



## Investigating potential MERGED and DISTINCT speakers in an ongoing merger—Evidence from production and perception

**Maria Evjen\***, Department of Linguistics and Scandinavian Studies, University of Oslo, Norway, [maria.evjen@iln.uio.no](mailto:maria.evjen@iln.uio.no)

**Timo B. Roettger**, Department of Linguistics and Scandinavian Studies, University of Oslo, Norway, [timo.roettger@iln.uio.no](mailto:timo.roettger@iln.uio.no)

**Sverre Stausland**, Department of Linguistics and Scandinavian Studies, University of Oslo, Norway, [sverre.stausland@gmail.com](mailto:sverre.stausland@gmail.com)

\*Corresponding author.

---

This study investigates whether it is possible to use evidence from production and perception to identify different types of speakers in the study of the ongoing merger of /j/ and /ç/ in Norwegian. Specifically, by investigating the acoustic and perceptual characteristics of the contrast between /j/ and /ç/, we examine whether we can draw a principled line between speakers who have the merger (MERGED speakers) and speakers who do not have the merger (DISTINCT speakers). Our findings indicate that speakers and listeners in part rely on the same parameters when producing and perceiving the contrast between /j/ and /ç/ in Norwegian, most notably spectral standard deviation and intensity. In both production and perception, speakers form continua, from having small or no acoustic differences between /j/ and /ç/ to having large differences, and from having productions which cannot be distinguished by listeners to having productions which are reliably distinguished by listeners. While we find an overlap between these distributions, indicating that some speakers at the extremities are MERGED and DISTINCT speakers, these continua do not allow for a clear boundary between groups. Any grouping of speakers will therefore have to rely on an arbitrary boundary between MERGED and DISTINCT speakers.

---



## 1. Introduction

Phonemic mergers constitute a rather common sound change, found in a range of languages at various points in time. Mergers thus provide ample data for investigating why and how languages change. The study of ongoing mergers in particular allows us to investigate different factors relevant to our understanding of sound changes, such as their driving forces, their outcomes, and the rate and diffusion of change. Specifically, ongoing mergers make it possible to study speakers at different stages of the change, both speakers who have the merger, from now on referred to as MERGED speakers, and speakers who do not yet have the merger, from now on referred to as DISTINCT speakers, as well as speakers who might be somewhere in between these two stages. Comparing the phonetic and phonological patterns of these different types of speakers can provide crucial insights into the details of mergers in general.

In the study of ongoing mergers, then, researchers have made explicit their goal of identifying MERGED and DISTINCT speakers to analyze their patterns (e.g., van Dommelen, 2003, 2019; Jannedy & Weirich, 2017). However, doing so is no trivial task. First, speakers are not necessarily consistent in their behavior and therefore classifying them as MERGED and DISTINCT speakers is not straightforward. Furthermore, the methods used to identify different types of speakers, whether they rely on articulation, acoustics, or perception, can have a large impact on the resulting classification. Nevertheless, a substantial body of research has used these varying methods to classify MERGED and DISTINCT speakers in the investigation of ongoing mergers, for example, in Norwegian (Andrésen, 1980; Papazian, 1994; Dalbakken, 1996; van Dommelen, 2003; 2019), German (Jannedy & Weirich, 2017), and Taiwan Mandarin (Lee-Kim & Chou, 2022).

In the current study, the merger of interest is an ongoing merger of the two voiceless fricatives /ʃ/ and /ç/ in Norwegian. This merger is of particular interest because of its unique position in the public discourse on Norwegian language. The merger has been noted for almost a century (*Undervisningsplan for Oslo folkeskole*, 1940, p. 54) and has been subject to a wide public debate, often portraying the merger and MERGED speakers in a negative light (Simonsen & Moen, 2004, p. 606). As a result, there is a high level of sociolinguistic awareness of this sound change, which in turn might influence its characteristics and diffusion.

Despite this widespread attention in the general public, the merger has been subject to relatively few empirical studies. In these studies, analyses and descriptions of the merger have been based on groups of MERGED and DISTINCT speakers, mainly identified through perception. However, the methods employed to categorize speakers have often relied on subjective measures, such as the researcher's own perception of the fricatives (Papazian, 1994; Dalbakken, 1996; van Dommelen, 2003; 2019), or they have relied on a small number of listeners (Andrésen, 1980).

Research on this and other mergers, then, would benefit from an explicit discussion of the approaches that can be taken to study speakers at different stages of the change in the situation of an ongoing merger. The present paper will present acoustic and perceptual data on the

Norwegian merger to shed light on its phonetic properties, both for individual language users and the population as a whole. This data will contribute to our understanding of the phonetic nature of ongoing mergers and to the discussion of whether it is possible to identify MERGED and DISTINCT speakers in a reliable way.

## 1.1. The Norwegian merger of /ʃ/ and /ç/

In many Norwegian dialects, there is a contrast between the voiceless fricatives /ʃ/, as in /ʃin/ *skinn* ‘skin’, and /ç/, as in /çin/ *kinn* ‘cheek.’ The merger of these two fricatives has been observed for close to a century, but whereas the first observations referred primarily to merging among children, the merger has in recent decades also been noted among adult speakers (Papazian, 1994). This phenomenon is observed in different dialect areas in Norway, such as Oslo (Papazian, 1994), Bergen (Johannessen, 1983), and Trondheim (Dalbakken, 1996). The merger of /ʃ/ and /ç/ is salient both to lay people and linguists, and previous research has documented the occurrence of the merger from both a perceptual and an acoustic perspective. Andrésen (1980) showed that for some speakers, listeners failed to correctly classify their productions of /ʃ/ and /ç/, while van Dommelen (2003) showed that the acoustic characteristics of some speakers’ productions of /ʃ/ and /ç/ were not significantly different. These studies provide evidence for the merger, although we will argue in Section 1.2.1 that the categorization of MERGED and DISTINCT speakers on which these studies base their analyses is potentially unreliable.

The general impression among lay people and researchers is that the fricatives tend to merge as /ʃ/, and most researchers have assumed that this impression holds true in their analysis of the merger (Papazian, 1994; Torp, 1999; van Dommelen, 2019). However, some researchers observe that the fricatives can merge in the opposite direction as well, towards /ç/ (Johannessen, 1983, p. 13), or that the merged product might fall somewhere in between the two fricatives (Andrésen, 1980, p. 94). Not much research has been conducted to thoroughly investigate the outcome of the merger, and the research conducted on this matter has based its analysis on a somewhat subjective method of identifying MERGED and DISTINCT speakers (van Dommelen, 2003). Based on this lack of instrumental research, the question of the outcome of the merger is left open in the current study. This topic will be addressed in Evjen and Stausland (2025), where the insights gained in the current study regarding the identification of MERGED and DISTINCT speakers will be used in the investigation of the outcome of the merger.

## 1.2. Identifying MERGED and DISTINCT speakers in previous research

### 1.2.1. The Norwegian merger

Previous research on the Norwegian merger of /ʃ/ and /ç/ has used varying methods of identifying MERGED and DISTINCT speakers, most of which have relied on perception. Although these studies

provide some useful insights into the behavior of potential MERGED and DISTINCT speakers, they have certain methodological weaknesses that potentially affect their conclusions.

Papazian (1994) conducted a study of the distribution of the merger among 11, 15 and 18 year olds in Oslo using a questionnaire and a reading test. The questionnaire asked speakers to state whether they pronounce eight minimal pairs containing /ʃ/ and /ç/ the same or differently. In the reading test, a subset of the speakers who filled out the questionnaire were asked to read aloud sentences containing these minimal pairs. Papazian then noted whether he believed that the speaker had produced /ʃ/ or /ç/ for each word. This reading test was intended to be a control for whether or not the answers given in the questionnaire accurately represented the speakers' productions.

This method of finding MERGED and DISTINCT speakers is vulnerable to possible biases of the speakers and the researcher. It is widely recognized that when conducting linguistic research, asking people how they speak does not necessarily give an accurate representation of how they actually speak (Labov, 1972, p. 213). This is particularly the case when the object of study is sociolinguistically disfavored, as is the case with the Norwegian merger (Torp, 1999, p. 344). Furthermore, researchers might be subject to confirmation bias (Nickerson, 1998), leading to selectivity in the interpretation and use of evidence to support a hypothesis that the researcher already believes to be true.

Similarly, Dalbakken (1996) relied on her own perception to identify MERGED and DISTINCT speakers in her study of spontaneous and read speech among teenage speakers from Trondheim. These data were subsequently used by van Dommelen (2003) in an acoustic study of the merger. Van Dommelen made some slight modifications to this classification based on his own judgment of the fricatives, and his analysis therefore relied on his and Dalbakken's perception. Relying on the perception of two listeners is vulnerable when classifying the productions of /ʃ/ and /ç/. Specifically, in those cases where the two researchers disagreed, the reader cannot know which of them is more accurate. This point can be anecdotally illustrated by the fact that in our own research on the merger of /ʃ/ and /ç/, the two native speaker authors of this paper have disagreed on the classification of many of the tokens.

In a later acoustic analysis of the merger of /ʃ/ and /ç/, van Dommelen (2019) attempted to improve on this method. In this study, he used a speech corpus including speakers of various Norwegian dialects producing both read sentences and spontaneous speech. To identify MERGED and DISTINCT speakers, van Dommelen used the perception of multiple listeners to categorize the speakers' productions. Specifically, van Dommelen asked seven native Norwegian speakers to rate whether the productions of /ç/ were realized as [ç] or [ʃ]. Importantly, however, van Dommelen only used this method when classifying the productions of /ç/, not the productions of /ʃ/. He states that all productions of /ʃ/ were realized as [ʃ], but it appears that this conclusion was reached based only on his own perception. As stated in Section 1.1, the outcome of the

merger has not been sufficiently documented, and it is possible that some speakers merge the fricatives in the direction of /ç/. Because van Dommelen used his own perception alone when classifying the productions of /ʃ/, such speakers might have been overlooked.

Finally, similarly to van Dommelen (2019), Andrésen (1980) used a perception test to investigate the Norwegian merger of /ʃ/ and /ç/. Andrésen recorded three 13-year-old speakers from Bergen, producing two sets of minimal pairs containing /ʃ/ and /ç/. These productions were subsequently categorized by the three speakers and two researchers trained in phonetics through a perceptual identification task. That is, for each realization, the listeners were asked to report whether they heard /ʃ/ or /ç/. Importantly, the productions of both /ʃ/ and /ç/ were rated. Andrésen did not have an expressed goal of determining whether the three speakers were MERGED or DISTINCT speakers, but the results of the perception study reveal how accurately listeners could classify speakers' productions of /ʃ/ and /ç/. Speakers whose productions were mostly identified correctly probably had a distinct pronunciation, whereas speakers whose productions were often misperceived most likely had a merged pronunciation. This study thus demonstrates a possible approach to identifying MERGED and DISTINCT speakers using a perception study.

### 1.2.2. Other mergers

In attempting to explore whether acoustic measurements and perception studies allow us to identify MERGED and DISTINCT speakers in Norwegian, it is relevant to assess the methods that have been used to find two such groups in the study of other ongoing mergers. The merger of /ʃ/ and /ç/ in German and the merger of alveolars and retroflex sibilants in Taiwan Mandarin serve as relevant examples from other languages (for other studies investigating the degree of merging among speakers in a situation of an ongoing merger see, e.g., Cheng et al. [2022] on Hong Kong Cantonese, Hay et al. [2006] on New Zealand English, and Irons [2007] on Kentucky English).

Jannedy and Weirich (2017) investigated the production and perception of /ʃ/ and /ç/ in certain varieties of German, where, as in Norwegian, the two fricatives are merging. They recorded speakers from three areas: Kreuzberg (a neighborhood in Berlin), Berlin (excluding Kreuzberg), and Kiel. The merger of /ʃ/ and /ç/ is typically associated with the multiethnolect variety of German spoken in Kreuzberg, while speakers from other parts of Berlin and Kiel are not expected to merge the fricatives to a notable extent (Jannedy & Weirich). The speakers were recorded producing minimal pairs containing the fricatives /ʃ/ and /ç/, and in a perception task, 12 listeners were then given the task of identifying the recorded words. The accuracy with which listeners correctly identified the speakers' productions was calculated for each speaker group, and the results for the Kreuzberg speakers were compared with the results for the Berlin and Kiel speakers.

Jannedy and Weirich (2017) thus based their initial categorisation of speakers on dialectal and sociolectal background. Speakers were grouped as Kreuzberg speakers, Berlin speakers, or Kiel speakers, and the perceptual identification task was conducted to assess the accuracy of identification within these groups. Individual speakers' degree of merging was not investigated, and so there might be speakers in the Kreuzberg group who did not merge and speakers in the Berlin and Kiel groups who did. This approach is useful if the purpose of the investigation is to gain an overview of trends within different groups of speakers. However, if the purpose of the study is to identify the degree of merging in individual speakers, one would have to investigate the accuracy scores for each speaker, similar to the method used by Andrésen (1980) in his small-scale study of the Norwegian merger.

Moving on to another merger, Lee-Kim and Chou (2022) investigated the so-called sibilant merger in Taiwan Mandarin. Whereas Standard Mandarin has a contrast between the alveolars /s ts ts<sup>h</sup>/ and the retroflexes /ʂ tʂ tʂ<sup>h</sup>/, Taiwan Mandarin speakers often merge this contrast, resulting in a deretroflexion of the retroflex category (Kubler, 1985). Lee-Kim and Chou recorded speakers of Taiwan Mandarin producing four words containing /s ʂ ts<sup>h</sup> tʂ<sup>h</sup>/, and the recordings were then analyzed acoustically. The mean of the spectral energy distribution (center of gravity) was measured for each fricative. For each speaker, the difference in spectral mean between the alveolars and the retroflexes was calculated. The resulting value was referred to as spectral distance, and Lee-Kim and Chou used this value to identify MERGED and DISTINCT speakers.

Using spectral distance as a measure of speakers' degree of merging builds on previous research on the acoustics of voiceless fricatives in this variety. The spectral mean, the acoustic parameter examined by Lee-Kim and Chou (2022), has been found to be an important acoustic parameter to the distinction between alveolars and retroflexes in Taiwan Mandarin (Lin & Wu, 2023). This parameter has also been found to pattern with listeners' perception of the sibilants /s ʂ/ (and /ç/) in an identification task with Taiwan Mandarin speakers and listeners (Chiu et al., 2020).

