University of Amsterdam
Faculty of Humanities

The Social Life of the Fronted [s]: Acoustic Patterns Across Finland

Ella Heinävaara (14233584)
BA thesis Linguistics
University of Amsterdam
Supervised by Cesko Voeten
22.06.2025

Abstract

This thesis investigates the acoustic variation of the Finnish sibilant [s] by measuring center of gravity (CoG) values across speakers from 30 locations in Finland. The study examines whether the perceived *sharpness* of [s], often called *Helsinki [s]*, varies regionally and socially, particularly by gender. This study uses cleaned audio data from a corpus and analyzes [s] segments using Praat and R scripts to compute CoG. A threshold of 6000Hz from previous research (Koivisto, 2022) is used to classify a fronted [s] acoustically. Contrary to prior assumptions, findings show no statistically significant difference in [s] sharpness between men and women in the dataset. The results find that while fronted [s] were present in the dataset, no location had statistically significant fronted [s]. Perceptually, fronted [s]'s were found in the dataset, with some being under the threshold and thus challenging the previously set threshold. Additionally, this research finds that larger municipalities have lower CoG values, challenging the widespread association of urban environments and the fronted [s]. These findings call into question the geographic and social exclusivity of the so-called Helsinki [s].

Table of Contents

1. Introduction

1.1. Finnish /s/

Among the various realizations of /s/, a particularly notable variant of this sound is the fronted [s], which has become the subject of both linguistic and sociolinguistic studies in Finland. The fronted [s], sometimes referred to as the *Helsinki [s]*, *City [s]*, or even *Gay [s]*, is characterized by a more advanced articulation of the sound in the mouth, resulting in a sharper and more sibilant quality. While it is usually associated with Helsinki, this study intends to challenge the idea by studying the regional variation of this fronted [s] realization.

Scholars disagree on the exact articulation of [s] in Finnish. According to Sovijärvi (1979), [s] is pronounced with the tip of the tongue usually placed on the medialveolar ridge. Karlsson (1983, as cited in Koivisto, 2022), in turn, describes the Finnish [s] sound as dental. On the other hand, in Iivonen (2009, as cited in Halonen et al., 2020), [s] is a voiceless predorsal alveolar fricative. Lastly, Suomi et al. (2008, p. 27) state that it can be anywhere between [s] and [ʃ]. This range of descriptions points to the fact that the Finnish [s] is not fixed to a single articulatory category but instead shows intra- and inter-speaker variation, which opens up the possibility for certain variants, such as the fronted [s], to acquire sociolinguistic significance.

Acoustically, the majority of the energy of the Finnish [s] sound typically falls within the range of approximately 3500–4000 Hz (Suomi, 1990, as cited in Koivisto, 2022), or, according to another source, slightly above or below 4000 Hz (Lauttamus, 1981). Niemi et al. (2006, as cited in Koivisto, 2022) found that the energy of the Finnish [s] is concentrated in the range of approximately 4000–5500 Hz. However, this study only included male speakers, whose CoG values tend to be lower than those of female speakers, depending on the surrounding phonetic environment. Karlsson (1983, as cited in Koivisto, 2022) places the energy of the Finnish [s] sound within the broader range of 4000–8000 Hz.

Due to being fronted, the variant is thought to be characterized by a higher spectral concentration, typically above 6000Hz. However, due to the limited number of previous studies and the absence of a unified phonetic definition, this threshold remains a tentative guideline rather than an established criterion. This study adopts the same threshold for comparability. The decision to use this value is talked about in more detail in part 1.5.

1.2. History of the Sound

According to Halonen et al. (2020), the fronted [s] originates from Swedish. Its history dates back to the early 1800s, when Helsinki was appointed the capital of the Grand Duchy of Finland—an autonomous state under the Russian Empire—by Alexander I. At the time, Swedish was the dominant language among the city's residents (Halonen et al., 2020). However, with the rise of nationalism across Europe, the tsar sought to promote Finnish over Swedish, the language of the former rulers. The goal was to establish Finnish as the primary language of administration, education, and business, starting with the capital, in hopes that the rest of the country would follow.

As part of this linguistic shift, the Finnish National Theatre was founded, where Swedish-speaking actors and actresses were required to perform in Finnish. In doing so, they carried over their fronted [s] pronunciation (Halonen et al., 2020)[1]. Interestingly, this feature was primarily adopted by women, though the reasons for this remain unclear. One explanation was presented in Halonen et al. (2020), where during the emergence of the Helsinki [s] phenomenon, as efforts were made to develop a refined Finnish spoken language, women were viewed as the *weak link*, seen as less able than men to uphold an educated or prestigious way of speaking.

Throughout the 1900s, the sharp [s] sound continued to be recorded, particularly among Schlager singers. Many of these singers happened to be native Swedish speakers and used a fronted [s], which became part of their recognizable style. As Finnish (female) musicians began imitating these popular performers, they adopted the fronted [s]. In fact, a satirical cartoon humorously claimed that one couldn't be a Schlager singer without this *speech defect* (Vaattovaara & Halonen, 2024).

The sharp [s] sound did not disappear in the 20th century but has persisted into the 21st century, where it remains primarily associated with women, particularly with the so-called *pissis*

[1] According to Riad (2013), the /s/ in Swedish is either (apico)dental or (predorso)alveolar. The current author does not believe that the Finnish variant is particularly dental or fronted today. However, over a century ago, the realization of /s/ could plausibly been dental, since the fronting supposes a dental or similarly advanced articulation.

girls, young women from subcultures often perceived as embodying exaggerated femininity, often considered the Finnish equivalent of Californian *valley girls* (Koivisto, 2022). Additionally, this variant has become linked to the gay community, earning the nickname *gay [s]*. Over time, the sharp [s] has been increasingly recognized as a sociolinguistic marker, reflecting both regional identity and broader cultural trends.

It is essential to note that while Finnish historically has only one phonemic sibilant [s], loanwords from Swedish, Russian, and English have introduced the postalveolar [ʃ], as in *shakki/šakki*[2] ('chess') and *shoppailla* ('to shop'). However, this sound remains largely restricted to borrowings and does not function as a distinct phoneme in native Finnish vocabulary. While its use has increased in contemporary speech, particularly among younger and bilingual speakers, it is unlikely to have influenced the sharp [s] variant examined in this study. Due to the data analyzed covering speech up to the year 1979, the presence of [ʃ] is not expected to be a relevant factor in this study. Additionally, some people, especially older speakers, usually use [s] instead of [ʃ] with the loanwords as well (Abondolo & Valijärvi, 2023, p. 152).

1.3. Perception of the Sound

As Koivisto (2022) notes, the media have had a big part in shaping the image of the sound. Different publications in the late 1800s published texts bashing the actresses and people with this fronted variant about their pronunciation, and many satirical magazines also made fun of it (Halonen et al. 2020). At the time, the sound was described as hissy, prissy, sharp, and fronted. This all contributed to the emergence of a linguistic stigma.

Vattovaara & Halonen (2015) conducted a study where they played a sound clip of a famous singer with this fronted [s] as well as some clips for control to people from all around the country and asked them to say what they thought of it. Vaattovaara & Halonen (2021) presented the study in a podcast episode, and in that episode, they played a snippet of the singer with the fronted [s]. Although the audio was not ideal since there was background noise, the current author wonders if the [s] produced by the singer was actually a fronted [s]. They touch on the topic of finding it difficult to find a snippet, especially since it has not been phonetically researched much,

[2] Both forms are generally accepted. The former is more common and preferred by media and chess players (Nurmi, 2004)

nor was there a clear definition for the sound. Despite the challenges and the current author's own reservations, they managed to find interesting results. They did not directly concentrate the focus on the [s], but most people picked it out of the audio. Most people associate the [s] with Helsinki, but sometimes with the capital region or more generally with the southern part of the country. It is also interesting that a man from Helsinki claimed that the woman sounded like she was from Espoo, a city right next to Helsinki. It is as if one does not want to associate oneself with that sound. As a man, he might not want to associate with the sound since he might fear being considered feminine or emasculated. One participant connected the [s] to any bigger city, not necessarily in the South. Other descriptions of the [s] included *sharpness,* "*shushiness*", *uneducated*, and an association with a younger speaker, possibly a teenager.

Interestingly, in their podcast episode, Vaattovaara & Halonen (2021) mentioned that when a participant deemed that the speaker was from Helsinki, they pointed out the [s] even though the [s] was not fronted.

Many other folk linguistics studies have concluded that the fronted [s] is *posh*, *conceited*, *arrogant*, or *irritating* (Mielikäinen and Palander, 2002; Vaattovaara, 2013a, as cited in Halonen & Vaattovaara, 2017).

1.4. Challenging the *Folk Linguistic Fact*: A Response to Halonen & Vaattovaara

Halonen & Vaattovaara (2017, 2021) have stated multiple times that the fronted [s] is not a *linguistic fact*. Instead, they claim that it is a *folk linguistic fact*. Although they do not clearly define the term, the phrase *linguistic fact* is most likely a reference to Labov (1975). In his framework, a linguistic fact is not merely an abstract structural regularity, but a verifiable pattern grounded on observable behavior and shared understanding of a speech community. From this perspective, the assertion that a fronted [s] lacks the status of a linguistic fact is subject to empirical scrutiny.

Labov would first and foremost demand observational evidence. If the claim implies that a fronted [s] is never actually produced, this would be directly refuted by acoustic studies, which can reliably demonstrate the physical existence and range of [s] productions, like in Koivisto

(2022). If speakers consistently produce [s] sounds that are acoustically distinct in terms of their fronting, then its phonetic existence cannot be denied.

