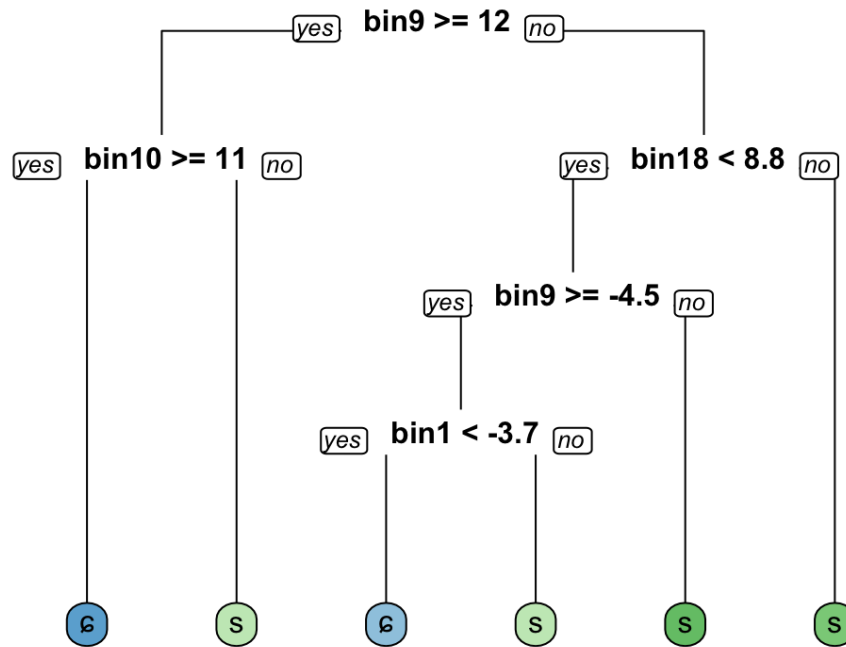
```

single tree

```

library(rpart)
sibtreeY <- rpart(formula = Sibilant ~ .,
                  data = young,
                  method = "class")
library(rpart.plot)
rpart.plot(x = sibtreeY, yesno = 2, type = 0, extra = 0)