The results of the production study indicate that speakers formed a continuum from speakers with a spectral distance around 0—indicating a merged pronunciation—to speakers with a spectral distance of 4000 Hz—indicating a distinct pronunciation. To divide the speakers on this continuum into two groups, Lee-Kim and Chou (2022) investigated the spectral energy distribution of alveolars and retroflexes for each speaker. They then investigated the differences between these distributions using a Bayes factor test. Speakers who had no significant difference between the two categories were classified as MERGED speakers, and speakers who had a significant difference were classified as DISTINCT speakers.

This approach thus relies on acoustic measurements rather than perception in the identification of MERGED and DISTINCT speakers. An important factor in this analysis is that previous research has investigated which acoustic parameter is most relevant to the distinction

at hand. The speakers in Lee-Kim and Chou's study were ranked in a certain order according to spectral distance, from speakers with smaller values (the MERGED speakers) to speakers with larger values (the DISTINCT speakers). However, it is not a given that this order would be the same if one were to investigate the differences between alveolars and retroflexes based on other acoustic parameters. In attempting to identify MERGED and DISTINCT speakers, then, it is an advantage to know which acoustic parameter is most important in signaling the relevant contrast in the language or variety under investigation. If this is not known, it is necessary to measure a range of acoustic parameters and investigate their importance to the contrast at hand.

However, even if it is possible to obtain a ranking of participants based on the size of their acoustic distance for a given parameter, it is not necessarily the case that one can divide these speakers into groups in a non-arbitrary way. In Lee-Kim and Chou's (2022) study, speakers' spectral distances formed a continuum, and there was no obvious cutoff point for inclusion in the MERGED and DISTINCT speaker groups. A boundary between speakers was obtained using statistics, but because there was no evidence for a clear separation of speakers in the data, such a boundary is not necessarily informative. If speakers form a continuum rather than separate groups, then, it is possible to categorize speakers, but not without making arbitrary decisions.

### 1.3. The current study

Following the review in Section 1.2 of the methods used to categorize speakers as MERGED and DISTINCT speakers in the study of the ongoing mergers in Norwegian, German, and Taiwan Mandarin, it appears that previous research has relied on listener perception or acoustic measurements. Acoustic measurements, on the one hand, constitute an instrumental approach to exploring the patterns of different speakers. However, if it is not known beforehand which acoustic parameters contribute the most to the separation between the two categories under investigation, it is necessary to first investigate a range of parameters and statistically analyze their contribution to the contrast at hand. Perception studies, on the other hand, can provide useful insights into which speakers do and do not produce a contrast from the perspective of listeners. For such studies to be useful, it is necessary to include a large number of listeners.

In the current analysis, we aim to characterize speakers' realization of the contrast between /ʃ/ and /ç/ in Norwegian, from both an acoustic and a perceptual perspective, and subsequently to compare the results of the two analyses. Our goal is to evaluate whether it is possible, on acoustic or perceptual grounds, to identify MERGED and DISTINCT speakers in the case of the merger of /ʃ/ and /ç/.

The paper is structured as follows. Section 2 presents a production study which includes an acoustic analysis of a range of parameters and a statistical analysis of the degree of separability each parameter contributes to the production of the contrast between /ʃ/ and /ç/. Section 3 presents a perception study in which listeners were tasked with identifying speakers' fricatives.

The results were submitted to analyses of the level of correct classification of the two fricatives, along with a statistical analysis of which acoustic parameters contribute the most to the separation between /ʃ/ and /ç/ in perception. Section 4 provides a comparison of the production and perception results, as well as a discussion of whether we can identify MERGED and DISTINCT speakers based on the current findings. Section 5 concludes the paper.

All stimuli, data tables, statistical models, and scripts used in the current study are publicly available here: <https://osf.io/e3jn5/>.

## 2. Production study

In this experiment, native speakers of Norwegian were recorded, and their speech was acoustically analyzed. The results of the acoustic analysis were submitted to statistical analyses with the aim of exploring which acoustic parameters contribute more and less to the separation between /ʃ/ and /ç/ in production. Based on the findings of the statistical analysis, speakers' acoustic differences between /ʃ/ and /ç/ were evaluated for those parameters which were found to give the largest degree of separation.

### 2.1. Methods

#### 2.1.1. Participants

Forty-two native speakers of Norwegian participated in the experiment (32 female, 10 male; aged 18–52, mean age 24). Because the realization of /ʃ/ and /ç/ varies between different Norwegian dialects, /ʃ/ being produced as [sj] in certain dialects and /ç/ being produced as [tʃ] in certain dialects (Torp, 1999), participants' native dialect was restricted to East Norwegian. In this dialect, both phonemes are invariably produced as fricatives. Participants were classified as speakers of East Norwegian if their home municipality was within a 120 kilometer radius from the city center of Oslo. Furthermore, only participants whose only native language is Norwegian, and whose parents also have Norwegian as a native language, were included. Participants were first-year students of Scandinavian studies completing an introductory course in Norwegian grammar, and they received course credits for participation.

#### 2.1.2. Materials

Participants were asked to read out loud a set of nonce words of a CV structure. To conceal the purpose of the experiment, the nonce words contained not only /ʃ/ and /ç/ as C, but all possible consonants of Norwegian. As V, the nonce words contained either of the vowels /i a u/. This resulted in 51 nonce words (17 consonants × 3 vowels), and each of these was read twice. Each participant thus read 102 words in total.

The purpose of using nonce words rather than real words was to access speakers' abstract categories for the phonemes /ʃ/ and /ç/, rather than their word-level representations of real



words. Furthermore, using nonce words allows us to control for lexical frequency to a greater extent, a factor which has been shown to affect the rate and diffusion of sound changes (Todd et al., 2019). Some of the nonce words happened to be real words of Norwegian, such as /sa/ and /ʃi/. However, the nonce words were presented in such a way that there was no indication that the words carried any meaning. They were also not necessarily spelled in the same way as the corresponding Norwegian word. For example, the Norwegian word /ʃi/ 'ski' is spelled «ski», while the nonce word /ʃi/ was spelled «sj» (see Section 2.1.3). These factors are likely to minimize the possible effect of nonce words corresponding to real words.<sup>1</sup> The usefulness of using nonce words in the investigation of mergers has been demonstrated by Hay et al. (2013).

### 2.1.3. Recording

The participants were recorded in the Sociocognitive Lab at the University of Oslo. The recordings were made using a head-mounted AKG MicroMic C544 L condenser microphone and a Focusrite Scarlett 2i2 3rd generation audio interface. The recordings were made in Audacity version 3.4.2 with a sampling rate of 41 000 Hz. Each participant was seated opposite the researcher and informed about the task they were to complete. Participants were then placed in front of a laptop where they were shown a PowerPoint presentation. Each slide in the presentation contained one nonce word, presented in a randomized order, and the participants were asked to read the word out loud. The presentation ran automatically once the participant clicked start, and the 51 nonce words were displayed with a two-second interstimulus interval. Participants were then offered to take a break before the 51 nonce words were repeated in the same manner, in a new randomized order. Each participant was presented with a different order.

The orthographic forms «sj» and «kj» were used to prompt participants to produce /ʃ/ and /ç/. The two fricatives can be spelled in different ways in Norwegian, and no letter can signal /ʃ/ or /ç/ alone. The forms «sj» and «kj» are the most common representations of these phonemes, and they are often referred to as the sj-sound and the kj-sound among the general public. What's more, «sj» and «kj» unambiguously signal /ʃ/ and /ç/, respectively, in Norwegian. For these reasons, this orthography was chosen.

## 2.2. Analysis

### 2.2.1. Segmentation

The fricatives /ʃ/ and /ç/ were identified in the recorded sound files and subsequently segmented and annotated by the first author using Praat (Boersma & Weenink, 2022). Because it was expected that some speakers would have the merger, it was necessary to consult the PowerPoint

---

<sup>1</sup> Among the nonce words containing /ʃ/ and /ç/, only /ʃi/ is a real word in Norwegian. Because this is the only word in the materials in which we find /ʃ/ before /i/, it was not possible to separate potential effects of /ʃi/ being a real word from potential effects of the vowel context.

presentation presented to each participant to determine which fricative the participant had been instructed to produce. The annotations reflected the intended fricative rather than the perceptual impression. Both the waveform and the spectrogram were consulted when identifying the boundaries of the fricatives. The onset of each fricative was marked at the onset of high-frequency energy as indicated in the spectrogram, and at the onset of rapid zero-crossings in the waveform. Similarly, the end of the fricative was marked at the offset of high-frequency energy and the zero-crossing preceding either the onset of periodicity associated with the following vowel or a period of aspiration. Norwegian fricatives are not phonologically aspirated, but in the recorded materials, many of the fricatives appear to be phonetically aspirated, possibly as a result of hyperarticulation. This aspiration was identified as a period of low-intensity energy following the offset of the fricative and preceding the onset of periodicity of the vowel. Aspiration was not segmented as part of the fricative.

In total, the experiment resulted in 504 tokens (42 speakers  $\times$  2 fricatives  $\times$  3 vowel contexts  $\times$  2 repetitions). However, three speakers were excluded because they reported to have a speech impairment affecting the production of sibilants. Furthermore, some speakers were unsure of how to interpret the <sj> and <kj> prompts, and they therefore produced these sequences as [sj] and [kj] rather than [ʃ] and [ç] in some cases. One of these participants was excluded because all of their /ç/ tokens were produced as [kj]. The other participants who occasionally produced [sj] and [kj] were included in the analysis, although these specific tokens were excluded. The exact number of productions of each fricative in each vowel context thus varied slightly between participants. The resulting number of fricatives submitted for analysis, excluding the [sj] and [kj] tokens, was 449, produced by 38 speakers.

In addition to the fricatives, the vowel following the fricative was segmented and annotated. The onset of the vowel was identified as the onset of periodicity following the preceding fricative, or in those cases where the fricative was phonetically aspirated, following the aspiration. The end of the vowel was identified as the offset of periodicity at the end of the word. The purpose of segmenting the vowels was to conduct analyses of the second formant (F2) in the transition between the fricatives and the following vowels.

### 2.2.2. Acoustic analysis

Prior research on which acoustic parameters contribute most to the distinction between /ʃ/ and /ç/ in Norwegian is limited. Only van Dommelen (2003, 2019) has investigated the acoustics of these fricatives, measuring duration, relative root-mean square (rms) amplitude, and spectral characteristics such as the spectral peak and center of gravity. However, as argued in Section 1.2.1, van Dommelen's analyses in large part rely on his own perception in categorizing speakers' productions of /ʃ/ and /ç/, possibly affecting the results. Previous research on the distinction between /ʃ/ and /ç/ in German has investigated spectral moments and discrete cosine

transformation (DCT) coefficients and found that DCT coefficients appear to more accurately capture the acoustic differences between the fricatives (Jannedy & Weirich, 2017). Research on Luxembourgish, however, measured the center of gravity and spectral peak of /ʃ/ and /ç/, finding that these parameters adequately differentiate between the fricatives (Conrad, 2023).

Looking at previous research examining voiceless fricatives in general, most studies have measured all or a subset of the following acoustic parameters: spectral moments, spectral peak, duration, F2 transitions, and intensity or rms amplitude (e.g., Forrest et al. [1988] and Jongman et al. [2000] on English, Nirgianaki [2014] on Greek, and Wikse Barrow et al. [2022] on Swedish). Because of the scarcity of research on Norwegian, and for comparability with studies of voiceless fricatives in a range of languages, we chose to focus on these acoustic parameters in the current study.

Starting with the first four spectral moments, then, these measure a spectrum's energy distribution, where the spectrum is treated as a probability density distribution (Forrest et al., 1988). The measurements consist of 1) the spectral mean, often referred to as the center of gravity, representing the mean of the energy distribution, 2) the spectral standard deviation, referring to the spread of the energy with regard to the mean, 3) the skewness of the distribution, indicating whether the distribution is asymmetric towards higher or lower frequencies, and 4) the kurtosis of the distribution, referring to whether the distribution is more or less peaky than a normal distribution.

In addition to spectral moments, we measured the fricatives' duration and intensity, and the F2 at the point of transition between the fricative and the vowel (Jongman et al., 2000). Furthermore, because the fricatives showed signs of phonetic aspiration, the duration of aspiration and its distribution across fricatives were analyzed.<sup>2</sup>

Spectral moments were analyzed in Praat using a script developed by DiCanio (2013). The script computed discrete Fourier transforms (DFTs) from a set of windows placed across the duration of the fricative. The windows were then time-averaged (Shadle, 2012), meaning that the DFTs were averaged for each token before measuring and collecting the spectral moments. This way, the measurements were not as sensitive to fluctuations in the signal as they would be if they were based on one window. In our analysis, we extracted 6 Hann windows of 15 ms evenly spaced across the middle 80% of the fricative. The script used a low-pass filter of 300 Hz. In

---

<sup>2</sup> We also measured the frequency of the spectral peak, as this parameter has been found to distinguish between English voiceless fricatives (Jongman et al., 2000). Spectral peak was measured using a script developed by Wikse Barrow et al. (2022), extracting a 20 ms Hann window from the midpoint of the fricative, and computing a long-term average spectrum (LTAS) object from which the maximum frequency was measured. However, this parameter was excluded from the analysis, as we discovered that numerous measurements contained much lower values than the majority and that these values were often exactly the same across different tokens. Because it was not clear what caused this peculiarity, we decided to exclude the spectral peak measurements.

addition to the spectral moments, the script measured the average intensity across the extracted windows and the total duration of the fricative.

Aspiration was measured using a modified version of the Praat script developed for the acoustic analysis in Wikse Barrow et al. (2022). This script measured the duration of the interval between the offset of the fricative and the onset of the following vowel. In those cases where the fricative was not phonetically aspirated, then, aspiration was measured to be zero. As for F2 transitions, these were measured using the script developed by García (2017). Measurements were collected from a 25 ms window placed at the onset of the vowel.

Two of the measured acoustic parameters, aspiration and kurtosis, showed positively skewed distributions and were therefore log-transformed (see Winter [2020, pp. 90–91] for a discussion of log-transformation of skewed data). The kurtosis variable contained negative values, and consequently a constant was added to each measurement prior to the log-transformation. The log-transformed measurements for aspiration and kurtosis were used in all subsequent analyses in the study. Moreover, in the acoustic analysis, all parameters were standardized for speaker and vowel using z-scores to avoid speaker-specific and vowel-specific idiosyncrasies influencing the analysis and to allow for comparison across the different acoustic parameters. Z-scores were calculated by subtracting the mean and dividing by the standard deviation within each grouping, with the result that all variables have a mean of zero and a standard deviation of 1.