Beyond mere phonetic existence, Labov's framework would assess its linguistic significance. He emphasizes that *some differences don't make a difference* in a linguistic sense, but this principle extends beyond phonemic contrast. Even while a fronted [s] does not serve to distinguish word meanings (i.e., it's not a separate phoneme), its patterned variation can still constitute a linguistic fact. If the use of fronted [s] is systemically correlated with social factors such as the speaker's gender, location, or age, then it carries sociolinguistic meaning. This systematic correlation transforms a raw phonetic detail into a socially charged linguistic variable. For Labov, the very fact that a phonetic variant serves to index social identity or group affiliation makes it a central component of a community's phonetic inventory and, therefore, a robust linguistic fact.

The robustness of these socio-indexical associations depends on their consistency, recognizability, and social understanding. While the acoustic fronting of [s] can be objectively measured, its linguistic relevance lies in whether speakers and listeners consistently associate the variant with specific social meanings. In this context, fronted [s] variants have been linked to urban speech, feminity, youth, and queer identity (e.g. Vaattovaara & Halonen, 2017). These associations are reinforced through media representation and metalinguistic commentary, which give rise to shared perceptions, the very perceptions that can be referred to by the term *folk linguistic fact*. But importantly, such perceptions are not necessarily false; they may reflect real linguistic behavior, albeit through a socially filtered lens.

Therefore, labeling fronted [s] a *folk linguistic fact* risks overlooking its empirical grounding and complex social role. While the label may be intended to highlight the role of ideology or popular perception, it does not invalidate the linguistic reality of the feature. A variant can be both a folk linguistic construct and a linguistic fact if it is both perceived and produced in socially meaningful ways. The folk-linguistic awareness of a variant is often an indicator, not a contradiction, of its sociolinguistic salience.

1.5. Research Aim and Hypothesis

These associations raise an important question: Is the fronted [s] truly confined to the southern parts of Finland, or does it appear elsewhere? Are the associations with specific social groups and urban regions, particularly Helsinki, based on production differences, or are they largely a result of cultural associations?

The present study aims to answer the first question, namely, to examine the acoustic properties of the fronted [s] variant across 30 different locations in Finland, comparing these with data from Helsinki. By analyzing the acoustics of speakers from regions around the country, this study investigates whether the fronted [s] is a phenomenon exclusive to Helsinki and its surroundings or whether it is present in other parts of the country. The association with southern Finland is strengthened by Aittokallio's (2002) study, which found evidence of fronted [s] in speakers from Turku, a southern city two hours away from the capital, indicating that the phenomenon may also occur in other parts of southern Finland.

In Aittokallio's study, they studied the acoustic realizations of the [s]'s of girls from three age groups: primary schoolers, middle schoolers, and high schoolers, across three conditions: speaking to a boy, to a girl, and to a teacher. All the participants were chosen based on the fact that some authority at their school, for example, a speech therapist or a special needs teacher, had noticed their fronted [s] realizations. Interestingly, the study found that for primary schoolers, the CoG values are higher when speaking with another girl, whereas for middle schoolers, the CoG values are higher when speaking with a boy or a teacher. For high schoolers, there was no difference between the conditions. The mean CoGs across the conditions for the three age groups were 7700Hz, 6200Hz, and 7100Hz, respectively. These findings illustrate that CoG values associated with fronted [s] can vary significantly depending on social context, age, and potentially other factors, suggesting that fronting is not necessarily uniform even among speakers perceived to have it.

The Helsinki data from *The Longitudinal Corpus of Finnish Spoken in Helsinki* has been analyzed by Koivisto (2022), and their results shall be the comparison point for this study. The study found that the speaker's gender significantly influenced the acoustic sharpness of the long [s] sound, as measured by the Centre of Gravity (CoG) value. Female speakers produced [s] sounds with higher CoG values, indicating a sharper pronunciation compared to male speakers. This supports the idea that the sharp [s] is more associated with femininity, a characteristic historically linked to women's speech.

The threshold for a fronted [s] in Koivisto's study was 6000Hz, and the same threshold shall be applied here to compare the findings. The value was chosen in their paper due to the *normal [s]* being around 4000Hz is notably lower than 6000Hz. Though that logic is not the most sound, 6000Hz seems to be an appropriate value, since in Aittokallio's (2002) study, the mean CoGs for a fronted [s] were between 6200–7700Hz.

Dialectally, Finland is often split into east and west. This split is primarily marked by differences in vocabulary and vowel length, with speakers from each region using distinct words and exhibiting unique vowel duration patterns (Kotus, n.d). However, this dialectal variation has not been attested to extend to the [s] sound. While vocabulary and vowel length differences are significant in many aspects of the language, it is unlikely that they will have a noticeable effect on the sharpness or acoustics of the [s] sound.

Both Helsinki and Turku have notable Swedish-speaking populations to this day, with around 5% of the population in both cities (Tilastokeskus, n.d.) [3]. Given the historical influence of Swedish phonetics on Finnish speech. A secondary question explored in this study is whether the proportion of Swedish-speaking residents in a given location correlates with higher CoG values for [s].

A further line of inquiry is the relationship between population size and fronted [s]. Since the variant is often associated with urban areas, this study explores whether speakers from larger municipalities produce higher CoG values, which could provide insight into the sociolinguistic dynamics of urban speech patterns. If a link between population size and frontedness is found, it could support the view that urbanity plays a role in the spread of this feature.

Building on these findings, this study hypothesizes that the fronted [s] is primarily a feature of Helsinki speech but may also appear in other regions, particularly in southern Finland. However,

[3] Although this amount might seem small, it reflects the national average, as Swedish speakers make up around 5% of the country's total population.

if the fronted [s] is found more widely across Finland, this could challenge the idea that it is exclusively a Helsinki-based phenomenon. Such findings might suggest that the term Helsinki [s], the most commonly used term for the phenomenon, does not fully capture its distribution, potentially reshaping how it is described in future research. Ultimately, the true aim of this study may just be to figure out what to actually call this phenomenon.

2. Methods

2.1. Dataset

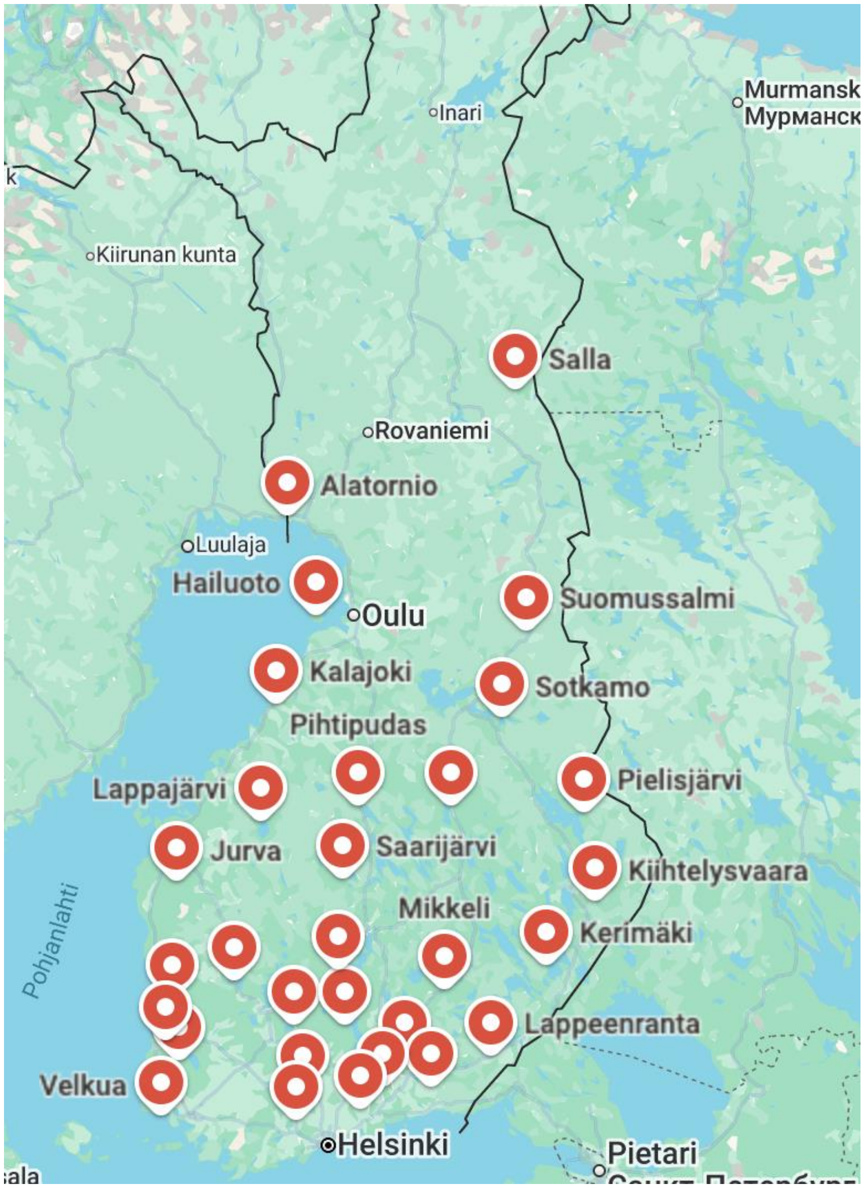
The dataset used is the Samples of Spoken Finnish by the Institute for the Languages of Finland. It is freely downloadable from the website of Kielipankki – Language Bank of Finland. The original dataset contains 50 locations. To keep the workload manageable, only 30 were analysed. For most of the places, there are 2 recordings, all around an hour long, recorded between 1960 and 1979. The recordings consist of semi-structured interviews in which participants are asked to discuss life in their area, providing natural speech patterns representative of regional dialects. These recordings are excerpts from longer recordings, parts of which were not made publicly available. Three locations, Pälkäne, Alatornio, and Salla, only had one usable audio because the other audio was too poor to use.

Table 1 provides the full list of the places in their regions, grouped by region and geographical area. Below, a visual representation of the locations is provided in Map 1.