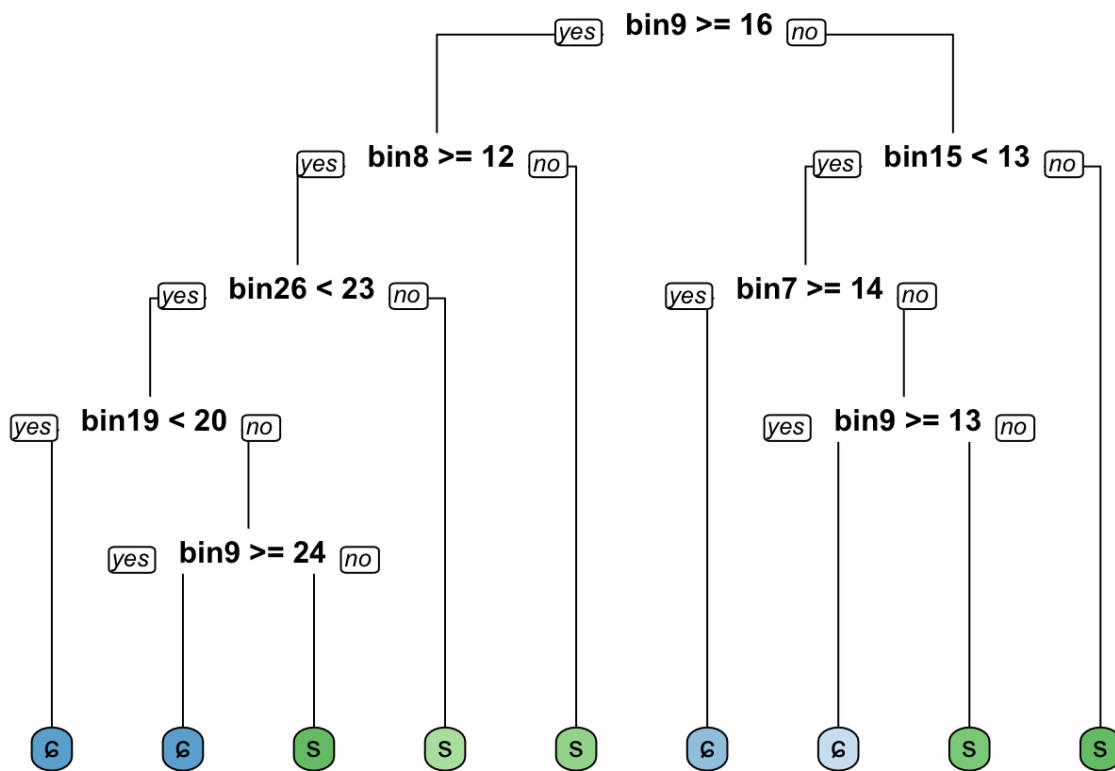
```



```

sibtree0 <- rpart(formula = Sibillant ~ .,
                  data = old,
                  method = "class")
rpart.plot(x = sibtree0, yesno = 2, type = 0, extra = 0)

```



Actual training of single trees:

Partitioning:

```

yn <- nrow(young)
yn_train = round(0.8*yn)
#set seed for reproducibility
set.seed(1830)
y_train_indices <- sample(1:yn, yn_train)
y_train <- young[y_train_indices,]
y_test <- young[-y_train_indices,]

#same for the other age group
on <- nrow(old)
on_train = round(0.8*on)
#set seed for reproducibility
set.seed(1834)
#split
o_train_indices <- sample(1:on, on_train)
o_train <- old[o_train_indices,]
o_test <- old[-o_train_indices,]
  
```

The models:

```

y_tree <- rpart(formula = Sibilant ~ .,
                data = y_train,
                method = "class")
y_tree
  
```

```
## n= 377
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 377 152 s (0.40318302 0.59681698)
##   2) bin9>=12.1626 96   8 c (0.91666667 0.08333333)
##     4) bin10>=10.70791 89   3 c (0.96629213 0.03370787) *
##     5) bin10< 10.70791 7    2 s (0.28571429 0.71428571) *
##   3) bin9< 12.1626 281  64 s (0.22775801 0.77224199)
##     6) bin17< 8.561642 111  54 s (0.48648649 0.51351351)
##       12) bin8>=-3.876849 57   4 c (0.92982456 0.07017544) *
##       13) bin8< -3.876849 54   1 s (0.01851852 0.98148148) *
##     7) bin17>=8.561642 170  10 s (0.05882353 0.94117647) *
```

```
o_tree <- rpart(formula = Sibilant ~ .,
                data = o_train,
                method = "class")
o_tree
```

```
## n= 356
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 356 141 s (0.39606742 0.60393258)
##   2) bin9>=16.21045 125  20 c (0.84000000 0.16000000)
##     4) bin7>=11.45434 108   9 c (0.91666667 0.08333333) *
##     5) bin7< 11.45434 17    6 s (0.35294118 0.64705882) *
##   3) bin9< 16.21045 231  36 s (0.15584416 0.84415584)
##     6) bin15< 12.81072 86   31 s (0.36046512 0.63953488)
##       12) bin7>=7.709157 38  12 c (0.68421053 0.31578947)
##         24) bin9>=11.88803 17   0 c (1.00000000 0.00000000) *
##         25) bin9< 11.88803 21   9 s (0.42857143 0.57142857)
##           50) bin15< 7.032271 7   0 c (1.00000000 0.00000000) *
##           51) bin15>=7.032271 14   2 s (0.14285714 0.85714286) *
##       13) bin7< 7.709157 48   5 s (0.10416667 0.89583333) *
##     7) bin15>=12.81072 145   5 s (0.03448276 0.96551724) *
```

Evaluating the trees by computing a confusion matrix:

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```



```

#generating thee predicted classes:
class_predictionY <- predict(object = y_tree,
                             newdata = y_test,
                             type = "class")
class_prediction0 <- predict(object = o_tree,
                             newdata = o_test,
                             type = "class")
#calculating the confusion matrix for the test set
confusionMatrix(data = class_predictionY,
                 reference = y_test$Sibilant)

```