### 2.2.3. Statistical analysis

The contribution of each acoustic parameter to the distinction between /ʃ/ and /ç/ was estimated by modeling the effect of fricative on the different parameters and subsequently ranking the estimates. Parameter estimation and inference was performed within the Bayesian framework. In Bayesian models, we approximate a posterior distribution, that is, a distribution of plausible parameter values given the data, the model and any prior assumptions about the estimated parameter values. Because the approximation of a posterior distribution is computationally costly, we use algorithms to sample values from the posterior distribution. The practical consequence of this choice is that we do not calculate a single point estimate for a parameter, but, in our case, we draw a sample of 8,000 plausible values for that parameter and quantify our uncertainty about it by summarizing the distribution of values (see Vasishth et al., 2018; Franke & Roettger, 2019, for tutorials using phonetic examples).

To examine which acoustic parameters speakers use to signal the contrast between /ʃ/ and /ç/, we fitted a set of Bayesian linear mixed models to the production data. The models were computed in R version 4.4.1 (R Core Team, 2021) using the package brms version 2.21.0 (Bürkner, 2017). The brms package allows the user to compute models with Stan, making use of Markov Chain Monte Carlo (MCMC) sampling (Stan Development Team, 2024). The R script

used for the analysis, including the computation of models, standardization of variables, and plotting, was adapted from Roessig et al.'s (2022) publicly available script.

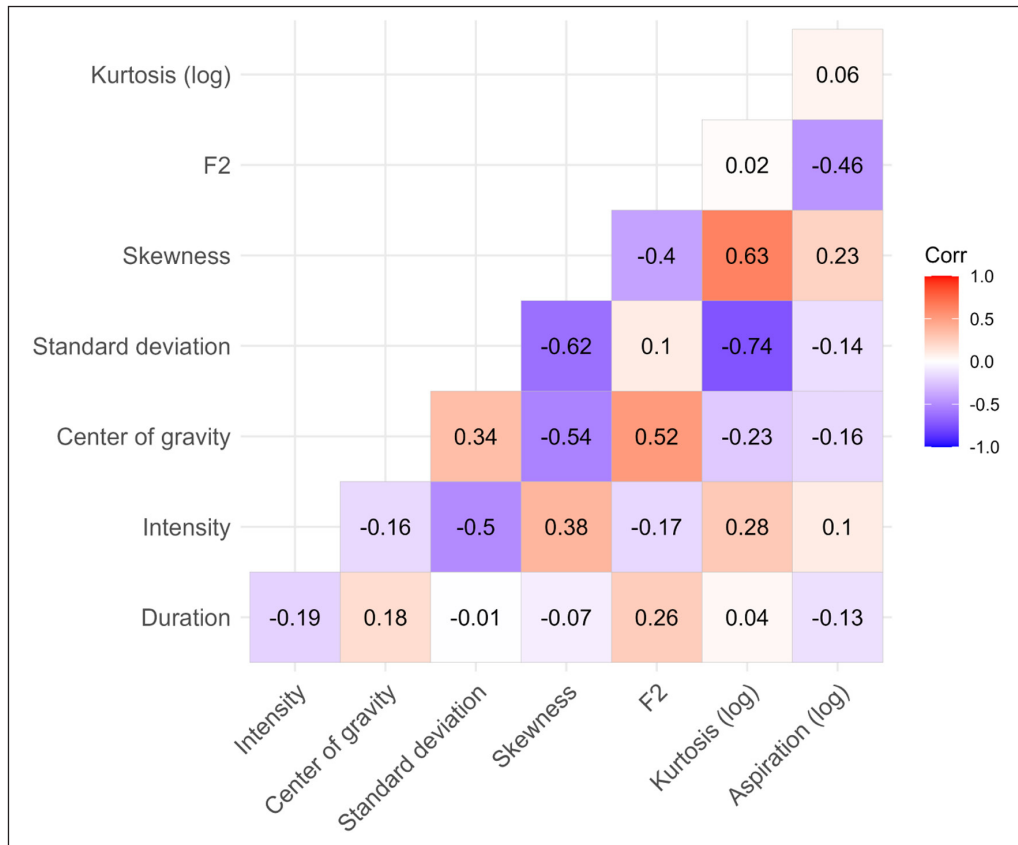
Separate models were computed for each of the eight acoustic parameters, following the analysis conducted in Roessig et al. (2022). The measurements of the acoustic parameters were standardized using z-scores before fitting the individual mixed-effects models. As stated in Section 2.2.2, z-scoring has the effect that all parameters have a mean of 0 and a standard deviation of 1, making it possible to compare the beta weights from each model. Each model contained the relevant acoustic parameter as the response variable, while fricative (/ʃ, ç/), the following vowel (/i, a, u/), and the interaction between them were entered as fixed effects. Because the nonce words were made up of only the fricative under investigation and the following vowel, the variable item was not entered as a random effect in the models. As random effects, the models included by speaker varying intercepts and varying slopes for fricative by speaker. The model structure was thus as follows:

$$acoustic\ parameter \sim fricative*vowel + (1 + fricative|speaker)$$

The approach of fitting separate models for each parameter rather than entering all the parameters as predictors in one model was chosen for two reasons: First, the different acoustic measures are correlated to differing degrees, leading to collinearity. The correlation matrix in **Figure 1** below illustrates the correlation between the eight acoustic parameters measured in the production study. The strongest correlation is found between spectral standard deviation and kurtosis, showing a correlation of  $r = -0.74$ , but several other parameters show correlations up to  $r = 0.63$ . As illustrated in Figure 3(a) in Schertz and Clare (2020, p. 5) (adapted from Clayards [2018]), correlated acoustic cues can play independent roles in signaling a contrast between two categories. However, correlations between parameters pose difficulties for statistical modeling. Specifically, collinearity between two or more predictors in a mixed model renders the model results uninterpretable. While there are suggested “remedies” for collinearity, statistically sound approaches such as random forests and supervised component regression come with their own limitations. Random forests cannot straightforwardly estimate variance arising from speakers who behave differently, typically incorporated as a random effect in a mixed model (Tomaschek et al., 2018, p. 265), while supervised component regression has low predictive accuracy (Tomaschek et al., p. 260). Moreover, these different approaches do not converge on the same results (Tomaschek et al.).

Second, a model with several acoustic parameters as fixed effects should ideally account for speaker-specific variation through random slopes to avoid inflated false positive rates (e.g., Barr et al., 2013). The complexity of such a model would quickly become computationally intractable. The approach of running separate models for each parameter, although not capturing correlation

and interactions between the parameters, was therefore chosen. The consequence of this choice is that we cannot isolate the impact of individual acoustic dimensions per se. Instead, the effect magnitude of a parameter represents the magnitude of this parameter along with its correlates. Since we do this for all parameters, they remain comparable.



**Figure 1:** Correlations between the eight acoustic parameters.

In the computation of the models, each model ran four MCMC chains for 4000 iterations with 2000 warmup iterations, thus resulting in 8000 posterior samples used for inference. We used regularizing, weakly informative priors for the models (Gelman et al., 2017). For the regression coefficient of interest, we used a normally distributed prior with a mean of 0 and a standard deviation of 0.5. For the intercept, we used a normally distributed prior with a mean of 0 and a standard deviation of 1. For the standard deviation of the random intercepts and slopes and the residual standard deviation, we used a Cauchy distributed prior with a mean of 0 and a standard deviation of 0.5. *Rhat* values were inspected and indicated convergence of each model, and all

models exhibited a high amount of effective samples for all parameters. We therefore concluded that all models were healthy.

In the interpretation of the output of the models, we compared the estimates of the posterior distributions for the two fricatives /ʃ/ and /ç/ for each acoustic parameter. Taking center of gravity as an example, our goal was to investigate whether /ʃ/ had smaller or larger measurements for this parameter than /ç/. We therefore extracted the posterior distributions for /ʃ/ and /ç/ from the model containing center of gravity, and we subtracted the values of the distribution for /ʃ/ from the values of the distribution for /ç/. The mean of the resulting distribution represents the estimated difference in center of gravity between the two fricatives. The 95% Credible Intervals of the distribution were extracted, representing the range of values for which we can be 95% certain we will find the true value of the parameter, given the data, the priors, and the model (Vasishth et al., 2018, p. 152). If the 95% Credible Interval does not contain zero, then, it is likely that there is in fact a difference between /ʃ/ and /ç/ for center of gravity. Importantly, the model estimates also allow us to evaluate effect sizes, which means that we can compare and rank the magnitudes of the differences between /ʃ/ and /ç/ for the different acoustic parameters.

## 2.3. Results

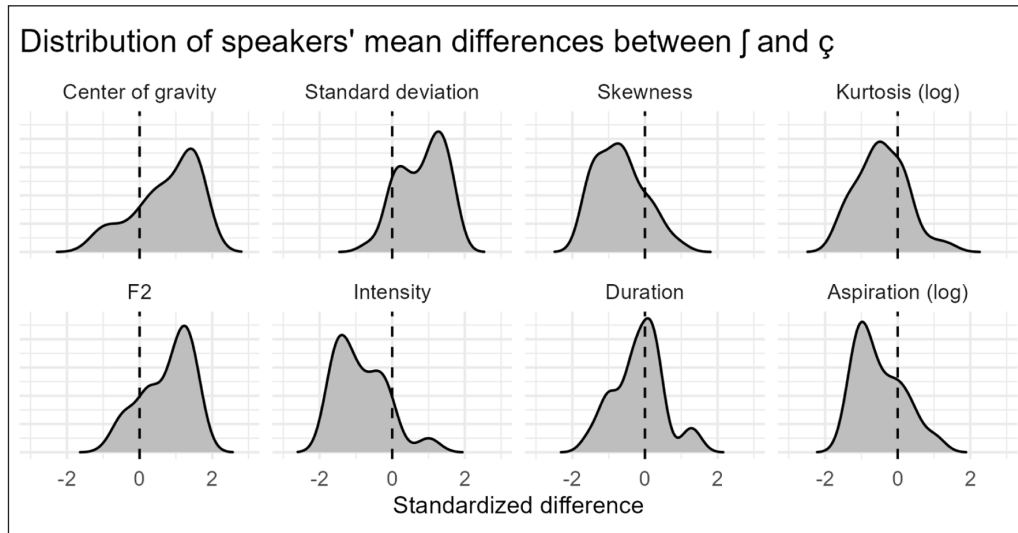
### 2.3.1. Acoustic results

As the purpose of this study is to explore whether it is possible to characterize speakers in our population as MERGED and DISTINCT speakers, the most important measurements for the descriptive analysis of the production data are not the recorded values for each acoustic parameter in themselves, but rather the differences between /ʃ/ and /ç/ for each speaker and each acoustic parameter. By studying the distributions of these differences, we can investigate the degree of merging in the population, in that larger differences are indicative of a contrast between the two fricatives, and smaller differences are indicative of a weakening of the contrast.

In order to perform a descriptive analysis of the differences between /ʃ/ and /ç/ for each parameter, speakers' mean values for the two fricatives were calculated, and the mean value for /ʃ/ was subtracted from the mean value for /ç/. **Figure 2** illustrates the mean differences between /ʃ/ and /ç/ for the acoustic parameters measured in Section 2.2.2. A positive difference indicates that /ç/ had a larger mean value than /ʃ/, while a negative difference indicates that /ʃ/ had a larger mean value than /ç/. Density plots were computed using ggplot2 from the Tidyverse package (Wickham et al., 2019) in RStudio (RStudio Team, 2019), and the default values for the smoothing bandwidth and kernel were used.

**Figure 2** illustrates that for all parameters but duration, a majority of speakers appear to produce a difference between the two fricatives, as indicated by peaks in the distributions left or

right of zero. However, we also find a number of speakers close to zero, indicating that certain speakers do not produce a difference between /ʃ/ and /ç/ for these parameters. For duration, on the other hand, a majority of speakers have a mean difference of zero.



**Figure 2:** Distributions of speakers' mean difference between /ʃ/ and /ç/ for each of the eight measured acoustic parameters. All acoustic parameters were standardized for speaker and vowel. A positive difference indicates that /ç/ had a larger mean value than /ʃ/, while a negative difference indicates that /ʃ/ had a larger mean value than /ç/.

Interestingly, speakers' differences are found on both sides of zero for most of the parameters, indicating that while some speakers have higher values for /ʃ/ than for /ç/ for a given parameter, others have the opposite pattern. This finding could possibly indicate that the different acoustic parameters interact in ways we do not fully understand. For example, speakers could differ in what acoustic cue they emphasize, with the result that their less important cues fluctuate somewhat randomly around zero.