Table 1: *Municipalities in the dataset by region and geographical area*

Region	Locations (Former Municipalities)	Geographical Area
Kainuu	Suomussalmi, Sotkamo	Eastern Finland
North Ostrobothnia	Hailuoto, Kalajoki	Western Finland
Lapland	Alatornio (now part of Tornio), Salla	Northern Finland
North Karelia	Kiihtelysvaara (now part of Joensuu), Pielisjärvi (now part of Lieksa)	Eastern Finland
South Karelia	Lappee (now part of Lappeenranta)	Southeastern Finland
South Savonia	Mikkeli, Kerimäki (now part of Savonlinna)	Southeastern Finland

Northern Savonia	Lapinlahti	Eastern Finland
Central Ostrobothnia	Pihtipudas	Central Finland
Central Finland	Jämsä, Saarijärvi	Central Finland
Uusimaa	Vihti, Askola	Southern Finland (near Helsinki)
Kanta-Häme	Loppi	Southern Finland
Päijät-Häme	Padasjoki, Artjärvi (now part of Orimattila)	Southern Finland
Kymenlaakso	Sippola (now part of Kouvola), Jaala (now part of Kouvola)	Southeastern Finland
Pirkanmaa	Pälkäne, Ikaalinen	Central Finland
Southwest Finland	Velkua (now part of Naantali), Eurajoki	Western Finland
Satakunta	Noormarkku (now part of Pori), Hinnerjoki (now part of Eura)	Western Finland
South Ostrobothnia	Jurva (now part of Kurikka), Lappajärvi	Western Finland



Map 1: *Locations in dataset on a map*

2.2. Data Processing

The recordings were imported into Praat (Boersma & Weenink, 2025). When the sound included audible static noise or the audio was too quiet, the audio files were cleaned and filtered beforehand using Python. For removal of static noise, a high-pass filter at 50 Hz was applied to remove low-frequency background noise, followed by noise reduction using the noisereduce library with a moderate suppression setting. To eliminate high-frequency artifacts, a low-pass filter at 12,000 Hz was applied. To make the audio louder, its volume was adjusted using a separate function. The audio was loaded via the librosa library, and all volume calculations were performed with NumPy. This function either normalized the audio to its maximum peak amplitude or amplified it by a specific gain in decibels (dB). The decision between normalizing and amplifying was made based on which option provided a more perceptually balanced result. The code used for this preprocessing is provided in Appendix A.1.

To identify each instance of [s], a Praat script was designed to automate the process of aligning segments of speech with linguistic annotations. It takes an audio file and a corresponding annotation file that marks the timing and content of spoken words. For each word, it extracts the relevant portion of audio and uses a speech synthesizer to generate a reference pronunciation. The script then aligns the real speech segment with the synthesized version to produce more detailed annotations. Next, the words with [s] were selected, and only they were saved to save storage. To manage the workload, only the first 500 words with [s] were analysed. The script can be found in Appendix A.2. The alignments were manually checked by the author.

The Centre of Gravity for each selected [s] instance was calculated using a Praat script. This code can be found in Appendix A.3.

Once the CoG values for all 30 locations were obtained, these values were compared across regions. To investigate regional and gender-based variation in more detail, a linear mixed-effects model (LMER) was fitted using the lme4 package (Bates et al., 2015) in R (R Core Team, 2024), with location and gender as fixed effects and speaker as a random intercept to account for multiple [s] tokens per speaker. Prior to model fitting, categorical predictors were sum-coded to facilitate interpretation of fixed effects as deviations from the overall mean.

To facilitate comparison with previous research, all CoG values were centered by subtracting 6000 Hz, following the threshold used by Koivisto (2022) to distinguish fronted [s] realizations. To aid interpretation, estimated marginal means (EMMs) were computed using the emmenas package (Lenth, 2025). These allowed for the calculation of adjusted CoG values by gender (controlling for location) and by location (controlling for gender), as well as pairwise contrasts between gender groups. While the main effect of gender was available from the model coefficient directly due to sum coding, EMMs were used to report adjusted group means and visualize differences more clearly.

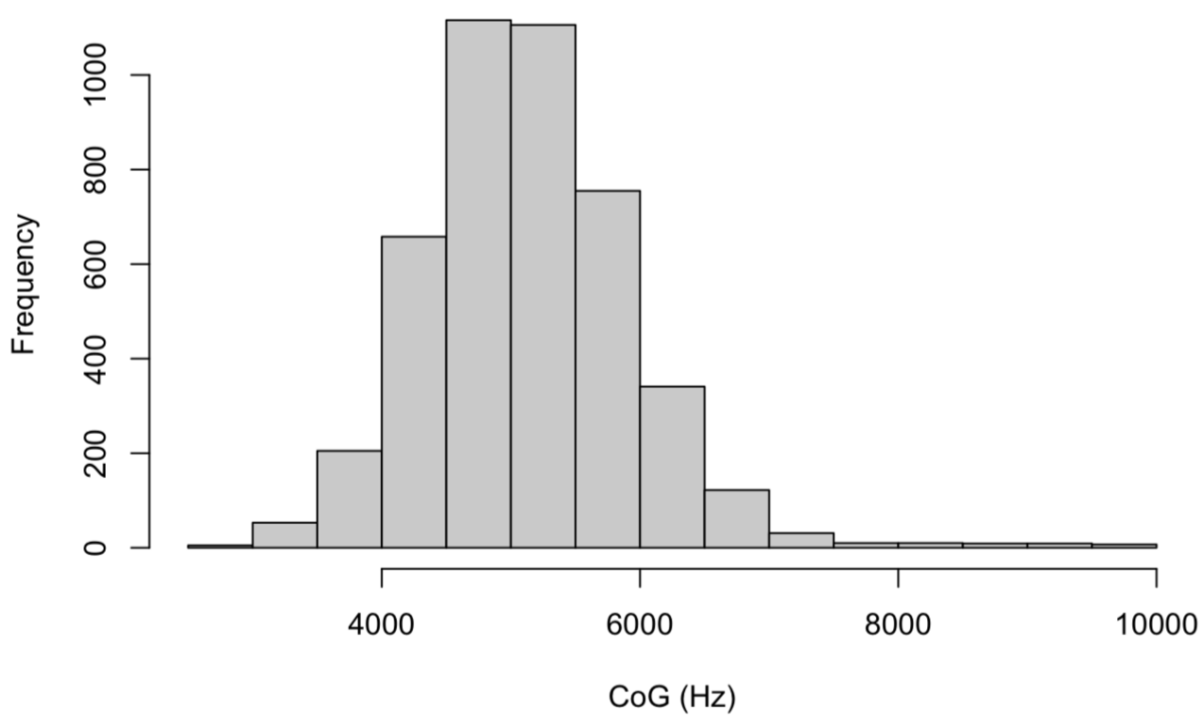
Additionally, to explain the data, one more linear mixed effects model was done with location population, region population, and the percentage of Swedish-speaking population in the location as fixed effects and gender and speaker as random effects. The population data was gathered from *Statistics Finland* (Tilastokeskus, 2023). The population data from the closest year to the recording year was used when possible.

3. Results

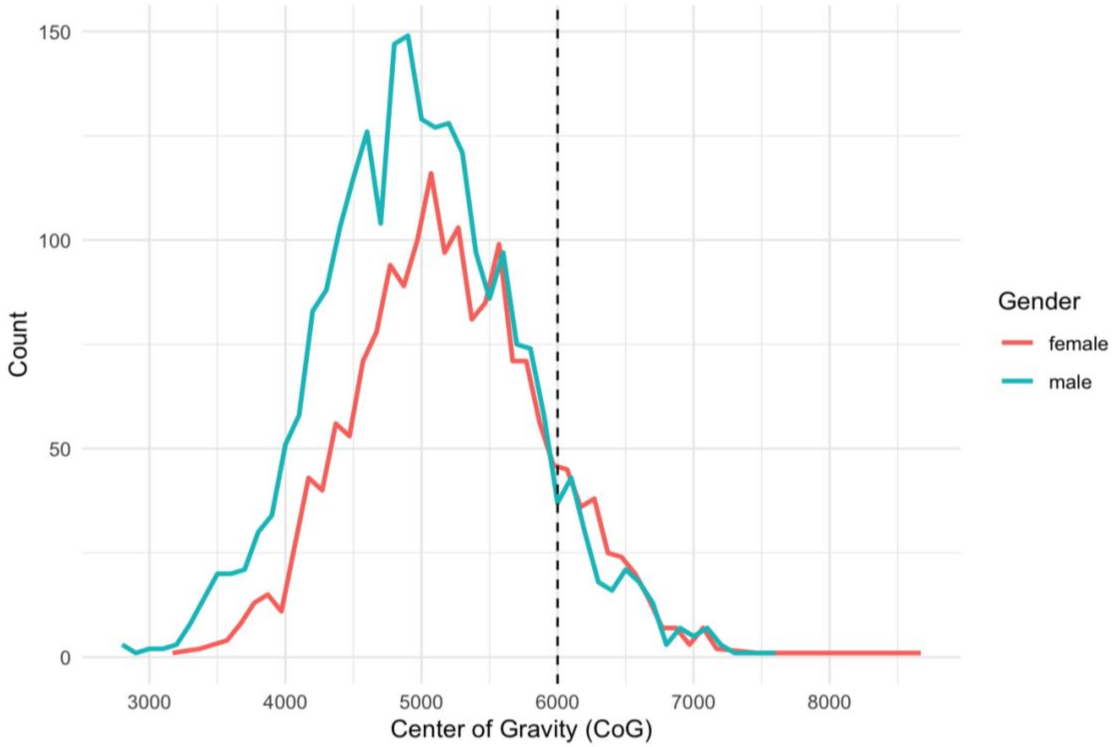
3.1. Overview

In total, 4168 [s] tokens were analyzed across all speakers and locations, with an average of 73.1 [s] per participant. The participant pool included 32 men and 24 women.