```

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  C   s
##           C 33  2
##           s  7 52
##
##           Accuracy : 0.9043
##           95% CI : (0.826, 0.9553)
##           No Information Rate : 0.5745
##           P-Value [Acc > NIR] : 1.948e-12
##
##           Kappa : 0.8009
##
## Mcnemar's Test P-Value : 0.1824
##
##           Sensitivity : 0.8250
##           Specificity : 0.9630
##           Pos Pred Value : 0.9429
##           Neg Pred Value : 0.8814
##           Prevalence : 0.4255
##           Detection Rate : 0.3511
##           Detection Prevalence : 0.3723
##           Balanced Accuracy : 0.8940
##
##           'Positive' Class : C
##

```

```

confusionMatrix(data = class_prediction0,
                 reference = o_test$Sibilant)

```

```

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  C   s
##           C 36  4
##           s  7 42
##
##           Accuracy : 0.8764
##           95% CI : (0.7896, 0.9367)
##           No Information Rate : 0.5169
##           P-Value [Acc > NIR] : 6.293e-13
##
##           Kappa : 0.752
##
##           Mcnemar's Test P-Value : 0.5465
##
##           Sensitivity : 0.8372
##           Specificity : 0.9130
##           Pos Pred Value : 0.9000
##           Neg Pred Value : 0.8571
##           Prevalence : 0.4831
##           Detection Rate : 0.4045
##           Detection Prevalence : 0.4494
##           Balanced Accuracy : 0.8751
##
##           'Positive' Class : C
##

```

Bagging:

```

library(ipred)
#setting seed for reproducibility
set.seed(1943)
y_bag <- bagging(formula = Sibilant ~ .,
                 data = y_train,
                 coob = TRUE)
y_bag

```

```

##
## Bagging classification trees with 25 bootstrap replications
##
## Call: bagging.data.frame(formula = Sibilant ~ ., data = y_train, coob = TRUE)
##
## Out-of-bag estimate of misclassification error: 0.0477

```

```

o_bag <- bagging(formula = Sibilant ~ .,
                 data = o_train,
                 coob = TRUE)
o_bag

```

```
##
## Bagging classification trees with 25 bootstrap replications
##
## Call: bagging.data.frame(formula = Sibilant ~ ., data = o_train, coob = TRUE)
##
## Out-of-bag estimate of misclassification error: 0.1067
```

Evaluating the bagged trees:

```
library(caret)
#generating thee predicted classes:
class_predictionYbag <- predict(object = y_bag,
                               newdata = y_test,
                               type = "class")
class_predictionObag <- predict(object = o_bag,
                               newdata = o_test,
                               type = "class")
#calculating the confusion matrix for the test set
confusionMatrix(data = class_predictionYbag,
                 reference = y_test$Sibilant)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  C  s
##           C 38  2
##           s  2 52
##
##           Accuracy : 0.9574
##           95% CI : (0.8946, 0.9883)
##           No Information Rate : 0.5745
##           P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.913
##
## Mcnemar's Test P-Value : 1
##
##           Sensitivity : 0.9500
##           Specificity : 0.9630
##           Pos Pred Value : 0.9500
##           Neg Pred Value : 0.9630
##           Prevalence : 0.4255
##           Detection Rate : 0.4043
##           Detection Prevalence : 0.4255
##           Balanced Accuracy : 0.9565
##
##           'Positive' Class : C
##
```

```
confusionMatrix(data = class_predictionObag,
                 reference = o_test$Sibilant)
```

```

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  c  s
##           c 41  3
##           s  2 43
##
##           Accuracy : 0.9438
##           95% CI : (0.8737, 0.9815)
##           No Information Rate : 0.5169
##           P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.8876
##
##           Mcnemar's Test P-Value : 1
##
##           Sensitivity : 0.9535
##           Specificity : 0.9348
##           Pos Pred Value : 0.9318
##           Neg Pred Value : 0.9556
##           Prevalence : 0.4831
##           Detection Rate : 0.4607
##           Detection Prevalence : 0.4944
##           Balanced Accuracy : 0.9441
##
##           'Positive' Class : c
##

```

^improved upon the single trees.

Calculating the AUC of the bagged trees:

```

probpredY <- predict(object = y_bag,
                    newdata = y_test,
                    type = "prob")
probpred0 <- predict(object = o_bag,
                    newdata = o_test,
                    type = "prob")
library(Metrics)

```

```

##
## Attaching package: 'Metrics'

```

```

## The following objects are masked from 'package:caret':
##
##   precision, recall

```

```

auc(actual = ifelse(y_test$Sibilant == "s", 1, 0),
     predicted = probpredY[,"s"])

```

```

## [1] 0.9928241

```

```
auc(actual = ifelse(o_test$Sibilant == "s", 1, 0),
     predicted = probpred0[, "s"])
```

```
## [1] 0.983822
```

^ very close to 1

k-fold cross-validation to reduce variance of estimates:

```
#specifying the training configuration
ctrl <- trainControl(method = "cv", #cross validation
                    number = 5, #number of k in "k-fold"
                    classProbs = TRUE,
                    summaryFunction = twoClassSummary)#calculating the ROC curve for
AUC
#training configuration
#reproducibility
set.seed(2009)
ytreebag <- train(Sibilant ~ .,
                 data = y_train,
                 method = "treebag",
                 metric = "ROC", #for AUC
                 trControl = ctrl)

ytreebag
```

```
## Bagged CART
##
## 377 samples
## 38 predictor
## 2 classes: 'c', 's'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 301, 302, 301, 302, 302
## Resampling results:
##
## ROC Sens Spec
## 0.9941099 0.9546237 0.9911111
```

```
otreebag <- train(Sibilant ~ .,
                 data = o_train,
                 method = "treebag",
                 metric = "ROC",
                 trControl = ctrl)

otreebag
```

```
## Bagged CART
##
## 356 samples
## 38 predictor
## 2 classes: 'c', 's'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 284, 285, 285, 285, 285
## Resampling results:
##
## ROC          Sens          Spec
## 0.9551581 0.8431034 0.9488372
```

```
#generating predictions
predbagY <- predict(object = ytreebag,
                    newdata = y_test,
                    type = "prob")
predbagO <- predict(object = otreebag,
                    newdata = o_test,
                    type = "prob")

#computing the AUC
auc(actual = ifelse(y_test$Sibilant == "s", 1, 0),
     predicted = predbagY[, "s"]) #young
```

```
## [1] 0.9923611
```

```
auc(actual = ifelse(o_test$Sibilant == "s", 1, 0),
     predicted = predbagO[, "s"]) #old
```

```
## [1] 0.9843276
```

Random Forest:

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
## margin
```

```
## The following object is masked from 'package:dplyr':
##
## combine
```

```
set.seed(2029)
youngForest <- randomForest(formula = Sibilant ~ .,
                             data = y_train) #sqrt(38)=6.16, default mtry = 6
youngForest
```