Speakers' mean differences thus show that for most parameters, we find some speakers who produce a clear difference between /ʃ/ and /ç/, while other speakers produce only a small difference or no difference. There is, however, no clear grouping of speakers. If the speakers in this study formed two distinct groups of MERGED and DISTINCT speakers, we would expect the distribution of at least some of these parameters to fall into two separate bins, with one group of speakers with a mean difference around zero, and another group of speakers centered on a larger positive or negative value. Since we do not find such a bimodal distribution in the data, we cannot identify MERGED and DISTINCT speakers based on the measured acoustic differences and, furthermore, we cannot determine which of these acoustic parameters are most important



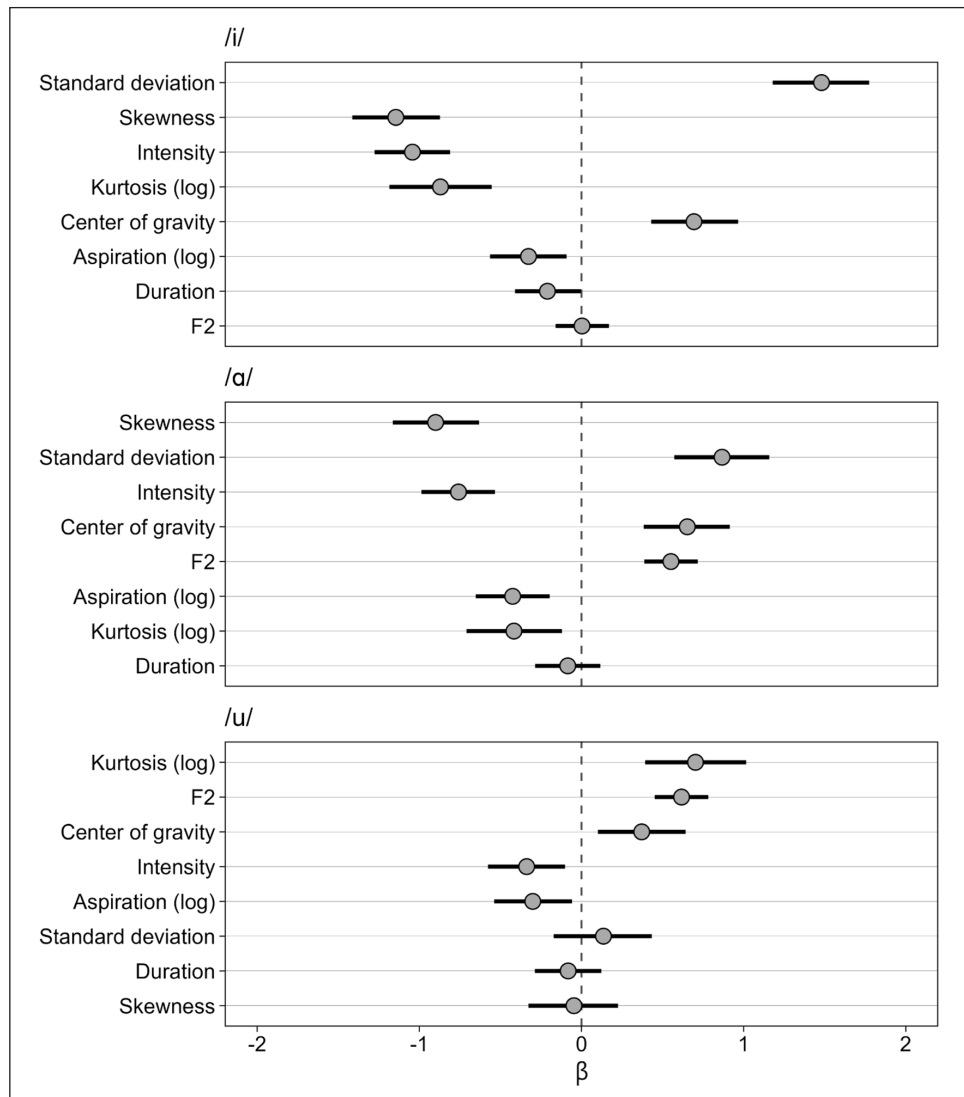
to the production of the contrast between /ʃ/ and /ç/. Consequently, we turn to the statistical analysis investigating which of the measured acoustic parameters contribute more and less to the production of this contrast.

### 2.3.2. Ranking of cues

The differences between the two fricatives were modeled based on the change from /ʃ/ to /ç/, and the magnitude of the difference for each acoustic parameter was represented by the size of the coefficients for each model. A positive coefficient indicates that the measurements for /ç/ were higher than those for /ʃ/ for the relevant parameter, while a negative coefficient indicates that the measurements for /ʃ/ were higher than those for /ç/. The coefficients were ranked to compare their effect magnitude. We will discuss the effect magnitudes alongside their uncertainty by the posterior means and their 95% Credible Interval. Only a subset of the effect magnitudes are reported here, but posterior means and 95% Credible Intervals for all parameters in all vowel contexts can be found in Appendix A.

**Figure 3** illustrates this ranking of coefficients, separated by vowel context. Overall, /ʃ/ and /ç/ seem to be well separated in production, with large effect magnitudes for all vowel contexts. However, as illustrated, the ranking varies substantially depending on the vowel, with /u/ standing out in particular from /i/ and /a/. The results indicate that the separation between /ʃ/ and /ç/ is greatest before the vowel /i/ and smallest before the vowel /u/. While the largest beta weight for /i/ is standard deviation with a size of 1.48 [95% CrI: 1.18, 1.77], the largest beta weight for /u/ is kurtosis with a size of 0.70 [0.39, 1.02]. The effect magnitude for the rest of the parameters are also overall larger before /i/ than before /u/. As for the /a/ context, the largest beta weight is skewness (-0.90 [-1.16, -0.63]), and the effect magnitudes for the remaining parameters appear to fall in between those for /i/ and /u/.

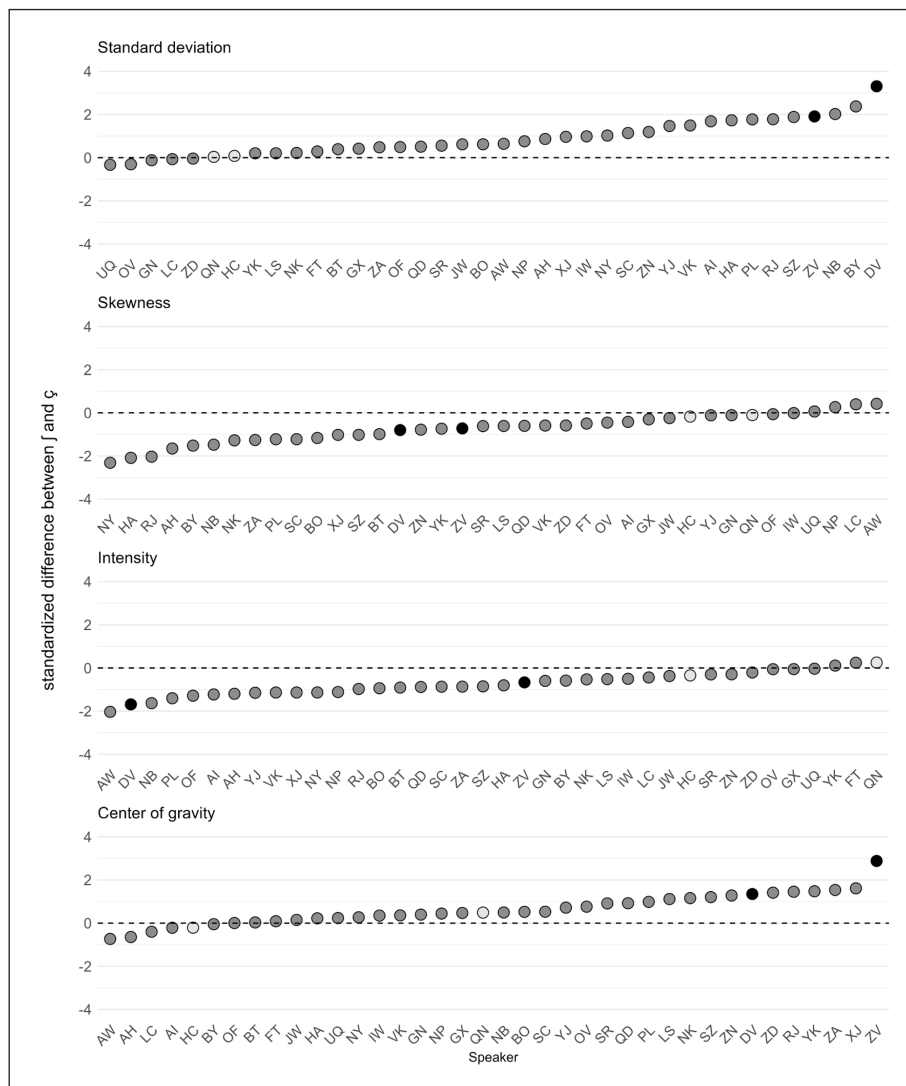
Looking at the overall separability of the different acoustic parameters, we find some patterns despite the variation by vowel context. The parameters that appear to give a reliable separation between /ʃ/ and /ç/ across all three vowel contexts are intensity, center of gravity, and kurtosis. The largest effect magnitudes, however, are found for standard deviation and skewness, but only in the /i/ and /a/ contexts. These two parameters do not appear to separate between /ʃ/ and /ç/ in the /u/ context. F2 has a high ranking before /u/ and is found in the middle of the ranking before /a/, although the effect magnitude in the two contexts are similar. Before /i/, however, F2 appears to give no separation. A further discussion of these context effects will follow in Section 4.2. Common to all three vowel contexts is the finding that aspiration and duration are found towards the bottom of the rankings with relatively small beta weights. We can therefore conclude that these parameters do not contribute notably to the separation between /ʃ/ and /ç/ in production.



**Figure 3:** Ranking of the coefficients from the individual regression models, indicating the effect magnitudes of the different parameters. Whiskers indicate 95% Credible Intervals.

Having identified certain parameters which contribute more to the contrast than others, we can investigate speakers' mean differences between /j/ and /ç/ for these parameters specifically. Choosing the four parameters which appear to give the largest degree of separation for the three vowel contexts combined, namely standard deviation, skewness, intensity, and center of gravity, we examined speakers' mean differences. **Figure 4** presents the mean difference for each speaker for these four parameters. As in **Figure 2** above, speakers found around zero do not have a notable difference for the relevant acoustic parameter, whereas speakers found to have larger positive or negative values produced a difference for this particular parameter. **Figure 4** makes

it possible to compare speakers' differences across parameters, and we can observe that for some speakers, we find some consistencies. Speakers QN and HC, for example, have differences close to zero across these parameters. Speakers, DV and ZV, on the other hand, have large differences for two of these four parameters. However, there is also a lot of variation between parameters, and the distributions form continua, making it impossible to establish non-arbitrary boundaries between speakers.



**Figure 4:** Individual speakers' mean difference between /j/ and /ç/ for the four parameters, which seem to contribute most to the separation between these phonemes in production. Examples of speakers with smaller differences (QN, HC) across the different parameters are indicated with light gray dots, and examples of speakers with larger differences (DV, ZV) are indicated with black dots.

We also note that **Figure 4** does not take vowel context into account. It is possible that speakers' mean differences between /ʃ/ and /ç/ for the different parameters are the result of more complex within-speaker patterns depending on the following vowel. These potential patterns are not explored further here, but figures illustrating the rankings of speakers' mean differences for all acoustic parameters separated by vowel context can be found in Appendix B.

The results of the production study, then, point to certain speakers who have smaller and larger differences, respectively, between /ʃ/ and /ç/. However, despite certain patterns, it is clear that there is no undisputable way to group the speakers in this sample as MERGED and DISTINCT speakers based on acoustics alone.

### 3. Perception study

We now turn to a perceptual assessment of the contrast between /ʃ/ and /ç/ as produced by the speakers in the production study. In this experiment, a perceptual identification task similar to the studies by Andrésen (1980) and Jannedy and Weirich (2017) was conducted. Norwegian listeners were exposed to the nonce words recorded in the production study, and for each word, they were asked to identify whether the first sound of the word was /ʃ/ or /ç/. The aims of the experiment were to analyze the level of correct identification of productions of /ʃ/ and /ç/, first from the perspective of the listeners and subsequently from the perspective of the speakers, and to study which of the acoustic parameters analyzed in Section 2 contributed more and less to listeners' choice of fricative.

#### 3.1. Methods

##### 3.1.1. Participants

Sixty-four native speakers of Norwegian participated in the experiment (49 female, 15 male; aged 18–50, mean age 24). As in the production study, participants were first-year students of Scandinavian studies completing an introductory course in Norwegian grammar, and they received course credits for participation. The criteria for inclusion were less strict in this study than in the production study. As well as East Norwegian speakers, speakers with other Norwegian dialects and other native languages in addition to Norwegian were included. As East Norwegian constitutes the standard spoken language in Norway, all participants have been extensively exposed to this variety throughout their lifetime, and they all lived in Oslo at the time of participation. It is therefore reasonable to assume that their perception of /ʃ/ and /ç/ is comparable to the perception of East Norwegian speakers.<sup>3</sup>

---

<sup>3</sup> An investigation of the overall identification scores of listeners with other language backgrounds as compared to the overall identification scores of East Norwegian listeners indicated that these groups were in fact comparable (mean proportion of correctly identified tokens of 74.4% and 71.2%, respectively).

Furthermore, participants were asked to report whether they had any speech or hearing impairments. One participant reported to have a speech impairment, but because only perception was under investigation in the current experiment, this participant was nonetheless included. None of the participants reported to have a hearing impairment.

### 3.1.2. Materials

The experiment consisted of a practice phase and a test phase, and the materials used in each phase were different. In the practice phase, the materials consisted of nonce words produced by three native speakers of East Norwegian who did not participate in the production study (2 female, 1 male; aged 30–42). The nonce words were identical to the nonce words containing /ʃ/ and /ç/ in the production study, giving 6 words with a CV structure, where C was /ʃ/ or /ç/ and V was /i/, /a/, or /u/. Each speaker produced these 6 nonce words once, giving 18 words in total in the practice phase. The speakers were aware of the purpose of the experiment, and they were recorded under the same conditions as the participants in the production study. The purpose of including this practice phase was to allow the listeners to familiarize themselves with the task they were going to complete.

In the test phase, the recorded nonce words containing /ʃ/ and /ç/ from the production study were used as stimuli. To ensure that an equal number of productions of each fricative was included for each speaker, the speakers who occasionally produced [sj] and [kj] for /ʃ/ and /ç/ in the production study (see Section 2.2.1) were excluded. This left 34 speakers whose productions were included as stimuli in the perception study. The materials consisted of 6 words, repeated twice, for each of the 34 speakers, resulting in 408 tokens in total.

### 3.1.3. Procedure

The experiment took place in the Sociocognitive Lab at the University of Oslo, and the data were collected through a Praat MFC Experiment (Boersma & Weenink, 2022). Participants were seated in front of a laptop computer and informed that they would be participating in a listening experiment. As mentioned above, the experiment included a practice phase of 18 trials and a test phase of 408 trials. In both phases, the nonce words were presented in a randomized order. The test phase contained 6 blocks of 68 trials, with breaks between each block.

Participants were instructed that they would hear a series of words and that their task was to choose which sound each word started with. They were told that the first sound would always be one of two sounds: the *sj*-sound, corresponding to /ʃ/ or the *kj*-sound, corresponding to /ç/. As the participants were not necessarily familiar with the phonetic symbols for these sounds, the orthographic representations ⟨sj⟩ and ⟨kj⟩ were used when displaying the two options on the screen (see Section 2.1.3).

To choose between the two sounds, participants were told to press the S key on the laptop keyboard for the *kj*-sound and the L key for the *sj*-sound. Participants could also press the spacebar to replay a word. This option was included in case participants did not hear a given word, either because of noise from the environment or because they lost focus. Participants were informed that this option was intended for cases such as the ones mentioned, rather than to repeat every word they were unsure of.<sup>4</sup>

Participants wore Koss UR40 over ear headphones during the experiment, and they were told to adjust the volume to a comfortable level. The experiment lasted approximately 25 minutes.

### 3.2. Statistical analysis

A first point of interest when analyzing the results of the perception study is the level of correct identification of the two fricatives among listeners, as it allows us to gain an understanding of whether the contrast between /ʃ/ and /ç/ is well perceived. This can in turn inform us about the degree of merging among the speakers in our sample.