Plot 1 shows a histogram that visualizes the distribution of all CoGs in the raw dataset. The distribution of CoG values shows a median of 5065 Hz and a mean of 5095 Hz, with observed values ranging from a minimum of 2751 Hz to a maximum of 8833 Hz. Female speakers produced higher average CoG values (mean = 5223Hz, median = 5193Hz, range 3124–8833Hz) compared to male speakers (mean = 5001 Hz, median 4968Hz, range 2751–7829Hz). These findings are consistent with prior findings suggesting gender-related physiological and articulatory differences influence the spectral characteristics of sibilants (Fuchs & Toda, 2010). 454 values were over 6000Hz, which amounts to just over 10 percent of all the values. Plot 2 visualizes the gender differences. The dotted line shows the 6000Hz mark.



Plot 1: *Histogram of all CoG values*

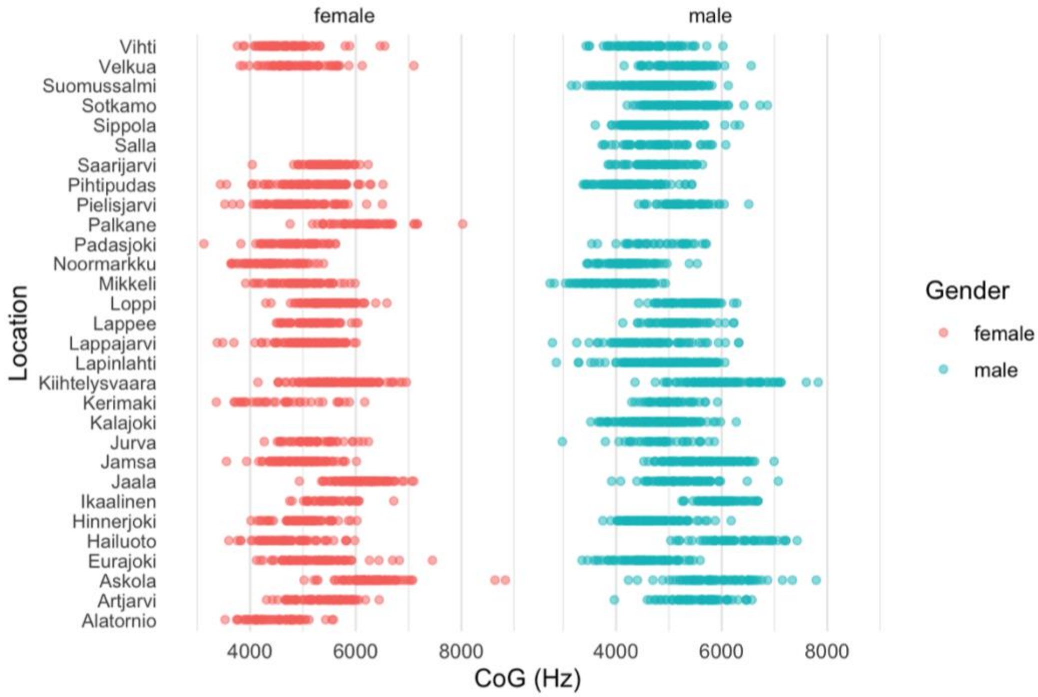


Plot 2: *Distribution of CoG by gender*

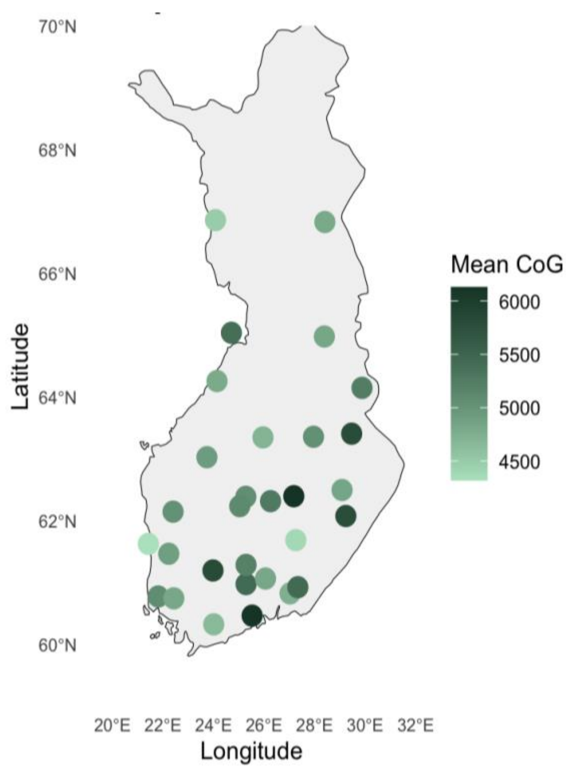
3.2. Regional patterns

Plot 3 visualizes the distribution of CoGs per location by gender. It shows that women and men have similar patterns in the same locations. This suggests that regional variation in [s] sharpness affects both genders in comparable ways, indicating that place of origin plays a stronger role than gender in shaping the acoustic characteristics of [s] across locations.

Plot 4 visualizes the mean CoG on a map. Although there is a lot of variation, no clear pattern can be seen on the map.



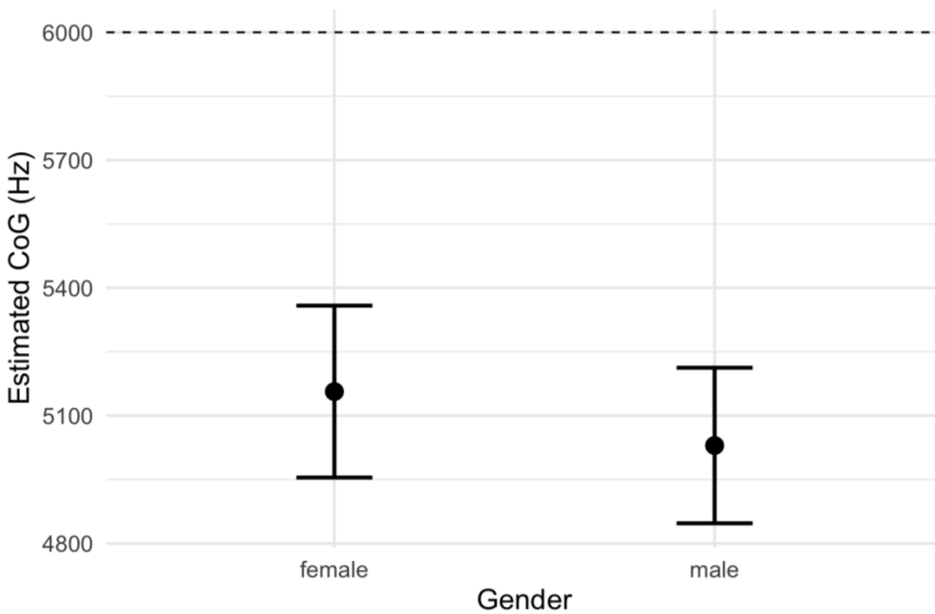
Plot 3: *All CoGs per location by gender*



Plot 4: *Mean CoG per location on a map*

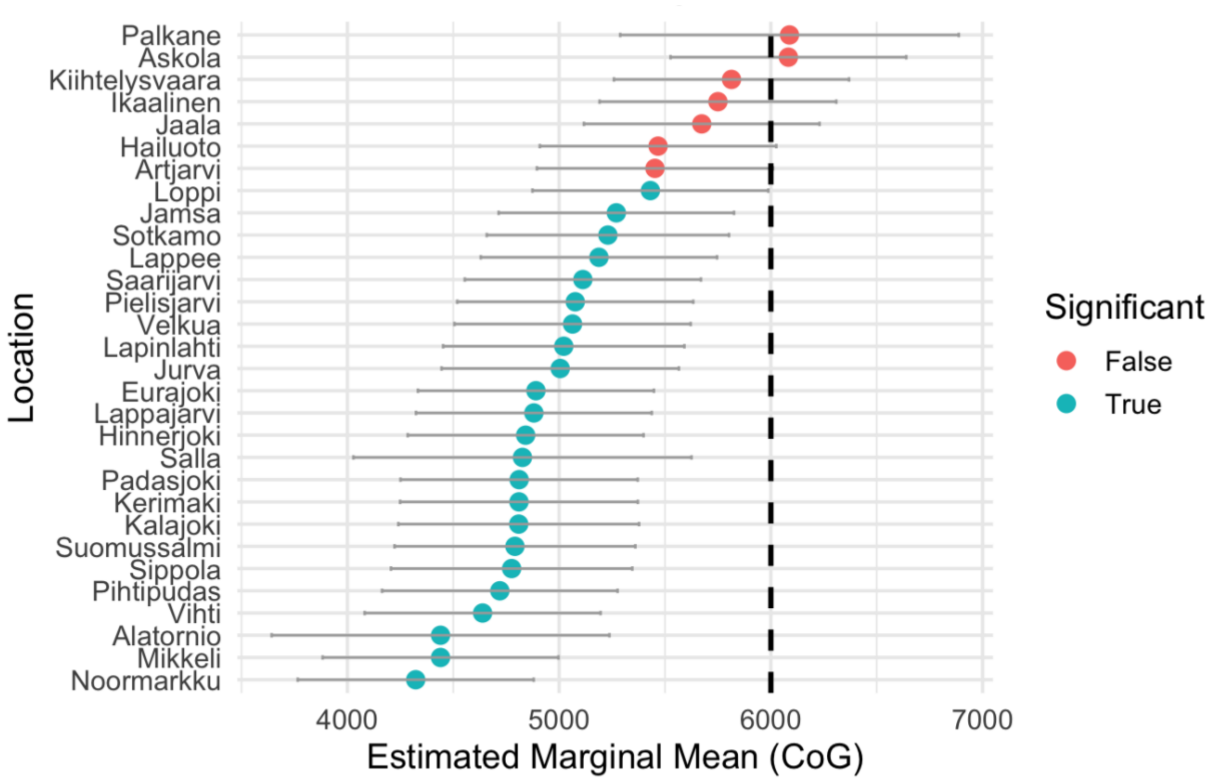
3.3. Statistical testing

Estimated marginal means of the CoG, adjusted for location, indicated that females had a mean-centered CoG of 5131.98 Hz (SE = 87.69), while males had a slightly lower mean-centered CoG of 5049.67 Hz (SE = 75.72). The 95% confidence intervals for females ranged from 4960.11 Hz to 5303.84 Hz, and for males from 4901.25 Hz to 5198.08 Hz. The full data is visible in Appendix B.1. Pairwise contrasts between genders revealed a non-significant difference in CoG values (estimate = 82.31 Hz, SE = 121.30, z = 0.68, p = 0.497), indicating that gender did not have a statistically significant effect on the CoG once location was controlled for. The full data is visible in Appendix B.2. The code can be seen in Appendix C.1.