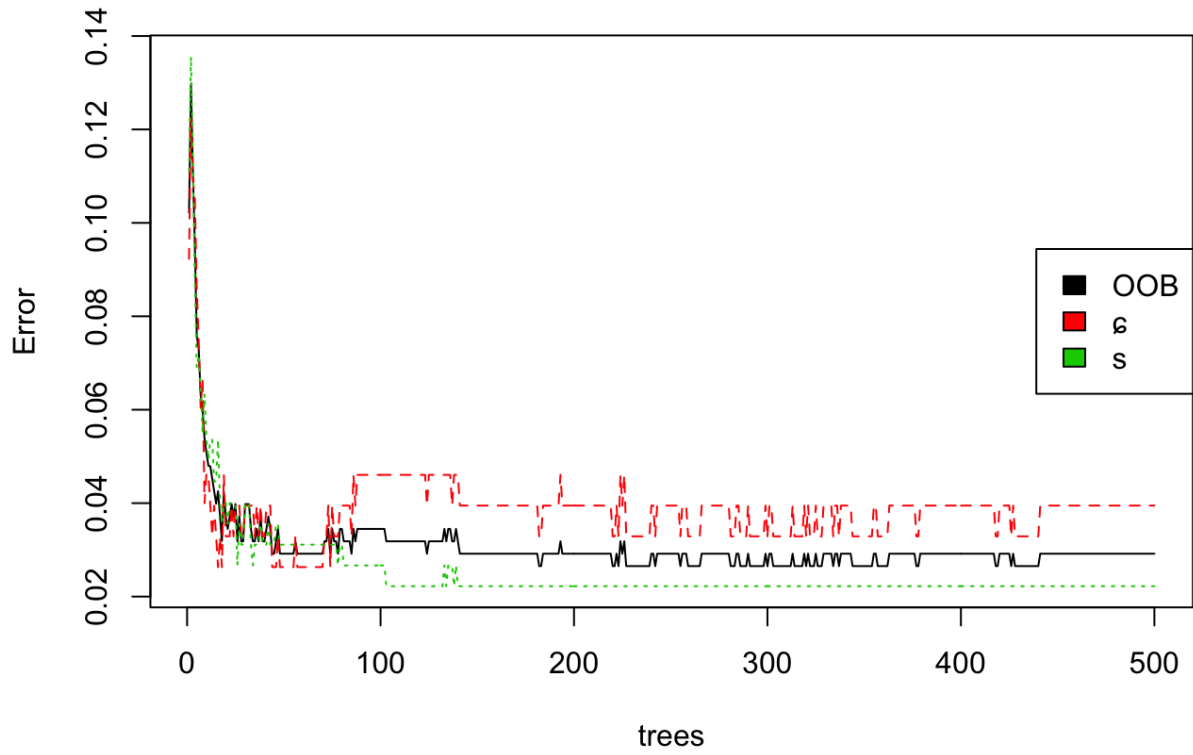
```
##
## Call:
## randomForest(formula = Sibilant ~ ., data = y_train)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 6
##
##           OOB estimate of error rate: 2.92%
## Confusion matrix:
##      c  s class.error
## c 146  6  0.03947368
## s   5 220  0.02222222
```

```
oldForest <- randomForest(formula = Sibilant ~ .,
                           data = o_train)
oldForest
```

```
##
## Call:
## randomForest(formula = Sibilant ~ ., data = o_train)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 6
##
##           OOB estimate of error rate: 5.9%
## Confusion matrix:
##      c  s class.error
## c 126 15  0.10638298
## s   6 209  0.02790698
```

```
errY <- youngForest$err.rate
errO <- oldForest$err.rate
plot(youngForest)
legend(x = "right",
       legend = colnames(errY),
       fill = 1:ncol(errY))
```

youngForest



```
plot(oldForest)
legend(x = "right",
      legend = colnames(err0),
      fill = 1:ncol(err0))
```



```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  C  s
##           C 41  1
##           s  2 45
##
##           Accuracy : 0.9663
##           95% CI : (0.9046, 0.993)
##           No Information Rate : 0.5169
##           P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.9325
##
## Mcnemar's Test P-Value : 1
##
##           Sensitivity : 0.9535
##           Specificity : 0.9783
##           Pos Pred Value : 0.9762
##           Neg Pred Value : 0.9574
##           Prevalence : 0.4831
##           Detection Rate : 0.4607
##           Detection Prevalence : 0.4719
##           Balanced Accuracy : 0.9659
##
##           'Positive' Class : C
##
```

```
oob_errY <- errY[nrow(errY), "OOB"]
oob_errO <- errO[nrow(errO), "OOB"]

#comparing test sets accuracy to OOB accuracy
paste0("Test Accuracy (younger group): ", ycm$overall[1])
```

```
## [1] "Test Accuracy (younger group): 0.98936170212766"
```

```
paste0("OOB Accuracy (younger group): ", 1-oob_errY)
```

```
## [1] "OOB Accuracy (younger group): 0.970822281167109"
```

```
paste0("Test Accuracy (older group): ", ocm$overall[1])
```

```
## [1] "Test Accuracy (older group): 0.966292134831461"
```

```
paste0("OOB Accuracy (older group): ", 1-oob_errO)
```

```
## [1] "OOB Accuracy (older group): 0.941011235955056"
```

Evaluating test set AUC:

```
#generating predictions on the test set
predfrstY <- predict(object = youngForest,
                    newdata = y_test,
                    type = "prob")
auc(actual = ifelse(y_test$Sibilant == "s", 1, 0),
    predicted = predfrstY[, "s"])
```

```
## [1] 0.9976852
```

```
predfrst0 <- predict(object = oldForest,
                    newdata = o_test,
                    type = "prob")
auc(actual = ifelse(o_test$Sibilant == "s", 1, 0),
    predicted = predfrst0[, "s"])
```

```
## [1] 0.9853387
```