To examine the level of correct identification, a Bayesian logistic regression model was fitted. As the response variable, the model included the variable *asIntended*, indicating whether, for a given token, listeners' choice of fricative in the identification task corresponded with the intended fricative—that is, the fricative that prompted the speaker in the production study. The intended fricative, the following vowel, and the interaction between them, were entered as fixed effects in the model. As random effects, we included random intercepts for speaker and listener, along with random slopes for the intended fricative by speaker and listener. The model thus had the following structure:

$$asIntended \sim fricative*vowel + (1 + fricative|speaker) + (1 + fricative|listener)$$

In addition to the identification model, the results of the perceptual identification task were entered into a statistical analysis of listeners' use of the acoustic parameters analyzed in Section 2. Specifically, this analysis modeled listeners' choice of fricative in perception as a function of different acoustic parameters, with the aim of identifying which acoustic parameters serve as the most important perceptual cues to the contrast between /ʃ/ and /ç/.

As in the analysis of the production data in Section 2.2.3, the analysis was carried out using a set of Bayesian mixed models. In the current analysis, the question of interest is how much each acoustic parameter affects listeners' choice of fricative in the identification task. That is, rather

---

<sup>4</sup> Given the length of the experiment (approximately 25 minutes), it is our impression that participants did not want to make their participation last longer than it needed to, suggesting that they did not use the replay option excessively.

than looking into the intended fricative or the accuracy of listeners, we are only concerned with whether listeners responded /ʃ/ or /ç/ for a given production.

Separate models were run for each acoustic parameter, as was the case in Section 2.2.3. Contrary to the analysis of the production data, however, the analysis of the perception data used each acoustic parameter as a predictor variable rather than the response variable. In each model the response variable was listeners' choice of fricative, /ʃ/ being coded as 0 and /ç/ being coded as 1. We therefore fitted a set of logistic regression models, and we entered the given acoustic parameter, the following vowel, and the interaction between them as fixed effects. Each model included random intercepts by speaker and listener, as well as varying slopes for the relevant acoustic parameter by both speakers and listeners. The model structure was as follows:

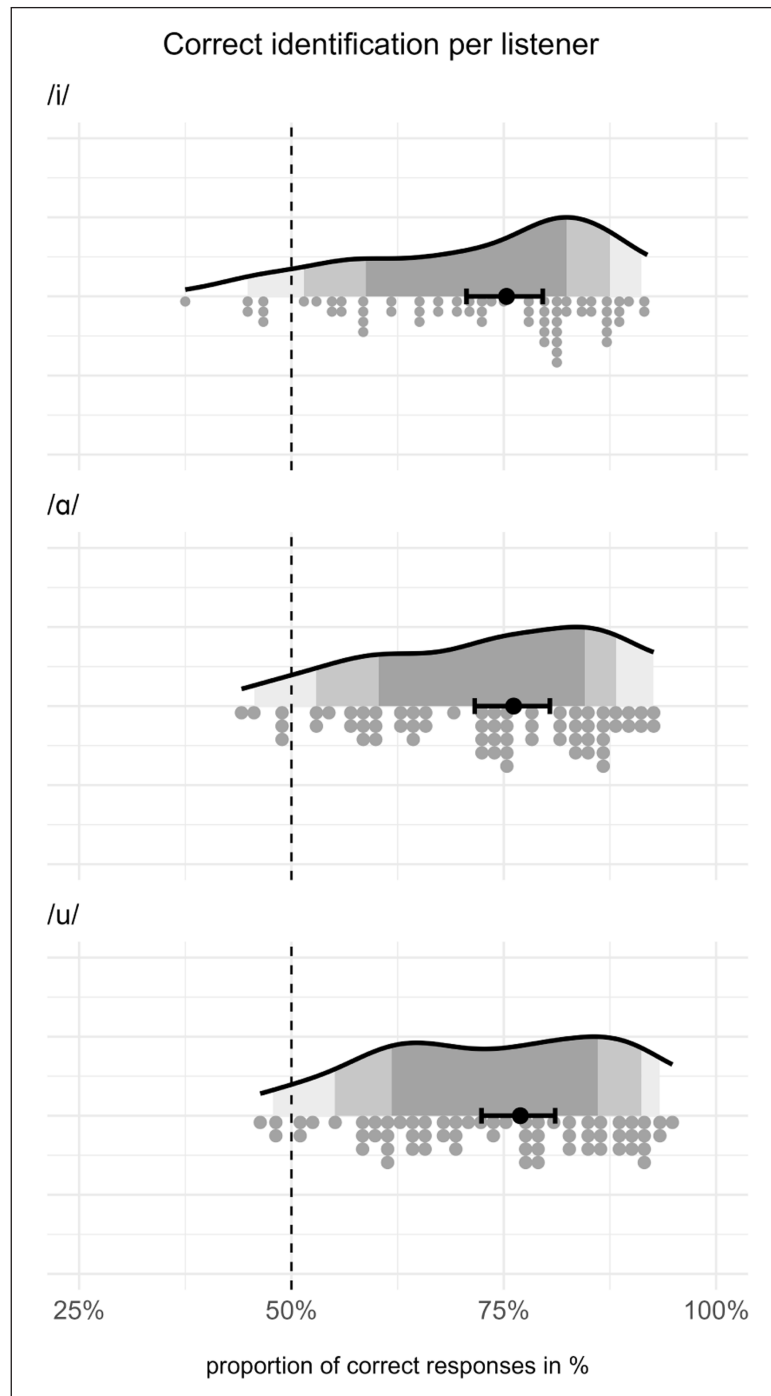
$$response \sim acoustic\ parameter * vowel + (1 + acoustic\ parameter | speaker) + (1 + acoustic\ parameter | listener)$$

All models were identical to the Bayesian regression models fitted in Section 2.2.3 with regards to the number of MCMC chains, iterations, and posterior samples used for inference. Convergence of the models was assessed in the same way as in Section 2.2.3. The models used regularizing, weakly informative priors (Gelman et al., 2017). For regression coefficients, we used a normally distributed prior with a mean of 0 and a standard deviation of 1. For the standard deviation of the random intercepts and slopes and the residual standard deviation, we used a Cauchy distributed prior with a mean of 0 and a standard deviation of 0.5. The prior for the intercept was normally distributed, with mean 0 and standard deviation 1.

### 3.3. Results

#### 3.3.1. Identification

Starting with the level of correct identification, **Figure 5** illustrates the distributions of listeners' identification scores. That is, the proportion of correctly identified tokens for each listener, for each of the three vowel contexts. The error bars represent the posterior means and the 95% Credible Intervals for correct identification, as estimated in the logistic regression model described in Section 3.2. The posterior means and 95% Credible Intervals in the /i/ context (0.75 [0.71, 0.80]), the /a/ context (0.76 [0.72, 0.80]), and the /u/ context (0.77 [0.72, 0.81]) indicate that, based on the model, the data, and the priors, it is plausible that listeners as a population were above chance in identifying speakers' productions in all vowel contexts. Furthermore, as the 95% Credible Intervals largely overlap for the three vowel contexts, it is plausible that there was no difference in identification of /ʃ/ and /ç/ depending on the following vowel.



**Figure 5:** Proportions of correctly identified responses for listeners as a population in the three vowel contexts /i/, /a/, and /u/. Individual dots are listener averages. Density represents smoothed kernel density of these averages. Shading corresponds to the estimated middle 50 / 80 / 95% of the distributions. Point corresponds to posterior means. Error bars correspond to the 95% Credible Intervals around the posterior means.



Overall, then, /ʃ/ and /ç/ are well separated in perception. Note, however, that the distributions overlap with an area around chance performance. In other words, while the majority of listeners perceived the fricatives above chance, the lower tails of the distributions contain listeners who appear to have guessed which fricative to respond, indicating that they did not accurately perceive the fricatives.

### 3.3.2. Accuracy

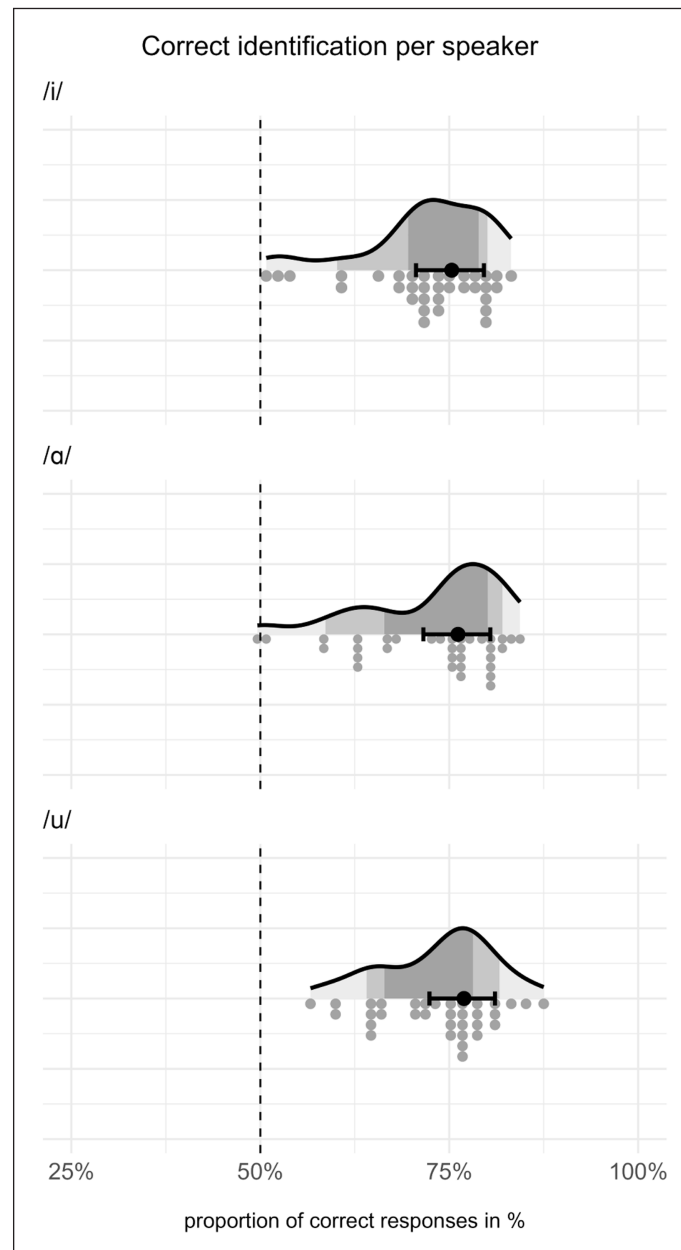
We now turn to an investigation of the level of correct identification by speaker, a measure we will refer to as *accuracy*. This measure demonstrates to what extent listeners perceived intended /ʃ/ as /ʃ/ and intended /ç/ as /ç/ for a given speaker, giving an indication as to the degree of merging among speakers as seen from a perceptual perspective. Imagine, for example, that listeners correctly identified 75% of a speaker's productions of /ʃ/ and 70% of the same speaker's productions of /ç/. This speaker's overall accuracy score would be the mean of these two scores, namely 72.5%. Speakers with high accuracy scores of 90%, for example, are likely to have produced a contrast, as the level of accuracy is high for both fricatives. Speakers with accuracy scores around 50%, on the other hand, most likely have not produced a perceptible contrast. This score could be the result of different underlying patterns.

One type of speaker could have a high accuracy score for one of the fricatives and a low accuracy score for the other. For instance, if, for a given speaker, listeners identified intended /ʃ/ as /ʃ/ 80% of the time, but in contrast identified intended /ç/ as /ç/ 20% of the time, the total accuracy score for this speaker would be 50%. This score reflects a scenario where the speaker has a merged pronunciation, and the result of the merger is /ʃ/. That is, intended /ʃ/ is perceived as /ʃ/, and intended /ç/ is also perceived as /ʃ/.

Another imaginable scenario is that a speaker has an overall accuracy score of 50%, just as in the previous scenario, but that the accuracy score for both fricatives is around 50%. In this case, listeners were at chance when identifying whether the speaker produced /ʃ/ or /ç/, indicating that the speaker's productions of both fricatives were not easily categorized as either /ʃ/ or /ç/. This speaker is likely to have a merged pronunciation, but the result of the merger might be somewhere in between the two fricatives.

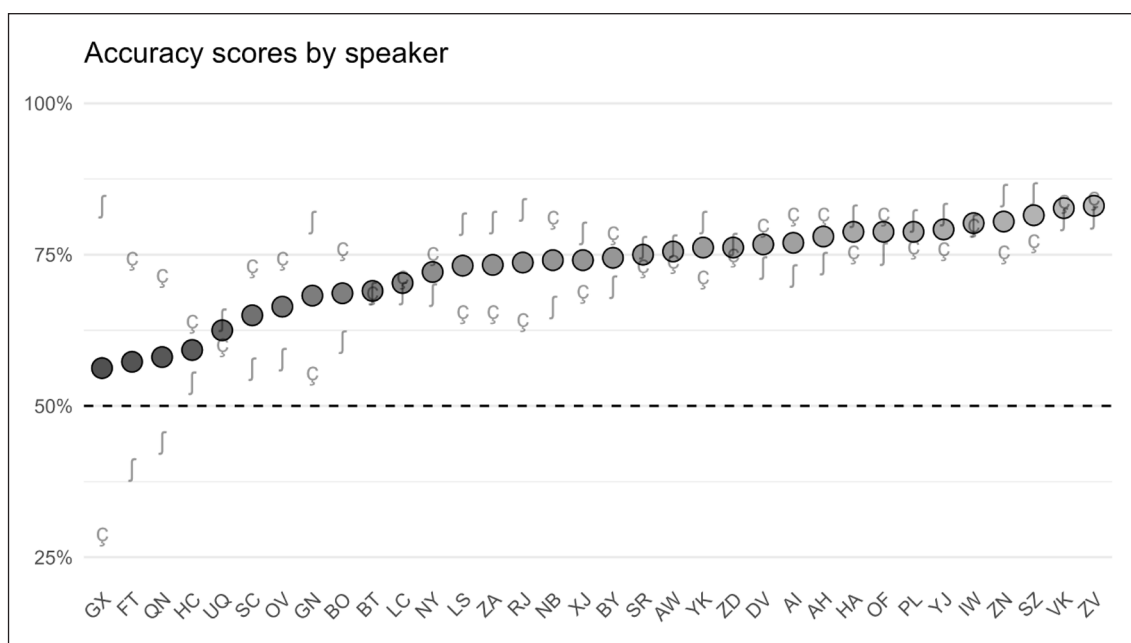
**Figure 6** illustrates the distributions of accuracy scores for the population of speakers, separated by vowel context. As in **Figure 5**, the error bars indicate the posterior means and 95% Credible Intervals for correct identification based on the Bayesian logistic regression model described in Section 3.2 (0.75 [0.71, 0.80] in the /i/ context, 0.76 [0.72, 0.80] in the /a/ context, and 0.77 [0.72, 0.81] in the /u/ context). It becomes clear that a majority of speakers have accuracy scores well above 50% in all vowel contexts, indicating that they produced a perceivable contrast between /ʃ/ and /ç/. A smaller number of speakers, however, are found in

an area around 50%, indicating that listeners did not reliably perceive a contrast between these speakers' fricative productions.