Plot 7: *Estimated Center of Gravity with SE by gender*

The LMER with EMMS, adjusted for gender, showed that several geographic locations have significantly lower CoG values than 6000Hz. For instance, Noormarkku (4323Hz), Mikkeli (4440Hz), and Vihti (4639Hz) showed among the most retracted [s] variants. As visible in plot 8, only the means of a few locations, namely Askola and Pälkäne, surpassed the 6000Hz threshold, though these estimates were not statistically distinguishable from the reference. The full table can be seen in Appendix B.3.



Plot 8: *Estimated Center of Gravity with confidence intervals by location*

The linear mixed-effects model revealed a significant negative effect of municipality population size on the CoG ($\hat{\beta} = -159.31$, SE = 80.04, p = 0.047), indicating that larger cities have lower CoGs. Region population size ($\hat{\beta} = 111.53$, SE = 78.56, p = 0.156), the percentage of Swedish speakers ($\hat{\beta} = -47.21$, SE = 76.79, p = 0.539), and gender (male; $\hat{\beta} = -123.19$, SE = 119.33, p = 0.302) did not show statistically significant effects. The full data tables can be seen in Appendix B.4. The code can be found in Appendix C.2.

4. Discussion

4.1. Overview of Findings

This study set out to explore whether the fronted [s] variant is exclusive to Helsinki or the southern part of the country, or whether it occurs more broadly across Finland. The results indicate that no region showed a significantly higher CoG than 6000Hz, indicating that fronted [s] is not present in the speech of the locations used in this dataset. While listening through the tokens, some participants definitely exhibited a fronted [s], but it was not provable statistically, meaning that there were not enough tokens over 6000Hz for the average to be statistically higher than 6000Hz.

While the geographic distribution (Plot 4) did not show a clear North-South or East-West gradient or other clear regional divide, some individual locations did stand out. For instance, Askola and Pälkäne had the highest mean CoG values, though not statistically distinct from the 6000 Hz reference. Askola's proximity to Helsinki could explain its high CoG values. However, Pälkäne's elevated CoG is less easily accounted for by geographical influence. There was also only one participant in that location, a female speaker, whose data may have skewed the results, even despite the normalization.

The findings of the latter linear mixed effects model suggest that municipality population has a negative influence on CoG, while broader regional population, Swedish-speaking population, and gender were not significant predictors in this dataset. The influence on CoG is surprising since the fronted [s] has been previously found in the biggest city, Helsinki, and the sixth biggest city, Turku (Kuntaliitto, n.d.). The finding is most likely due to all the locations being quite small and the biggest ones in this dataset not having the fronted variant, rather than a bigger pattern.

Although women produced slightly higher average CoG values than men, consistent with previous literature (e.g., Podesva & Van Hofwegen, 2015), the difference was not statistically significant once location was controlled for. This result is surprising, given that the more fronted variant has previously been associated with women, as discussed earlier and as observed in Koivisto's analysis of Helsinki data. This might be due to the age of the participants. Although the specific age of the participants was not made available, it was evident from the audio that they were older; thus, physiological changes in the vocal tract may influence the acoustic properties of

[s]. According to UT Southwestern Medical Center (n.d.), age-related changes in the vocal tract can include atrophy of muscles and stiffening of connective tissues, causing higher pitch in men and lower pitch in women, as well as general loss of projection and resonance. This physiological change could have influenced this dataset and thus the results. A diachronic acoustic study by Rilliard et al. (2024) of French speakers found a notable lowering of pitch in older female speakers but not in men. While their study focused on pitch rather than CoG, it suggests that age and gender interact in complex ways in the acoustic signal, and such changes might also play a role in shaping [s] production over time. Further research would be necessary to explore this relationship in Finnish data.

It is also possible that the lack of significance can be explained by the relatively small participant pool. Alternatively, there is the possibility that gender-indexed variation is more salient in urban centers, which, unfortunately, were not represented in this dataset. Urban centers have long been identified as key sites of linguistic innovation (e.g., Labov, 2001), where linguistic innovation and social indexing are more likely to flourish due to larger, more diverse populations, increased social mobility, and greater exposure to media and external linguistic influences and thus amplify gender-based differences in speech patterns compared to smaller or more rural communities.

As a post-hoc analysis, normalized CoG values (using Z-score normalization) per participant were analyzed. This was done to rule out the possibility of not getting statistically significant results due to not normalizing. These models showed that gender effects were not significant ($p = 0.5$). None of the locations in the dataset were significant either. The code, the data of gender effects and location effects can be found in Appendices C.3, B.5, and B.6, respectively.

Overall, the results suggest that regional origin plays a stronger role than gender in shaping the CoG of [s]. Plot 3 showed similar CoG patterns for men and women within the same location, indicating that local phonetic norms may override gender-based patterns.

4.2. Comparison to Previous Literature

In Koivisto's study, 28% of the [s] tokens were fronted, or at least above 6000Hz. That is a larger percentage than in this study, which is not surprising since the fronted [s] is associated with Helsinki. Also, it must be noted that Koivisto had a much smaller sample, only around 600 [s] tokens, which could affect the robustness and generalizability of the results. Koivisto's study also found that the speaker's gender had a statistically significant effect on the CoG, with women's CoG being higher, while this study failed to do so.

The values in this study were quite a bit lower on average than the values in Aittokallio's (2002) research. This is not surprising since the participants in their study were young. Although their ages were not made available, the participants were all school age. It is unclear whether Aittokallio accounted for developmental differences in the vocal tract associated with age. They were also all girls, and girls are known to have higher voices than boys (Flipsen et al., 1999, as cited in Fuchs & Toda, 2010). Their values also varied a lot. Across the three age groups and all conditions, the values ranged from 3900Hz to 9300Hz. This means that even with perceived fronted [s], they produced [s]'s with a CoG under 6000Hz.

4.3. Reconsidering the 6000Hz Threshold

The 6000Hz threshold for classifying [s] as fronted, originally chosen by Koivisto, is based on its substantial elevation above the average of 4000Hz. However, this rationale is not empirically robust, and the present study's findings suggest that using a fixed threshold may be overly simplistic.

Snippets of the audios were relistened to, focusing on checking whether the [s] is fronted according to the author. Fronted [s] sounds were perceived in the speech of participants from the following locations:

- Pälkäne (6129Hz)
- Askola (6110Hz)
- Ikaalinen** (5803Hz)

- Kiihtelysvaara* (5797Hz)
- Jaala* (5783Hz)

The numbers in parentheses are the mean CoG values of each location. In the locations with one asterix (*), there was only one participant with the fronted [s], a woman, whereas in the locations with two asterixes (**), the fronted variant was present in the male speaker, not the woman. When comparing these locations to Plot 8, one can see that these are the same top five spots as in the plot. Also, there is a slight gap between the mean CoG values of Jaala and Hailuoto, indicating that the division of fronted and non-fronted might be somewhere there, at least it is for this dataset.

These observations suggest that while fronted [s] is not widespread across the dataset, it is not confined to Helsinki, and individual variation plays a significant role. The perceptual findings also indicate that fronting may occur even when CoG values do not consistently exceed 6000Hz. According to the current data, fronted [s] realizations may occur even at values just above 5500Hz, further questioning the appropriateness of a fixed cut-off point.

This interpretation is consistent with Aittokallio's (2002) findings, where values ranged from 3900Hz to 9300Hz, even among children thought to produce sharp [s] sounds. The wide range highlights the variability inherent in CoG values, even within relatively homogeneous speaker groups. Together, these observations suggest that a more nuanced approach is needed when identifying fronted [s], one that accounts for individual and contextual variation rather than relying on a fixed numerical threshold.

4.4. Limitations

There are multiple limitations to this study. First, the at times small and unbalanced sample size per location; some locations only had 54 instances of [s], some as many as 196. This limited the statistical power to detect regional effects and increased the potential for outliers to influence results, even though each [s] token was manually checked by the author. Additionally, only having at most two participants per location limits the statistical power. There is a chance that most people in that location do indeed have a fronted [s], but the participants in this dataset do not.

Second, the dataset included a limited selection of locations, with no major cities represented. The only relatively large city was Mikkeli, with just over 52000 inhabitants. To properly analyse whether this phenomenon is primarily associated with larger urban areas or with southern regions more generally, future analyses should include data from major cities. Additionally, Northern Finland was underrepresented. Although the northern region of Lappi is more sparsely populated, including larger hubs such as Rovaniemi and Kemi would have strengthened the analysis and would be an interesting addition for further research. If such research is done, the potential influence of Sámi languages, namely Northern Sámi and Inari Sámi, should also be considered.

Third, the age of the recordings must be considered. The data was mostly gathered before the year 1980, was recorded using older recording equipment, and under different technical conditions, which may affect the acoustic measurements. Although steps were taken to normalize the data and clean the audio, it is possible that residual differences in recording quality could impact the measurement of CoG.