Importance of bins:

```
importance(youngForest)
```

##	MeanDecreaseGini
## bin1	1.6380489
## bin2	1.3796836
## bin3	2.3024663
## bin4	1.2753666
## bin5	1.5647548
## bin6	2.0058502
## bin7	9.8308005
## bin8	26.8463727
## bin9	25.5383454
## bin10	19.9473681
## bin11	15.8953911
## bin12	7.4079618
## bin13	5.2388464
## bin14	1.6767303
## bin15	2.1080205
## bin16	1.7388535
## bin17	2.7811341
## bin18	4.6499822
## bin19	3.6744119
## bin20	4.1087982
## bin21	1.6340413
## bin22	1.3395234
## bin23	1.1293322
## bin24	1.3311670
## bin25	0.6755835
## bin26	1.2228560
## bin27	2.4254639
## bin28	1.8861379
## bin29	1.4909992
## bin30	2.8976410
## bin31	1.8670790
## bin32	2.1373712
## bin33	2.9276755
## bin34	2.1772970
## bin35	2.4411565
## bin36	2.8396451
## bin37	4.2191296
## bin38	4.8214350

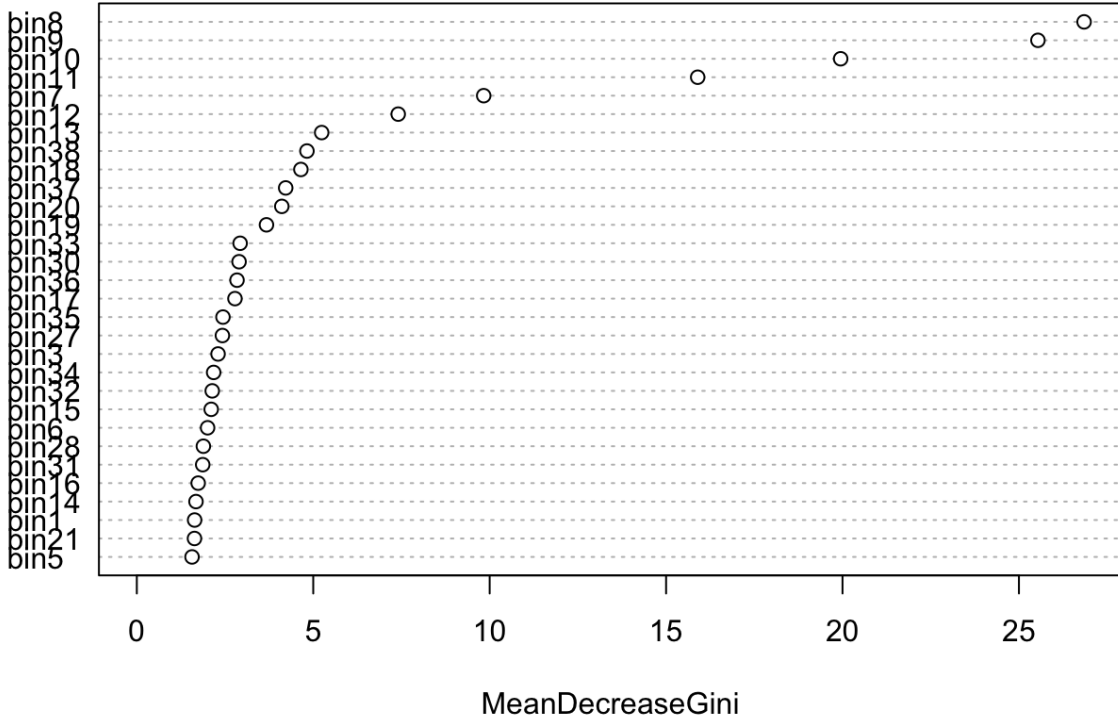
importance(oldForest)

```
##      MeanDecreaseGini
## bin1      2.3017923
## bin2      2.7118105
## bin3      2.3399730
## bin4      3.9826869
## bin5      4.1123224
## bin6      5.3333261
## bin7     21.8883512
## bin8     26.8176590
## bin9     26.9679821
## bin10    14.8008070
## bin11     3.9264227
## bin12     2.4751545
## bin13     1.7319649
## bin14     1.4303022
## bin15     2.9315024
## bin16     2.5822882
## bin17     3.8637288
## bin18     2.9835684
## bin19     1.8948187
## bin20     1.1127499
## bin21     1.5873533
## bin22     1.4883683
## bin23     1.5093237
## bin24     1.6983714
## bin25     1.0736549
## bin26     1.9340230
## bin27     1.9283709
## bin28     1.5414181
## bin29     1.3231258
## bin30     2.0241907
## bin31     1.9299164
## bin32     2.5981397
## bin33     2.3295162
## bin34     3.1588652
## bin35     2.7201841
## bin36     2.4618176
## bin37     1.4880216
## bin38     0.9524424
```