**Figure 6:** Proportion of correctly identified responses for speakers as a population, separated by vowel context. Individual dots are speaker averages. Density represents smoothed kernel density of these averages. Shading corresponds to the estimated middle 50 / 80 / 95% of that distribution. Point corresponds to posterior mean. Error bar corresponds to the 95% Credible Interval around the posterior mean.

Furthermore, **Figure 7** presents the accuracy scores of individual speakers, collapsed across vowel contexts as the Bayesian logistic regression model, and **Figures 5** and **6** all indicated that there were no notable differences in the degree of correct identification of /ʃ/ and /ç/ depending on the following vowel. **Figure 7** illustrates that speakers form a continuum, from scores around 55% to scores above 80%. It is clear from the distribution that there is no natural cutoff point between speakers which would allow us to single out separate groups. Moreover, **Figure 7** also presents speakers' accuracy scores for the two fricatives separately, and it is clear that among the speakers with the highest overall accuracy scores, there does not appear to be a notable difference between /ʃ/ and /ç/. Among the speakers with the lowest accuracy scores, we find both types of speakers described above. Looking at the five speakers with the lowest scores, GX, FT, and QN have a relatively high score for one fricative and a low score for the other, while HC and UQ have relatively low scores for both fricatives. Note that GX's scores indicate a merged pronunciation in the direction of /ʃ/, while FT and QN's scores are indicative of a merged pronunciation in the direction of /ç/.



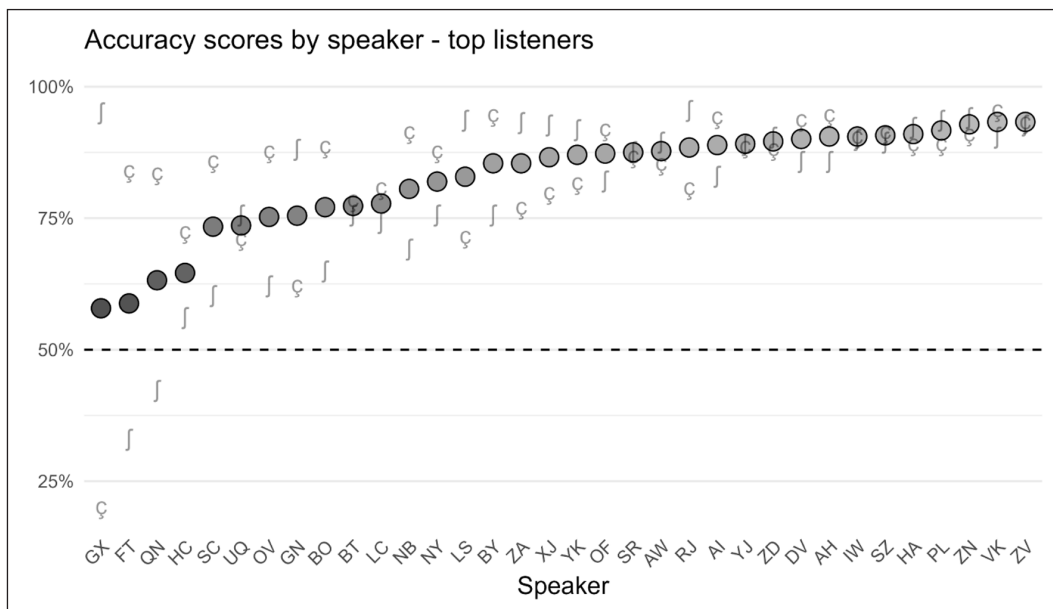
**Figure 7:** Proportion of correctly identified responses for each speaker. Proportions for the individual fricatives are indicated by their IPA symbols.

**Figure 7** also shows that no speaker has a higher overall accuracy score than 83%. If we take the findings in the identification analysis in Section 3.3.1 into account, it is reasonable to assume that speakers' scores are affected by the fact that some listeners were unable to correctly identify the fricatives (see **Figure 5**). These listeners were most likely at chance performance for

all speakers, and their effect on the accuracy scores is therefore expected to be the same across speakers. It is probable, then, that accuracy scores would be higher overall if it were not for these listeners.

We therefore carried out a descriptive analysis of the accuracy scores for individual speakers based only on the perception of the listeners who most likely were able to perceive the contrast between /ʃ/ and /ç/. We used 75% correct identification of fricative productions as a pragmatic threshold for having perceived the contrast (see Treutwein, 1995; Leek, 2001; Lesmes et al., 2015, for references to the use of 75% as a target performance level for the identification of stimuli in psychological and forced-choice experiments). This selection resulted in 34 listeners, who will be referred to as the top listeners.

As illustrated in **Figure 8**, accuracy scores are markedly higher when only top listeners are included. A majority of speakers have scores above 75%, and the highest ranked speakers have scores up to 93%. As predicted, then, the accuracy scores in **Figure 7** were lowered by the inclusion of listeners who did not reliably perceive the contrast between /ʃ/ and /ç/. Note, however, that the increase in accuracy scores in **Figure 8** compared to **Figure 7** is clearly smaller for the four lowest ranked speakers than for the rest of the speakers. The two speakers at the bottom, GX and FT, hardly improve at all, while the next two speakers in the ranking, QN and HC, show only slight improvements. The next two speakers, UQ and SC, on the other hand, show considerable improvements, along with the rest of the speakers.



**Figure 8:** Proportion of correctly identified responses for each speaker based on the perception of top listeners only. Proportions for the individual fricatives are indicated by their IPA symbols.

The comparison between **Figures 7** and **8** thus indicates that the four speakers at the bottom of the ranking likely did not produce a contrast between /ʃ/ and /ç/. That is, if these speakers in fact produce no difference between the two fricatives, their accuracy scores should not be much higher for top listeners than for all listeners. Even if listeners can reliably perceive the contrast, they cannot perceive a distinction which the speaker did not produce.

As in **Figure 7**, **Figure 8** presents speakers' accuracy scores for the individual fricatives as well as their mean accuracy score. These scores largely corroborate the scores in **Figure 7**.

### 3.3.3. Ranking of cues

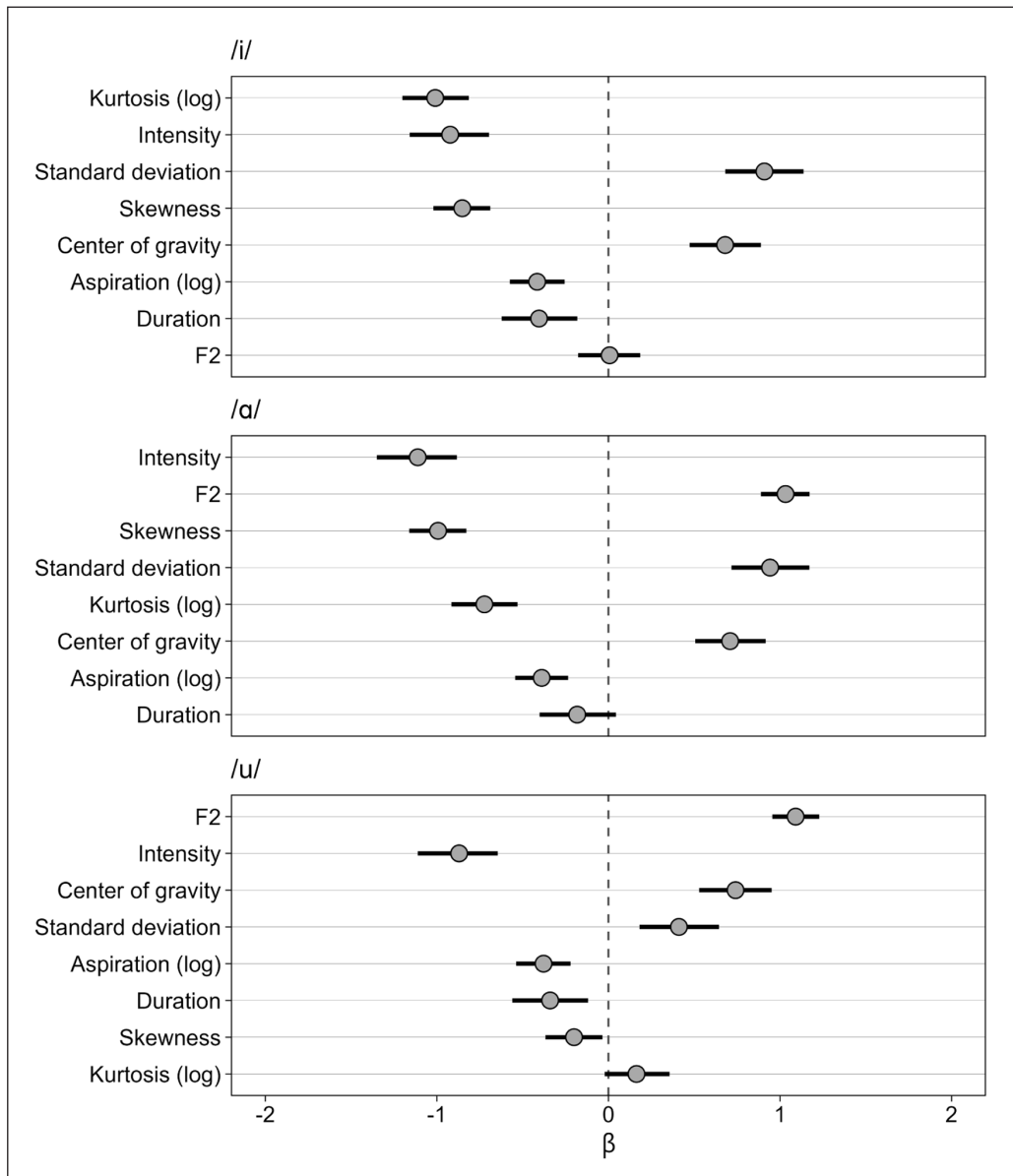
The results of the identification task presented in sections 3.3.1 and 3.3.2 illustrate the extent to which listeners accurately categorized speakers' fricative productions. We now turn to an investigation of which acoustic parameters listeners used when they performed this categorization. By studying the contribution of the different acoustic parameters measured in the production study to the perception of the contrast between /ʃ/ and /ç/, similar to the ranking of parameters performed in Section 2.2.3, we can compare whether speakers and listeners rely on the same parameters.

**Figure 9** presents the log odds of the logistic regression models for each acoustic parameter, separated by vowel context and ranked by size. As was the case in the corresponding analysis in Section 2.2.3, the differences between /ʃ/ and /ç/ were modeled as the change from /ʃ/ to /ç/. Again, positive coefficients indicate that the measurements for /ç/ were higher than those for /ʃ/, and negative coefficients indicate that the measurements for /ʃ/ were higher than those for /ç/. Effect magnitudes are reported alongside their uncertainty by the posterior means and their 95% Credible Interval. Only a subset of the effect magnitudes are reported here, but posterior means and 95% Credible Intervals for all parameters in all vowel contexts can be found in Appendix C.

**Figure 9** illustrates that the fricatives seem to be well separated in perception, as posterior means and 95% Credible Intervals do not include zero as a plausible value for most of the parameters. As in the production study, however, the ranking of parameters shows variation by vowel context. The highest ranked parameter in the /i/ context is kurtosis (-1.01 [-1.20, -0.81]), while in the /a/ context, it is intensity (-1.11 [-1.35, -0.88]), and in the /u/ context F2 (1.09 [0.96, 1.23]). Overall, the /a/ context seems to give the best separation between /ʃ/ and /ç/, while /i/ and /u/ have similar effect magnitudes. These vowel context effects will be discussed further in Section 4.2.

Assessing the overall ranking of the different parameters, intensity appears to be the parameter which gives the largest degree of separation between /ʃ/ and /ç/ in perception, as it is ranked highly for all vowel contexts. Standard deviation, center of gravity, and skewness also have a fairly high ranking across vowel contexts, although skewness has a low ranking before /u/.

Duration and aspiration are consistently ranked towards the bottom of the list, indicating that these parameters do not contribute notably to the separation between /ʃ/ and /ç/ in perception. The remaining parameters are more variable depending on the vowel context. F2 is ranked highly before /a/ and /u/, but it appears to be of no relevance before /i/. Conversely, kurtosis is the highest ranked parameter before /i/, while it is the lowest ranked parameter before /u/. Before /a/, it is found in the middle of the ranking.

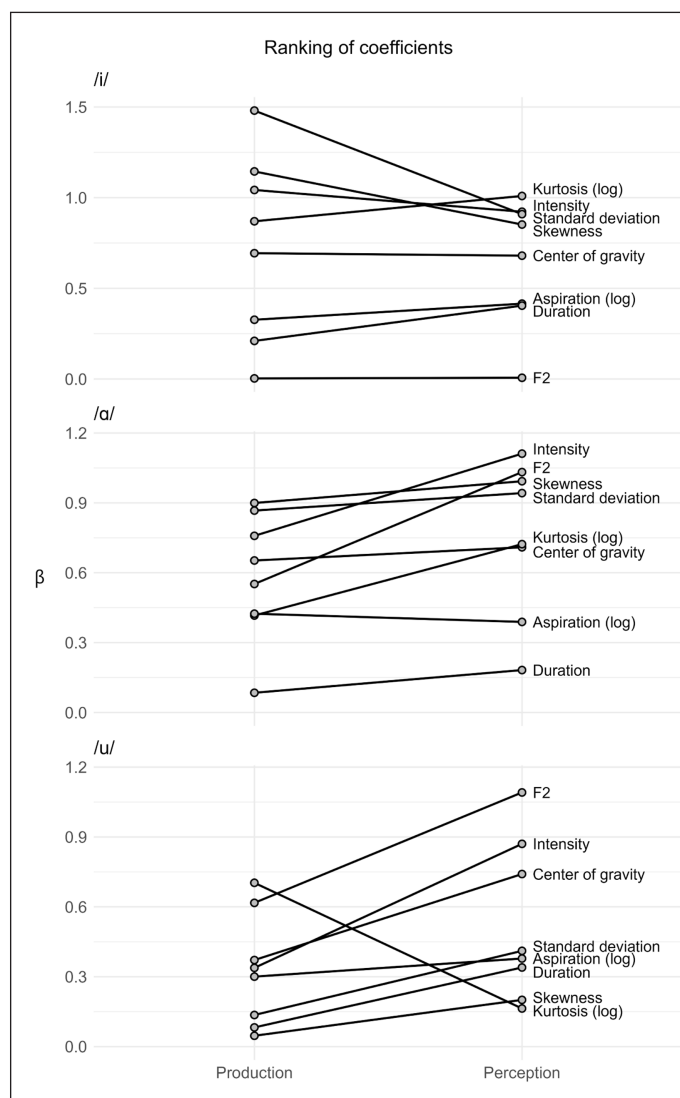


**Figure 9:** Ranking of the coefficients from the individual regression models, indicating the effect magnitudes of the different parameters. Whiskers indicate 95% Credible Intervals.