To further investigate whether the year of recording could account for variation in CoG across locations, a post hoc linear mixed-effects model was fitted with CoG as the dependent variable, year of recording as the fixed effect, and a random intercept for location ($N = 3912$). The model revealed no significant effect of year on CoG ($\hat{\beta} = 1.87$, $SE = 3.07$, $t = 0.609$). This suggests that, within the time frame of the recordings, there is no clear evidence that the sharpness of [s] changed linearly over time. This supports the conclusion that regional variation, rather than diachronic change, is the more prominent factor in shaping the acoustic realization of [s] in this dataset. The code and full data can be seen in Appendices C.4 and B.7, respectively.

Next, the phonetic environment of the [s] sounds included in the analysis must be considered. It is possible that, in some locations, [s] only occurred adjacent to front vowels, such as [i], which are known to naturally raise the CoG values (e.g., Koivisto, 2022; Lauttamus, 1981). This could result in higher average CoG values that do not necessarily reflect socially meaningful variation but rather the influence of local vowel contexts. To minimize this effect, as much data as possible was collected across different vowel environments. However, it remains a possibility that certain locations included proportionally more [s] sounds next to front vowels than others, which may have affected the results.

Lastly, as mentioned before in the introduction, [ʃ] also exists in the language nowadays, especially in loanwords. While this study did not focus on that phoneme, it is possible that some participants produced sounds closer to the post-alveolar fricative than the alveolar one. Previous research has stated that [ʃ] is usually anywhere between 2500Hz to 4000Hz in Finnish (Lauttamus, 1981). Although some of the values are within that range, while relistening to the audios, no clear instances of [ʃ] emerged.

4.5. Further research

To mitigate these issues in future work, more balanced and larger datasets across locations and speaker demographics are needed. In particular, increasing the number of speakers per location, ideally with both male and female participants and across multiple age groups, would provide a stronger basis for examining sociophonetic patterns with greater reliability. Additionally, collecting or re-recording data under standardized technical conditions could help reduce the confounding influence of recording quality on acoustic measurements.

Including detailed speaker metadata, such as age, gender, education level, and linguistic background, would enable more nuanced analyses of how social factors intersect with regional variation. Additionally, living in the era of social media and the internet, it would also be interesting to research whether that has made the sound more unified around the country. Especially among the younger generation, this increased exposure to national and even global speech patterns through platforms like TikTok, YouTube, and Instagram may lead to the adoption of socially marked variants, such as the fronted [s], or possibly the abandonment of the feature altogether, regardless of regional origin.

Future studies would also benefit from following participant recruitment methods such as those used by Aittokallio (2002), who targeted speakers perceived to produce fronted [s] sounds. This method, while more selective, allows for a more focused examination of the phonetic qualities of the variant. Such targeted sampling could serve as a valuable first step in identifying consistent acoustic characteristics, which in turn could help to establish a more empirically grounded threshold for classifying [s] as fronted. Once clearer acoustic criteria are established, subsequent studies could expand to broader speaker populations to explore the social distribution and variation of the feature in more depth.

The use of index scoring to quantify the degree of [s] fronting across speakers and contexts could be beneficial in future work. Index scores, as used in Stuart-Smith et al. (2020), capture acoustic variation on a continuous scale, reflecting subtle phonetic differences in the [s] sound. This approach allows researchers to correlate production with perception, track sociolinguistic salience, and examine to what extent features like the fronted [s] are meaningful. For example, Levon (2006) employed continuous listener ratings and controlled phonetic manipulation to investigate gendered and sexual identity through [s] variation, demonstrating that indexical meanings are highly context-dependent. Implementing index scoring in Finnish sociophonetic research could enhance the integration of acoustic measurement and social interpretation, particularly in studies exploring stylistic variation and identity construction.

Even though previous research (e.g., Fuchs & Toda, 2010) has found correlations between anatomical features and the acoustic realization of [s] across men and women, these patterns should not be interpreted as straightforward causal relationships. As Zimman (2017) argues, anatomical variation alone does not fully explain the gendered patterns observed in individual productions of [s]. Instead, such patterns are likely shaped by a combination of biological constraints and sociophonetic factors, including identity, style, and social context. Thus, gender differences in [s] realization should not be assumed to result directly from anatomical differences alone. In the context of this study, this issue is likely less relevant, since the data comes from older speakers and from a different time when sociolinguistic norms were most likely more rigid and conformity was socially more expected. Nevertheless, this remains an important consideration for future research, where newer data is used and where greater social acceptance has led to increased stylistic variation linked to identity.

5. Conclusion

The thesis set out to examine the acoustic properties and sociolinguistic significance of the fronted [s] sound in Finnish, often referred to as the *Helsinki [s]*. Through acoustic analysis of speech data from across Finland, we found that while variation in [s] sharpness exists, there is no statistically significant difference between genders. Furthermore, while the sharp [s] has been traditionally associated with the Helsinki area, the present study suggests that this feature is not exclusive to the capital. Despite the lack of statistically significant regional effects, perceptual analysis indicates that sharp [s] variants can be found across the country.

In addition, this study raises questions about the validity of using a fixed threshold to classify [s] as fronted. Several tokens perceived as fronted fell below the threshold of 6000Hz, which suggests that the threshold warrants further investigation.

Given these findings, the label *Helsinki [s]* may be too narrow. While Helsinki has been central to the social meaning associated with this variant, especially through historical ties to theatre, pop culture, and particular identity groups, the data revealed that this [s] is not exclusive to the city. By all accounts, *sharp [s]* or *fronted [s]* would be a more fitting name, since it doesn't restrict the sound to a geographical space or a specific social group.

References

Boersma, Paul & Weenink, David (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.27, retrieved 27 January 2025 from http://www.praat.org/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects models usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic Characteristics of /s/ in Adolescents. *Journal of Speech Language and Hearing Research*, *42*(3), 663–677. https://doi.org/10.1044/jslhr.4203.663

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Żygis (Eds.), Turbulent sounds: An interdisciplinary guide (pp. 281–302). De Gruyter Mouton. https://doi.org/10.1515/9783110226584.281

Google. (n.d.). List of Locations. Google Maps. Retrieved April 19, 2025, from https://www.google.com/maps/@/data=!3m1!4b1!4m3!11m2!2s3Bs87L-lRn-gJhCO4Kc0ww!3e3?entry=tts&g_ep=EgoyMDI1MDUxMy4xKgBIAVAD&skid=3290a99a-39b6-4017-92e0-8086a261e3a7

Halonen, M., Nyström, S., Paunonen, H., & Vaattovaara, J. (2020). *Stadin syntinen s* [The sinful [s] of Helsinki]. Art House.

Halonen, M., & Vaattovaara, J. (2017). Tracing the indexicalization of the notion "Helsinki s." *Linguistics*, *55*(5). https://doi.org/10.1515/ling-2017-0024

Halonen, M. & Vaattovaara, J. (Guest hosts). (2021, November 11) #9: Stadiässä (osa 1) Helsinkiläinen s kuultuna ja luultuna [Audio podcast episode]. *Murremyytin murtajat.* Spotify. https://open.spotify.com/episode/5KlVaCAYWE6GZY164Z6shI?si=e22db146d2e5496c&nd=1&dlsi=f43e70e2f8c94166

Iivonen, A. (1998, January 29). Suomessa "s" voi vaihdella paljonkin [In Finland, "s" can vary quite a lot]. *Helsingin Sanomat*. https://www.hs.fi/mielipide/art-2000003691552.html

Institute for the Languages of Finland. (2023). *Suomen kielen näytteitä, ladattava versio* [Samples of Spoken Finnish, Downloadable Version]. Kielipankki - The Language Bank of Finland. http://urn.fi/urn:nbn:fi:lb-2020112938

Koivisto, E. (2022). *Helsinkiläisen [s]-äänteen foneettinen tarkastelu: Onko "stadilaisella ässällä" foneettista pohjaa?* [A phonetic analysis of the Helsinki [s] sound: Does the "city s" have a phonetic basis?] Helsingin yliopisto. [University of Helsinki] http://hdl.handle.net/10138/350123

Kotimaisten kielten keskus. (Kotus) (n.d.). *Suomen murteet*. Retrieved April 5, 2025, from https://kotus.fi/sanakirjat/suomen-murteiden-sanakirja/suomen-murteet/

Kuntaliitto. (n.d.). *Kaupunkien ja kuntien lukumäärät ja väestötiedot* [Number and population data of cities and municipalities]. https://www.kuntaliitto.fi/kuntaliitto/tietotuotteet-ja-palvelut/kaupunkien-ja-kuntien-lukumaarat-ja-vaestotiedot

Labov, W. (1975). What is a linguistic fact? In *De Gruyter eBooks*. https://doi.org/10.1515/9783110882353

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social factors*. Oxford: Blackwell.

Lauttamus, T. (1981). Havaintoja suomen ja englannin sibilanttien kontekstuaalisesta variaatiosta. [Observations on the contextual variation of Finnish and English sibilants] Virittäjä, 85(4), 340. https://journal.fi/virittaja/article/view/37642

Levon, E. (2006). Hearing Gay: Prosody, Interpretation and the Affective Judgment of Men's Speech. American Speech, 81, 56-78. http://dx.doi.org/10.1215/00031283-2006-003

Lenth, R. V. (2025, May 4). *Estimated Marginal Means, aka Least-Squares Means [R package emmeans version 1.11.1]*. https://cran.r-project.org/web/packages/emmeans/index.html

Nurmi, T. (2004). Kirjoittajan ABC-kortti : Muut merkit. *Kirjoittajan ABC-kortti.* Retrieved June 11, 2025, from https://webcgi.oulu.fi/oykk/abc/kielenhuolto/oikeinkirjoitus/valimerkit/muut_merkit/#hattu

Podesva, R. J., & Van Hofwegen, J. (2015). /s/exuality in Smalltown California. In *Oxford University Press eBooks* (pp. 168–188). https://doi.org/10.1093/acprof:oso/9780190210366.003.0009

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Riad, T. (2013). The Phonology of Swedish. In *Oxford University Press eBooks*. https://doi.org/10.1093/acprof:oso/9780199543571.001.0001

Rilliard, A., Doukhan, D., Uro, R., & Devauchelle, S. (2024). Evolution of voices in French audiovisual media across genders and age in a diachronic perspective. arXiv preprint arXiv:2404.16104.