Plotting the importance:

```
varImpPlot(youngForest)
```

youngForest

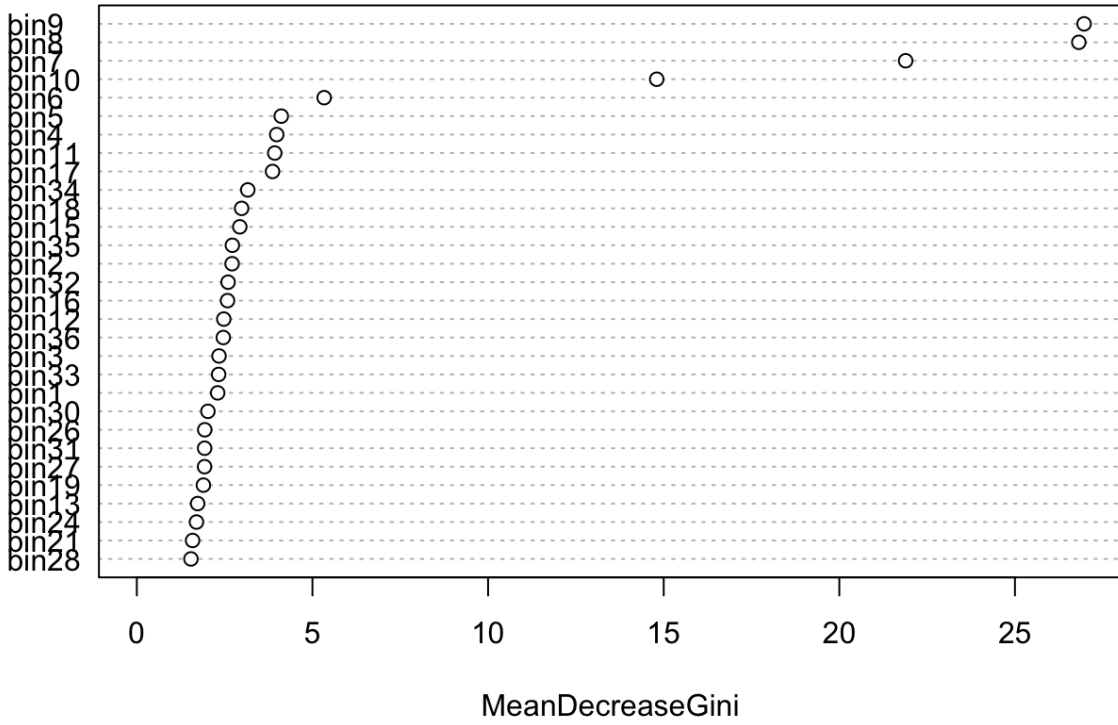


```
#png(filename="youngForest.png")  
#varImpPlot(youngForest)  
#dev.off()
```

^ most relevant bins: 8 >> 9 >> 10 >> 11 >> 7 >> 12

```
varImpPlot(oldForest)
```

oldForest



```
#png(filename="oldForest.png")  
#oForest  
#dev.off()
```

^ most relevant bins 9 >> 8 >> 7 >> 10 >> 6

G Full Results of Acoustic Measurements

184-page file. Downloadable at <https://www.fon.hum.uva.nl/archive/>