## 4. Discussion

### 4.1. Correlation between production and perception

The statistical analyses of the production data (Section 2) and the perception data (Section 3) allowed us to rank the acoustic parameters by their importance to both the production and the perception of the contrast between /j/ and /ç/. The rankings in **figures 3** and **9** are summarized for ease of comparison in **Figure 10**. In this figure, the absolute magnitudes of the coefficients are presented along with lines connecting corresponding parameters in production and perception. If the lines cross each other, it indicates that the rankings of the relevant acoustic parameters are different across the two studies.



**Figure 10:** Ranking of the coefficients from the individual regression models in production and perception. To compare the magnitude of the coefficients regardless of their sign, and to make the rankings comparable to **figures 3** and **9**, all coefficients are represented with positive values.

**Figure 10** illustrates that there are both similarities and differences between the rankings for production and perception. Three parameters appear to be relatively important overall, namely intensity, standard deviation, and skewness. Intensity is the parameter which has the highest ranking for the three vowel contexts, /i a u/, while standard deviation and skewness are more important before /i/ and /a/ than before /u/. F2 also appears to be of relative importance in both production and perception, but varies considerably by vowel context, being highly ranked before /a/ and /u/, but at the bottom of the rankings before /i/.

Center of gravity is consistently found in the middle of the rankings for all vowels in both production and perception, indicating that it is of some, but not great, importance to the contrast between /ʃ/ and /ç/. This finding is interesting given that center of gravity is one of the few parameters investigated by van Dommelen (2003, 2019) in the only acoustic analyses of /ʃ/ and /ç/ and the merger in Norwegian. This result thus illustrates that it is necessary to conduct further acoustic analyses of the two fricatives and the outcome of the merger based on the results of the current study.

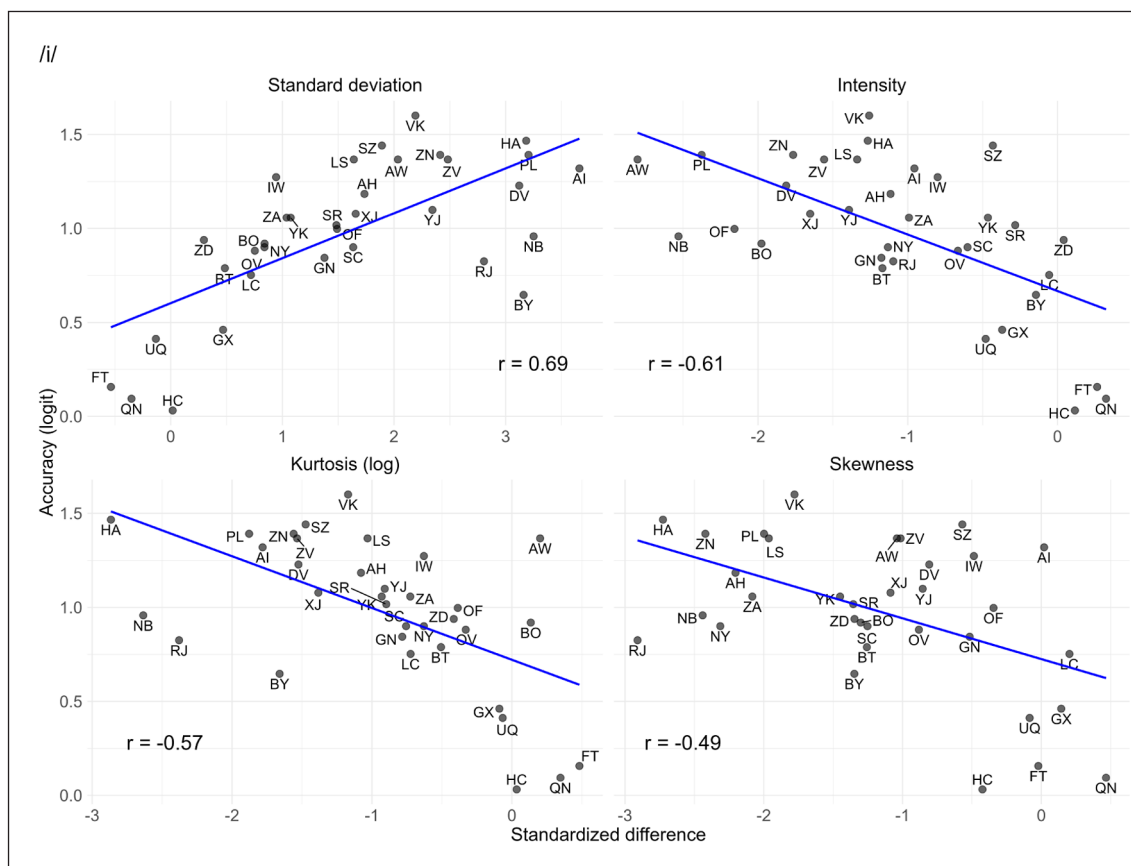
In the bottom of the rankings, we find duration and aspiration. These two parameters are not found to be important for any of the vowel contexts in either production or perception. The remaining parameter, kurtosis, is more variable across vowel contexts and between production and perception. Kurtosis is found to be relatively important before /i/, and of some importance before /a/. Before /u/, however, there is a discrepancy between production and perception, in that kurtosis is found to be the most important parameter in production, but the least important parameter in perception.

It appears, then, that there is a certain degree of overlap between production and perception when it comes to which acoustic parameters are relevant to the contrast between /ʃ/ and /ç/. To further assess the relationship between production and perception, we are also interested in the correlation between speakers' acoustic separation for the different parameters and their accuracy scores in perception. That is, do speakers who produce small or no acoustic differences between /ʃ/ and /ç/ also have low accuracy scores? And conversely, do speakers who produce large acoustic differences also have high accuracy scores?

To quantify the correlation between acoustic separation and perceptual accuracy, Pearson's product moment correlations were conducted. As we have highlighted the importance of vowel context in previous sections, correlations were assessed separately for /i/, /a/, and /u/. **Figure 11** plots speakers' mean difference between /ʃ/ and /ç/ for a selection of acoustic parameters against their accuracy score along with correlation coefficients. Only the four highest correlations overall are presented in **Figure 11**, but correlation coefficients for all acoustic parameters and all vowel contexts can be found in Appendix D. Note that the accuracy scores used in this analysis are based on the perception of all listeners. We also assessed the relationship between acoustics and



perception using the accuracy scores based on top listeners only, but these were not notably different and are therefore not reported.



**Figure 11:** Relationship between acoustic separation and accuracy in perception for the four acoustic parameters most strongly correlated with accuracy, all found in the /i/ vowel context. Acoustic measurements are standardized, and accuracy scores are presented in logits.

**Figure 11** illustrates the relationship between acoustic separation and accuracy in perception for the four parameters which have the strongest correlation with accuracy, all found in the /i/ context. The strongest correlation is found for standard deviation ( $r = 0.69$ ), followed by intensity ( $r = -0.61$ ), kurtosis ( $r = -0.57$ ), and skewness ( $r = -0.49$ ). In the /a/ and /u/ contexts, the highest correlations were found for standard deviation ( $r = 0.45$ ) and intensity ( $r = -0.44$ ). All remaining correlation coefficients were below  $r = 0.40$ .

**Figure 11** also illustrates the relationship between acoustics and perception for individual speakers. Assessing these plots together, it becomes clear that we can identify certain patterns in speaker positions. Most saliently, we find that speakers with the lowest accuracy scores

quite consistently also have the smallest acoustic differences. This is particularly the case for speakers GX, FT, HC, QN, and UQ. For speakers with the highest accuracy scores, there is more variability in the sizes of their acoustic differences. This result is expected, as we know that different speakers can rely on different acoustic cues when signaling a contrast (e.g., Brunelle et al., 2020). However, certain speakers have relatively large acoustic differences for a number of parameters combined with high accuracy scores, as illustrated, for example, with speakers ZN, ZV, HA, and PL.

Overall, we find that having larger acoustic differences to some degree maps onto greater separation in perception, but that this relationship can largely be attributed to the /i/ vowel context. In the /a/ and /u/ contexts, on the other hand, the mapping between acoustic separation and perceptual accuracy is less robust, with correlations consistently below  $r = 0.5$ . There are a number of possible explanations for this finding.

First, it is possible that there are acoustic parameters not measured in the current study, which in reality are the most important cue to a contrast between /ʃ/ and /ç/ in Norwegian. As mentioned in Section 2.2.2, for instance, evidence from acoustic analyses of the contrast between these fricatives in German indicate that DCT coefficients more accurately capture the spectral differences between these phonemes than spectral moments (Jannedy & Weirich, 2017).

Second, speakers and listeners might rely on a combination of parameters in distinguishing between the two fricatives, meaning that no parameter alone is of great importance. We note that several of the most highly ranked parameters in both production and perception, such as standard deviation, skewness, and kurtosis, are highly correlated, as indicated in **Figure 1**. It is possible, then, that speakers and listeners rely in large part on the interactions between these parameters. If so, these combined cues will not be accurately represented by the method employed in the current study, where each parameter is assessed separately.

Finally, it is also possible that the way we normalized the acoustic parameters does not map well onto the way listeners normalize speech in perception. That is, acoustic parameters can be normalized using different methods, and these methods do not necessarily provide equally accurate mappings onto perception (Persson & Jaeger, 2023). In the current study, acoustic parameters were normalized using z-scores, but it is possible that another normalizing account would more closely mirror the perceptual normalization strategy used by listeners, potentially leading to a more robust relationship between production and perception.

## 4.2. Vowel context

An important finding in both the production study and the perception study is the large effect of vowel context on the production and perception of the contrast between /ʃ/ and /ç/. This finding is corroborated in the analysis of the correlation between production and perception

in Section 4.1. Our findings clearly show that the rounded vowel /u/ has a different effect on the production and perception of the preceding fricative than the non-rounded vowels /i/ and /a/. First and foremost, this effect is apparent in that the effect sizes of the different acoustic parameters are smaller before /u/ than before the other vowels. In other words, /ʃ/ and /ç/ are more similar before /u/ than before /i/ and /a/, and this is particularly the case in production.

A possible explanation for this finding is the rounding of the vowel /u/. The fricative /ʃ/ has been described as rounded in Norwegian, whereas /ç/ is described as unrounded (Sivertsen, 1967, p. 79; Vanvik, 1979, pp. 39–40). When the following vowel is rounded, however, /ç/ might be anticipatorily rounded, as indicated by research showing that a rounded vowel can have a strong effect on the acoustic characteristics of a preceding fricative (Lulaci et al., 2024). This coarticulation would in turn lead to the loss of rounding as a cue to the contrast between /ʃ/ and /ç/. Because lip rounding has an effect on acoustic parameters such as center of gravity (Nittrouer et al., 1989), we can hypothesize that the differences between /ʃ/ and /ç/ for these parameters diminish when both fricatives are rounded, leading to less of a separation.

Another notable difference between the vowel contexts is the ranking of the F2 parameter. Before /u/, F2 is one of the most important cues to the contrast between /ʃ/ and /ç/, both in production and in perception. F2 is an important cue before /a/ as well, although not as important as before /u/. Before /i/, however, F2 appears to be of no importance, found at the bottom of the rankings for both production and perception. Similar results pointing to the importance of vowel context in fricative identification have been found for the contrast between /s/ and /ʃ/ in American English in Nittrouer and Studdert-Kennedy (1987) and Nittrouer (1992). Nittrouer and Studdert-Kennedy investigated the perception of the phoneme boundary between /s/ and /ʃ/ before the vowels /i/ and /u/, and their results indicate that F2 transition effects were smaller before /i/ than before /u/. Nittrouer examined the same effects before the vowels /a/ and /u/, and similarly found that the effect of F2 was smaller before /a/ than before /u/, although the difference was not as large as that between /i/ and /u/.

Nittrouer (1992) argues that this finding can be explained by the fact that the fricative-noise frequencies of the front fricatives /s/ and /ʃ/ are similar to the F2 of the front vowel /i/, whereas they are dissimilar to the F2 of the back vowel /u/. The transition from /s/ and /ʃ/ into /i/ therefore requires little change in F2, both in terms of frequency and in terms of time required to reach the target F2 of the vowel. The change required in both frequency and time to reach the target F2 of /u/, on the other hand, is much larger. Nittrouer also found that the time required to reach /a/ was shorter than the time required to reach /u/, providing an explanation for why the effect size of F2 before /a/ is found in between the effect sizes before /i/ and /u/.

It is plausible that these findings for /s/ and /ʃ/ in English are also applicable to the contrast between /ʃ/ and /ç/ in Norwegian. As /ʃ/ and /ç/ in Norwegian are also front fricatives, their

fricative-noise frequencies are expected to be more similar to the F2 of the front vowel /i/ than the F2 of the back vowel /u/. The rounding of the vowel /u/ has the effect of further lowering F2, which might be responsible for the difference between the back vowels /u/ and /ɑ/. We can thus hypothesize that the change in both frequency and time required to reach the F2 target of a following /i/ is likely to be smaller than the change required to reach the F2 target of a following /ɑ/, which in turn is likely to be smaller than the change required to reach the F2 target of a following /u/. It is possible, then, that the magnitude of the required change from fricative to vowel is an indication of the usefulness of F2 as a cue to the relevant contrast, providing an explanation for the observed ranking of cues in the current production and perception studies.