Sovijärvi, A. (1979). *Suomen kielen äännekuvasto* [The phonetic system of the Finnish language] (2nd ed.). Gummerus.

Stuart-Smith, J. (2020). Changing perspectives on /s/ and gender over time in Glasgow. Linguistics Vanguard, 6(s1). https://doi.org/10.1515/lingvan-2018-0064

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). FINNISH SOUND STRUCTURE. In *STUDIA HUMANIORA OULUENSIA 9*. https://oulurepo.oulu.fi/bitstream/handle/10024/36099/isbn978-951-42-8984-2.pdf?sequence=1

Tilastokeskus. (n.d.). *StatFin: Väestölaskenta – Väestön kieli, kansalaisuus ja ulkomailla syntyneet (passiivinen aineisto)*. Statistics Finland. Retrieved May 22, 2025, from https://pxdata.stat.fi/PXWeb/pxweb/fi/StatFin_Passiivi/StatFin_Passiivi__vaelaskp/?tablelist=true

Tilastokeskus. (n.d.). *Tilastokeskus*. https://stat.fi/tup/alue/kuntienavainluvut.html#?year=2025&active1=SSS

UT Southwestern Medical Center. (n.d.). *Aging Voice*. Retrieved May 29, 2025, from https://utswmed.org/conditions-treatments/aging-voice/

Vaattovaara, S. J., & Halonen, M. M. (2015). Missä on ässä? "Stadilaisen s:n" helsinkiläisyydestä. In M.-L. Sorjonen, A. Rouhikoski, & H. Lehtonen (Eds.), Helsingissä puhuttavat suomet : Kielen indeksisyys ja sosiaaliset identiteetit (pp. 40-83). (Suomalaisen Kirjallisuuden Seuran toimituksia; No. 1310). Suomalaisen Kirjallisuuden Seura.

Vaattovaara, J., & Halonen, M. (2024). **/s/* ideology in twentieth-century educational materials in Finland: On the language-political and sociocultural underpinnings.* Nordic Journal of Linguistics, Advance online publication.* https://doi.org/10.1017/S0332586524000131

Zimman, L. (2017). Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics*, *55*(5). https://doi.org/10.1515/ling-2017-0018

Appendix

Appendix A: Scripts

A.1: Python Script for Audio Cleaning

```
# 1. IMPORTS
# All necessary libraries for both cleaning and volume adjustment.
import soundfile as sf
import matplotlib.pyplot as plt
import librosa
import numpy as np
from scipy import signal
import noisereduce as nr

# 2. DEFINE THE CLEANING FUNCTION
def clean_audio(input_file, output_file, hp_cutoff=50, lp_cutoff=12000):

    # Load audio using librosa
    audio, sample_rate = librosa.load(input_file, sr=None)

    # High-pass filter to remove low-frequency hum
    sos_hp = signal.butter(4, hp_cutoff, btype='highpass',
fs=sample_rate, output='sos')
    filtered_audio = signal.sosfilt(sos_hp, audio)

    # Noise reduction to remove static and background noise
    # The 'prop_decrease' parameter controls how aggressively to reduce
noise.
    reduced_noise = nr.reduce_noise(y=filtered_audio, sr=sample_rate,
prop_decrease=0.5)

    # Low-pass filter to remove high-frequency artifacts
    sos_lp = signal.butter(10, lp_cutoff, btype='lowpass',
fs=sample_rate, output='sos')
    cleaned_audio = signal.sosfilt(sos_lp, reduced_noise)

    # Save the cleaned audio file
    sf.write(output_file, cleaned_audio, sample_rate, format='WAV')
    print(f"Cleaned audio saved to: {output_file}")
    return output_file


# 3. DEFINE THE VOLUME ADJUSTMENT FUNCTION
def adjust_volume(input_file, output_file, method='normalize',
gain_db=6.0):

    # Load audio using librosa
```

```
    audio, sample_rate = librosa.load(input_file, sr=None)

    adjusted_audio = None
    if method == 'normalize':
        peak_amplitude = np.max(np.abs(audio))
        if peak_amplitude > 0:  # Avoid division by zero
            adjusted_audio = audio / peak_amplitude
        else:
            adjusted_audio = audio # Audio is silent

    elif method == 'amplify':
        # Convert dB to a linear gain factor
        gain_factor = 10.0 ** (gain_db / 20.0)
        adjusted_audio = audio * gain_factor


    # Save the adjusted audio file
    sf.write(output_file, adjusted_audio, sample_rate, format='WAV')
    print(f" Volume-adjusted audio saved to: {output_file}")
    return output_file


# 4. EXAMPLE USAGE
if __name__ == "__main__":
    # Define input and output file paths
    original_file = "input_audio.wav"

    # Step 1: Clean the audio
    # The output of this step will be the input for the next
    cleaned_file = "temp_cleaned.wav"
    clean_audio(original_file, cleaned_file)

    # Step 2: Adjust the volume of the cleaned audio
    # Option A: Normalize the cleaned audio
    final_file_normalized = "final_normalized.wav"
    adjust_volume(cleaned_file, final_file_normalized,
method='normalize')

    # Option B: Amplify the cleaned audio by 6 dB
    final_file_amplified = "final_amplified.wav"
    adjust_volume(cleaned_file, final_file_amplified, method='amplify',
gain_db=6.0)
```

A.2: Praat Script for Alignment

```
#Set paths to the input files
audioFile$ = "/Users/your/audio.wav"
textGridFile$ = "/Users/your/annotation.TextGrid”
outputFolder$ = "/Users/your/folder"

# Create the speech synthesizer
Create SpeechSynthesizer: "Finnish", "Female1"  ; changed if male participant
speechSynth = selected("SpeechSynthesizer")

# Read the input files
Read from file: audioFile$
sound = selected("Sound")

Read from file: textGridFile$
tg = selected("TextGrid")

# Settings
wordTier = 8
numberOfWords = Get number of intervals: wordTier

# LOOP over all words
for i from 1 to numberOfWords
    selectObject: tg
    word$ = Get label of interval: wordTier, i
    startTime = Get start time of interval: wordTier, i
    endTime = Get end time of interval: wordTier, i
    duration = endTime - startTime

    if word$ <> "" and word$ <> "-" and startTime <= endTime and duration >= 0
        selectObject: sound
        Extract part: startTime, endTime, "rectangular", 1.0, "no"
        realSegment = selected("Sound")

        Create TextGrid: 0.0, duration, "word phoneme", " "
        wordTG = selected("TextGrid")
        Set interval text: 1, 1, word$

        alignedTG = 0
        selectObject: speechSynth, realSegment, wordTG
        alignedTG = To TextGrid (align): 1, 1, 1, -40.0, 0.1, 0.1

        if alignedTG <> 0
            selectObject: alignedTG
            numPhonemes = Get number of intervals: 2

            # Loop through all phonemes in tier 2
            has_s = 0
            for p from 1 to numPhonemes
                phoneme$ = Get label of interval: 2, p
```

```
                # Check for 's' phoneme in any form
                if index(phoneme$, "s") > 0
                    has_s = 1

                endif
            endfor

            if has_s = 1
                idx$ = string$(i)
                Save as text file: outputFolder$ + "/aligned-" + idx$ + ".TextGrid"
                selectObject: realSegment
                Save as WAV file: outputFolder$ + "/phoneme-" + idx$ + ".wav"
            endif

            removeObject: alignedTG
        endif

        removeObject: realSegment
        removeObject: wordTG
    endif
endfor
```

A.3: Praat script for Centre of Gravity

```
folder$ = "/Users/your/folder/"

Create Strings as file list: "List", folder$ + "phoneme-*.wav"
numFiles = Get number of strings
writeInfoLine:"There were", numFiles, " files found"
for i from 1 to numFiles
       selectObject: "Strings List"
    wavName$ = Get string: i
    idxStart = index(wavName$, "phoneme-") + length("phoneme-")
       idxEnd = index(wavName$, ".wav") - 1
       idx$ = mid$(wavName$, idxStart, idxEnd - idxStart + 1)
       soundName$ = "phoneme-" + idx$ + ".wav"
    tgName$ = "aligned-" + idx$ + ".TextGrid"
    Read from file: folder$ + soundName$
    sound = selected("Sound")
    Read from file: folder$ + tgName$
    tg = selected("TextGrid")
    segmentNumber = 1
    tier = 4
    numIntervals = Get number of intervals: tier
    for j from 1 to numIntervals
             selectObject: tg
        label$ = Get label of interval: tier, j
        if label$ = "s"  ; Check if interval is labeled "s"
            start = Get start time of interval: tier, j
            end = Get end time of interval: tier, j
```

```
                    selectObject: j
            # Extract the part of the sound that corresponds to the /s/ interval
            segment = Extract part: start, end, "rectangular", 1, "yes"
            # Convert the segment to a spectrum
            selectObject: segment
            To Spectrum: "yes"
            # Now, select the spectrum to calculate CoG
            selectObject: selected("Spectrum")
            cog = Get centre of gravity: 1.0
            # Output CoG for this /s/ interval
            appendInfoLine: "phoneme-", idx$, tab$, "segment ", segmentNumber,
tab$, "start: ", fixed$(start, 3), ", end: ", fixed$(end, 3), tab$, "CoG: ", cog
            # Clean up
            removeObject: selected("Spectrum")
            removeObject: segment
            # Increment segment counter for the next segment
            segmentNumber = segmentNumber + 1
        endif
    endfor
    removeObject: sound
    removeObject: tg
endfor
```