The results presented here thus suggest that in the study of the production and perception of the contrast between /ʃ/ and /ç/, taking vowel context into account is of great importance, both in the design stage and when conducting analyses and interpreting results. Specifically, researchers should include a range of vowel contexts when gathering data to ensure that analyses can be carried out both across vowel contexts and separately. Only then is it possible to accurately describe the fricatives and the contrast between them.

### **4.3. Is it possible to identify MERGED and DISTINCT speakers?**

In Section 1, we presented the motivation for conducting the current production and perception studies, namely to assess to what extent it is possible and useful to use acoustics and perception to group speakers into MERGED and DISTINCT speakers in a situation of an ongoing merger. We discussed a number of studies that have used acoustics and perception to investigate the merger of /ʃ/ and /ç/ in Norwegian and German, and the sibilant merger in Taiwan Mandarin. However, we argued that these studies have certain methodological shortcomings that lead to arbitrary boundaries between groups.

In the current study, our results indicate that neither acoustics nor perception allows us to group speakers without making arbitrary decisions ourselves. The distributions of speakers' differences between /ʃ/ and /ç/ for a range of acoustic parameters and their accuracy scores in perception indicate that thinking about speakers in terms of two separate groups is a misconception. Rather, what we find is a continuum of speakers whose acoustic differences between /ʃ/ and /ç/ range from 0 up to 2 standard deviations, and whose accuracy scores as determined by listeners range from 56% to 83% (or from 58% to 93%, as determined by the top listeners, see Section 3.3.2). The majority of speakers are found somewhere in between these tails of the distributions.

There is a degree of overlap between speakers who have small acoustic differences for the most important acoustic parameters and speakers who have low accuracy scores, and similarly between speakers who have larger acoustic differences and higher accuracy scores. It is possible, then, that these speakers are representative of a merged and a distinct pronunciation, respectively.

However, it would not be possible to draw a principled line separating these speakers from the rest of the population. Any grouping of speakers would rely on an arbitrary boundary where a certain number of speakers at the bottom and top of the one of the distributions for either some acoustic parameter or accuracy are chosen. It is clear, then, that although acoustics and perception can provide useful insights into the behavior of speakers on a population level, we do not find support in the data for grouping speakers as MERGED and DISTINCT speakers in a non-arbitrary way.

## 5. Conclusion

This study aimed at exploring the extent to which acoustics and perception allow us to identify MERGED and DISTINCT speakers in the ongoing merger of /ʃ/ and /ç/ in Norwegian. Through the use of a production study and a perception study, we investigated whether certain speakers stand out by having smaller and larger acoustic differences between the two fricatives, or by having smaller and larger accuracy scores in perception. Finally, we examined the degree of overlap in the results of the two studies.

Our findings indicate that there is indeed a degree of overlap, such that certain speakers appear to have smaller acoustic differences and smaller accuracy scores, and certain other speakers appear to have larger acoustic differences and larger accuracy scores. However, there was no clear separation of speakers which would allow us to draw a boundary between MERGED and DISTINCT speakers in a nonarbitrary way.

Furthermore, with the exception of the /i/ vowel context, the relationship between speakers' separation in acoustics and their accuracy score in perception was not particularly robust, indicating that listeners might rely mainly on a combination of the cues measured here or different cues altogether. Nevertheless, certain acoustic parameters, namely standard deviation and intensity, showed a notable correlation with perception, and further research on the contrast between /ʃ/ and /ç/ should therefore investigate the possible interactions between these parameters in particular.

The findings presented here are important to the study of ongoing mergers, as they provide an understanding of how speakers are distributed in a population where a merger is taking place. Whereas previous studies have categorized speakers as MERGED or DISTINCT speakers on acoustic or perceptual grounds, we argue that such a categorization is necessarily arbitrary as long as speakers form a continuum with regard to acoustic differences and accuracy in perception.

---

## Additional files

The additional files for this article can be found as follows:

- **Appendix A.** Posterior means and 95% Credible Intervals for each regression model in the production study, separated by vowel context. DOI: <https://doi.org/10.16995/labphon.18881.s1>
- **Appendix B.** Rankings of speakers' mean differences for all acoustic parameters, separated by vowel context. DOI: <https://doi.org/10.16995/labphon.18881.s2>
- **Appendix C.** Posterior means and 95% Credible Intervals for each regression model in the perception study, separated by vowel context. DOI: <https://doi.org/10.16995/labphon.18881.s3>
- **Appendix D.** Correlation coefficients for the correlations between each acoustic parameter and accuracy for each vowel context. DOI: <https://doi.org/10.16995/labphon.18881.s4>

## Acknowledgements

We would like to thank our research assistant Mari Eriksen Nordeng for help with data collection, and Professor James Kirby for providing thorough and valuable feedback on an earlier draft of this paper. We would also like to thank Associate Editor Yao Yao and Co-General Editor Lisa Davidson of *Laboratory Phonology*, and two anonymous reviewers for their constructive and insightful comments, leading to significant improvements to the paper.

## Competing interests

The authors have no competing interests to declare.

## Author contributions

Maria Evjen was responsible for project administration, conceptualization, methodology, data curation, investigation, acoustic analysis, formal analysis, visualization, and writing and editing of the original draft. Timo B. Roettger contributed with data curation, formal analysis, visualization, software, and review and editing of the draft. Sverre Stausland contributed to the conceptualization, methodology, supervision of the study, and review and editing of the draft.

---

## References

Andrésen, B. S. (1980). Palato-alveolarer i Bergensmålet. En begynnende systemisk forandring? [Palato-alveolars in the Bergen dialect. An incipient systemic change?] *Maal og Minne*, 1–2, 88–101.

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* [Computer program]. Retrieved October 6, 2022, from <http://www.praat.org/>
- Brunelle, M., Tân, T. T., Kirby, J., & Giang, Đ. L. (2020). Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology*, 11(1), 1–33. <https://doi.org/10.5334/labphon.278>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Cheng, L. S. P., Babel, M., & Yao, Y. (2022). Production and perception across three Hong Kong Cantonese consonant mergers: Community- and individual-level perspectives. *Laboratory Phonology*, 13(1), 1–54. <https://doi.org/10.16995/labphon.6461>
- Chiu, C., Wei, P.-C., Noguchi, M., & Yamane, N. (2020). Sibilant fricative merging in Taiwan Mandarin: An investigation of tongue postures using ultrasound imaging. *Language and Speech*, 63(4), 877–897. <https://doi.org/10.1177/0023830919896386>
- Clayards, M. (2018). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica*, 75(1), 1–23. <https://doi.org/10.1159/000448809>
- Conrad, F. (2023). Regional differences in the evolution of the merger of /ʃ/ and /ç/ in Luxembourgish. *Journal of the International Phonetic Association*, 53(1), 29–46. <https://doi.org/10.1017/S0025100320000407>
- Dalbakken, L. O. (1996). *Distinksjonen kje/sje i lydendringsperspektiv. En empirisk undersøkelse av barns og unges beherskelse av distinksjonen kje/sje i Trondheim.* [The kje/sje distinction in the perspective of sound change. An empirical investigation of childrens' and young people's mastery of the kje/sje distinction in Trondheim.] [Master's thesis, University of Oslo].
- DiCanio, C. (2013). *Spectral moments of fricative spectra script in Praat.* [https://www.acsu.buffalo.edu/~cdicanio/scripts/Time\\_averaging\\_for\\_fricatives\\_4.0.praat](https://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_4.0.praat)
- Evjen, M., & Stausland, S. (2025). *Analyzing the outcome of the Norwegian merger of /ʃ/ and /ç/.* [In preparation].
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84(1), 115–123. <https://doi.org/10.1121/1.396977>
- Franke, M., & Roettger, T. (2019). Bayesian regression modeling (for factorial designs): A tutorial. <https://doi.org/10.31234/osf.io/cdxv3>
- García, W. E. (2017). *vowelFormants v1.* [https://github.com/wendylviragarcia/vowels/blob/master/analyzes\\_vowels\\_extracts\\_f0\\_f1\\_f2\\_f3\\_f4\\_int\\_dur.praat](https://github.com/wendylviragarcia/vowels/blob/master/analyzes_vowels_extracts_f0_f1_f2_f3_f4_int_dur.praat)
- Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10), 555. <https://doi.org/10.3390/e19100555>
- Hay, J., Drager, K., & Thomas, B. (2013). Using nonsense words to investigate vowel merger. *English Language and Linguistics*, 17(2), 241–269. <https://doi.org/10.1017/S1360674313000026>

- Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4), 458–484. <https://doi.org/10.1016/j.wocn.2005.10.001>
- Irons, T. L. (2007). On the status of low back vowels in Kentucky English: More evidence of merger. *Language Variation and Change*, 19(2), 137–180. <https://doi.org/10.1017/S0954394507070056>
- Jannedy, S., & Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *Journal of the Acoustical Society of America*, 142(1), 395–405. <https://doi.org/10.1121/1.4991347>
- Johannessen, S. H. (1983). Om ‘skjendisar’ og ‘chipsreiarar’. Bruken av sje-lyd og kje-lyd i bergensmålet. [On ‘skjendisar’ and ‘chipsreiarar’. The use of the ‘sje’ sound and the ‘kje’ sound in the dialect of Bergen]. In H. Sandøy (Ed.), *Talemål i Bergen* (pp. 5–28, Vol. 1). Nordisk institutt, Universitetet i Bergen.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108(3), 1252–1263. <https://doi.org/10.1121/1.1288413>
- Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, 27(2), 156–176.
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8), 1279–1292. <https://doi.org/10.3758/BF03194543>
- Lee-Kim, S.-I., & Chou, Y.-C. (2022). Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Laboratory Phonology*, 13(1), 1–36. <https://doi.org/10.16995/labphon.6446>
- Lesmes, L. A., Lu, Z. L., Baek, J., Tran, N., Doshier, B. A., & Albright, T. D. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (d') in Yes-No and forced-choice tasks. *Frontiers in Psychology*, 6, 1070. <https://doi.org/10.3389/fpsyg.2015.01070>
- Lin, Y.-L., & Wu, J.-S. (2023). Sibilant production in Taiwan Mandarin: untangling the effects of linguistic and social variables. In C.-R. Huang, Y. Harada, J.-B. Kim, S. Chen, Y.-Y. Hsu, E. Chersoni, P. A. W. H. Zeng, B. Peng, Y. Li, & J. Li (Eds.), *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation* (pp. 456–463). Association for Computational Linguistics. <https://aclanthology.org/2023.paclic-1.45>
- Lulaci, T., Söderström, P., Tronnier, M., & Roll, M. (2024). Temporal dynamics of coarticulatory cues to prediction. *Frontiers in Psychology*, 15, 1446240. <https://doi.org/10.3389/fpsyg.2024.1446240>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037//1089-2680.2.2.175>
- Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *Journal of the Acoustical Society of America*, 135(5), 2964–2976. <https://doi.org/10.1121/1.4870487>
- Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20(3), 351–382. [https://doi.org/10.1016/S0095-4470\(19\)30639-4](https://doi.org/10.1016/S0095-4470(19)30639-4)



- Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30, 319–329. <https://doi.org/10.1044/jshr.3003.319>
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32(1), 120–132. <https://doi.org/10.1044/jshr.3201.120>
- Papazian, E. (1994). Om sje-lyden i norsk, og ombytinga av den med kje-lyden. [On the ‘sje’ sound in Norwegian and its interchange with the ‘kje’ sound.] *Norskraft*, 83.
- Persson, A., & Jaeger, F. T. (2023). Evaluating normalization accounts against the dense vowel space of Stockholm Swedish. *Frontiers in Psychology*, 14, 1165742. <https://doi.org/10.3389/fpsyg.2023.1165742>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence*, 5, 842546. <https://doi.org/10.3389/frai.2022.842546>
- RStudio Team. (2019). *Rstudio: Integrated development environment for r*. RStudio. <http://www.rstudio.com/>
- Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews. Cognitive Science*, 11(2), e1521. <https://doi.org/10.1002/wcs.1521>
- Shadle, C. H. (2012). Acoustics and aerodynamics of fricatives. In A. C. Cohn, C. Fougerson, M. K. Huffman, & M. E. L. Renwick (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 511–526). Oxford University Press.
- Simonsen, H. G., & Moen, I. (2004). On the distinction between Norwegian /ʃ/ and /ç/ from a phonetic perspective. *Clinical Linguistics & Phonetics*, 18(6–8), 605–620. <https://doi.org/10.1080/02699200410001703664>
- Sivertsen, E. (1967). *Fonologi: Fonetikk og fonemikk for språkstudenter*. [Phonology: Phonetics and phonemics for linguistics students.] Universitetsforlaget.
- Stan Development Team. (2024). *Stan modeling language users guide and reference manual*, 2.32.2. <https://mc-stan.org>
- Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, 185, 1–20. <https://doi.org/10.1016/j.cognition.2019.01.004>
- Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71, 249–267. <https://doi.org/10.1016/j.wocn.2018.09.004>
- Torp, A. (1999). Skarre-r og «skjøttkaker» – barnespråk, talefeil eller språkforandring? [Uvular r and ‘skjøttkaker’ – child language, speech impairment, or language change?] *Det Norske Vitenskapsakademi, Årbok*, 334–358.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35(17), 2503–2522. [https://doi.org/10.1016/0042-6989\(95\)00016-X](https://doi.org/10.1016/0042-6989(95)00016-X)

*Undervisningsplan for Oslo folkeskole. [Teaching plan for Oslo primary schools.]* (1940). H. Aschehough & Co. (W. Nygaard).

van Dommelen, W. A. (2003). An acoustic analysis of Norwegian /ç/ and /ʃ/ as spoken by young people. *Journal of the International Phonetic Association*, 33(2), 131–141. <https://doi.org/10.1017/S0025100303001245>

van Dommelen, W. A. (2019). Is the voiceless palatal fricative disappearing from spoken Norwegian? In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 765–769). Australasian Speech Science; Technology Association Inc.

Vanvik, A. (1979). *Norsk fonetikk: Lydlæren i standard østnorsk supplert med materiale fra dialektene. [Norwegian phonetics: The phonetics of Standard East Norwegian supplemented with materials from Norwegian dialects.]* Universitetet i Oslo, Fonetisk institutt.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147–161. <https://doi.org/10.1016/j.wocn.2018.07.008>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Wikse Barrow, C., Włodarczak, M., Thörn, L., & Heldner, M. (2022). Static and dynamic spectral characteristics of Swedish voiceless fricatives. *The Journal of the Acoustical Society of America*, 152(5), 2588–2600. <https://doi.org/10.1121/10.0014947>

Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. Routledge. <https://doi.org/10.4324/9781315165547>