Appendix B: Statistics

B.1 The results of the Linear Mixed Effects model for Gender

Estimated Marginal Means for Gender (Adjusted for Location)

	Gender	emmean	SE	asymp.LCL	asymp.UCL
1	female	-868.024	87.686	-1,039.886	-696.162
2	male	-950.333	75.723	-1,098.748	-801.919

B.2. Pairwise contrast for Gender

Pairwise Contrasts for Gender

Contrast	Contrast Estimate	Standard Error	df	Lower CI	Upper CI	z.ratio	p-value	Estimated Mean Female CoG (Hz)	Estimated Mean Male CoG (Hz)
female - male	82.310	121.299	Inf	−155.432	320.052	0.678567	0.497	5,131.976	5,049.667

B.3 The results of the Linear Mixed Effects model: location

Location	Estimated Mean	Standard Error	Lower CI	Upper CI
Alatornio	-1560.01951	406.9673	-2357.6608	-762.378185
Artjarvi	-547.46205	284.4194	-1104.9138	9.989691
Askola	82.75369	283.9977	-473.8716	639.378938
Eurajoki	-1109.54338	284.3038	-1666.7686	-552.318188
Hailuoto	-532.71999	284.8943	-1091.1025	25.662519
Hinnerjoki	-1158.08638	284.0908	-1714.8942	-601.278590
Ikaalinen	-250.05197	285.2211	-809.0751	308.971166
Jaala	-326.50121	283.9402	-883.0138	230.011419
Jamsa	-729.81995	283.5659	-1285.5989	-174.041031
Jurva	-994.91976	285.9099	-1555.2928	-434.546699
Kalajoki	-1190.84404	289.9794	-1759.1932	-622.494868
Kerimaki	-1189.66717	286.3991	-1750.9992	-628.335198
Kiihtelysvaara	-185.87843	283.2958	-741.1279	369.371044
Lapinlahti	-978.15282	290.6332	-1547.7834	-408.522225
Lappajarvi	-1119.30288	284.0299	-1675.9912	-562.614579
Lappee	-812.10700	284.6582	-1370.0269	-254.187147
Loppi	-568.98730	284.1958	-1126.0008	-11.973853
Mikkeli	-1560.03448	284.0829	-2116.8267	-1003.242278
Noormarkku	-1677.24725	284.2098	-2234.2881	-1120.206377
Padasjoki	-1188.87198	285.7862	-1749.0026	-628.741381
Palkane	87.99375	407.9844	-711.6410	887.628510
Pielisjarvi	-923.84217	284.3366	-1481.1317	-366.552602
Pihtipudas	-1280.04825	283.7153	-1836.1200	-723.976475
Saarijarvi	-887.96566	284.5076	-1445.5903	-330.340965
Salla	-1173.17059	407.3428	-1971.5478	-374.793404
Sippola	-1224.32346	290.7009	-1794.0868	-654.560150
Sotkamo	-769.75517	291.7944	-1341.6617	-197.848617
Suomussalmi	-1208.71317	290.1356	-1777.3685	-640.057870
Velkua	-936.65824	284.6322	-1494.5271	-378.789324
Vihti	-1361.41150	284.2214	-1918.4752	-804.347818

B.4. Results for linear mixed effects model with population data

	CoG			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	5147.00	104.88	4941.38 – 5352.62	**<0.001**
pop municipality z	-159.31	80.04	-316.23 – -2.39	**0.047**
pop region z	111.53	78.56	-42.49 – 265.54	0.156
swedish pct z	-47.21	76.79	-197.76 – 103.33	0.539
Gender [male]	-123.19	119.33	-357.15 – 110.77	0.302
Random Effects				
σ^2	246601.60			
$\tau_{00\ Folder}$	167020.79			
$\tau_{00\ location}$	89427.09			
ICC	0.51			
$N_{location}$	29			
N_{Folder}	55			
Observations	4025			
Marginal R^2 / Conditional R^2	0.085 / 0.552			

B.5. LMER with Z-normalized data: Gender

	Gender <fct>	emmean <dbl>	SE <dbl>	df <dbl>	asymp.LCL <dbl>	asymp.UCL <dbl>
1	female	1.415619e-16	0.02567916	Inf	-0.05033022	0.05033022
2	male	-3.848236e-17	0.02278997	Inf	-0.04466753	0.04466753

B.6. LMER with Z-normalized data: location

location	emmean	SE	df	asymp.LCL	asymp.UCL
Alatornio	4.70e-16	0.1240	Inf	-0.243	0.243
Artjarvi	4.43e-17	0.0843	Inf	-0.165	0.165
Askola	7.62e-17	0.0784	Inf	-0.154	0.154
Eurajoki	1.70e-16	0.0834	Inf	-0.163	0.163
Hailuoto	-1.51e-16	0.0906	Inf	-0.178	0.178
Hinnerjoki	3.05e-16	0.0791	Inf	-0.155	0.155
Ikaalinen	4.72e-16	0.0930	Inf	-0.182	0.182
Jaala	6.97e-16	0.0759	Inf	-0.149	0.149
Jamsa	3.19e-16	0.0725	Inf	-0.142	0.142
Jurva	-1.93e-16	0.1030	Inf	-0.202	0.202
Kalajoki	6.08e-16	0.0745	Inf	-0.146	0.146
Kerimaki	8.74e-17	0.1080	Inf	-0.211	0.211
Kiihtelysvaara	-4.44e-16	0.0678	Inf	-0.133	0.133
Lapinlahti	6.77e-16	0.0785	Inf	-0.154	0.154
Lappajarvi	2.63e-16	0.0782	Inf	-0.153	0.153
Lappee	-5.03e-16	0.0881	Inf	-0.173	0.173
Loppi	1.42e-16	0.0809	Inf	-0.159	0.159
Mikkeli	-2.16e-16	0.0796	Inf	-0.156	0.156
Noormarkku	-1.39e-16	0.0819	Inf	-0.161	0.161
Padasjoki	-4.14e-16	0.0993	Inf	-0.195	0.195
Palkane	-2.83e-16	0.1370	Inf	-0.268	0.268
Pielisjarvi	-1.96e-16	0.0834	Inf	-0.163	0.163
Pihtipudas	-2.56e-17	0.0747	Inf	-0.146	0.146
Saarijarvi	-5.91e-16	0.0861	Inf	-0.169	0.169
Salla	-4.56e-16	0.1290	Inf	-0.253	0.253
Sippola	-2.70e-16	0.0852	Inf	-0.167	0.167
Sotkamo	7.17e-16	0.0980	Inf	-0.192	0.192
Suomussalmi	1.69e-16	0.0734	Inf	-0.144	0.144
Velkua	1.11e-16	0.0878	Inf	-0.172	0.172
Vihti	9.91e-17	0.0822	Inf	-0.161	0.161

Results are averaged over the levels of: Gender
Degrees-of-freedom method: asymptotic
Confidence level used: 0.95

B.7. LMER with recording year

effect	term	estimate	std.error	statistic	conf.low	conf.high
<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
fixed	(Intercept)	5105.016070	92.669393	55.0884805	4923.387396	5286.644743
fixed	Year_centered	1.867739	3.066681	0.6090427	-4.142844	7.878323

Appendix C: R code

C.1 LMER code

```
# setting sum contrasts
contrasts(df$location) <-
contr.sum(length(levels(df$location)))
contrasts(df$Gender) <- contr.sum(length(levels(df$Gender)))

library(broom.mixed)
library(lme4)
library(emmeans)
# Fit the model
model <- lmer(CoG-6000 ~ Gender + location + (1 | Folder), data
= df)



# EMMs for location
emmenas_model <- emmeans(model, specs = ~location)
# EMMs for gender
emmeans_loc <- emmeans(model, specs=~Gender )
# Pairwise contrasts between genders
contrast_gender <- contrast(emmeans_loc, method = "pairwise")
```

C.2 LMER with population data

```
## Municipality population
# Standardize municipality population (z-score)
merged_data <- merged_data %>%
  mutate(pop_municipality_z = scale(pop_municipality))

# Join standardized population back to original df by
location for modeling
df <- df %>%
  left_join(merged_data %>% select(location,
pop_municipality_z), by = "location")

## Region population
# Standardize region population
merged_data <- merged_data %>%
  mutate(pop_region_z = scale(pop_region))

# Join standardized region population to df by location
df <- df %>%
  left_join(merged_data %>% select(location, pop_region_z),
```

```
by = "location")

## Swedish-speaking population

# merge in the Swedish-speaking percentage
merged_swedish <- merged_data %>%
  left_join(swedish_pct, by = "location")

df <- df %>%
  left_join(select(merged_swedish, location, swedish_pct), by
= "location") %>%
  mutate(swedish_pct_z = scale(swedish_pct, center = TRUE,
scale = TRUE))


## Fit the linear mixed-effects model
model_population <- lmer(CoG ~ pop_municipality_z +
pop_region_z + swedish_pct_z + Gender + (1 | location) +  (1
| Folder), data = df)
```

C.3 Post-hoc analysis: normalized data

```
#Normalize the data by speaker
df_normalized <- df %>%
  group_by(Folder) %>%
  mutate(
     CoG_z = (CoG - mean(CoG, na.rm = TRUE)) / sd(CoG, na.rm =
TRUE)
  ) %>%
  ungroup()


# Fit the model
model_normali <- lmer(CoG_z~ Gender+location +(1|Folder), data =
df_normalized)
# EMMs for location
emmenas_model_normali <- emmeans(model_normali, specs = ~location)

# EMMs for gender

emmeans_loc_normali <- emmeans(model_normali, specs=~Gender)
```

C.4 Post-hoc analysis: recording year

```r
library(lme4)

df <- df %>%
  mutate(
     year = as.numeric(year),
     Year_centered = year - mean(year, na.rm = TRUE)
  )
year_lmm <- lmer(CoG ~ Year_centered + (1|location), data =
df )
```