A perceptual investigation in spoken Trøndersk Norwegian

Eugenio Palombella

12245984

Research Master's in Linguistics

Graduate School of Humanities

Universiteit van Amsterdam

Supervisor: Dr Silke Hamann

Second reader: Dr Paul Boersma

24th July 2020

UNIVERSITEIT VAN AMSTERDAM

Abstract

The present study investigated the circumstances of perception around the merger of the voiceless palatal fricative /ç/ and the voiceless retroflex fricative /ʂ/ as uttered by speakers of the Trøndersk variety of Norwegian. Specifically, this study looked at the inference of speaker age on the perception of said fricatives. The investigation was carried out through the means of an online setup, in which 140 participants were exposed to two experimental conditions, each containing two speakers; the speakers uttered 5 semi-randomised tokens for each word of a minimal pair (*kjære* 'dear', and *skjære* 'to cut') whose initial fricative sounds were edited on a five-step continuum to assume the spectral features of the opposing phonemes, for a total of 10 tokens per speaker. The two aforementioned speakers were presented as 24 and 40 years old, alternating the declared age between conditions. The participants had to listen to the tokens and make a choice between the two orthographic forms of the minimal pair given above, in order to match what was heard to one of the two words. Given the current literature on the production of the merger, the expectations would have had younger speakers merge more (that is, to realise canonical palatals as retroflexes /ʂ/), which in turn would have led subjects to perceive a merged pronunciation more frequently in speakers labelled as "young", its presence notwithstanding. Nevertheless, our observations showed that perception was less accurate for speakers labelled as old. The only statistically significant result was that of better recognition rates for participants sorted in the first of the two conditions. It is posited that high perception rates for the palatal fricative /ç/ on young speakers are due to a perceived hyperarticulatory effect – that is, the perception of a starker contrast between phonemes in comparison to the older speakers' utterances – and to the attribution of higher prestige to the voiceless palatal fricative. The study gives indications for future lines of work.

Keywords: fricatives, Norwegian, perception, prestige.

Table of contents

1. Introduction

Norway is a multi-faceted country where the lack of a standardised language coexists with two written norms, arbitrarily chosen. If a Norwegian gets asked how many dialects are there in Norway, they are likely to answer, half-jokingly: "five million, one for each of us". And while this is an exaggeration, the language variation across geographical places and singular individuals is indeed enough to make a first-time learner's head spin.

This variation was fostered by a low population density, coupled with a large surface and a very loosely knit infrastructure system; the situation has started to change over the course of the 20th century, with the advancement of the railway and road system taking more and more space over the historical water transport. Such development has led to distant communities being gradually closer to each other, and to the sudden awareness of dialects – no more considered as mundane, daily and unexceptional – but rather as an identity and a cultural heritage to preserve and be proud of.

The development of infrastructures has led to mass immigration towards the biggest urban centres in the South of the country: Oslo, the capital city, has become a hub for people from all across the country. This has contributed to the growth and spread of Urban East Norwegian (UEN), a pseudo-standardised variety of the language, containing its own linguistic idiosyncrasies. One of these recent particularities is the ongoing merger of two phonemes, the voiceless palatal fricative /ç/ and the voiceless retroflex fricative /ʂ/, converging onto a retroflex realisation.

This peculiar pronunciation has spread outwards to the northernmost cities: one of these, Trondheim, has been the focus of a number of studies investigating at large the production of said merger. However, we are left with a gap concerning the perceptual side: this thesis aims to shed light on it by pioneering the circumstances of perception of merged pronunciations, and by giving an account of the prestige status of this linguistic change from the perspective of the listeners.

1.1. The Norwegian language

Norwegian is an Indo-European language belonging to the North Germanic family. As anticipated, Norway recognises two written norms (*bokmål* and *nynorsk*) but no official spoken language: the national territory is fractured in a dialectal continuum, in which every regional variety is mutually intelligible to the others (Venås & Skjekkeland, 2019). The unofficial standard (that is, the one taught to foreigners in and outside the country's borders; Vanvik, 1979) is the aforementioned UEN variety, also called *bokmålsnær* (*bokmål*-approximant) for its vicinity to one of the written conventions. Despite this unofficial status, UEN has not seeped into the larger speaking community as a recognised norm. Kristoffersen (2000: 7) points out that the more or less open prohibition of a standard spoken variety is imputable to the will to maintain every Norwegian dialect on an equal level.

A map of the Norwegian *fylker* (counties), shown in Figure 1, serves the purpose to localise where UEN is spoken.

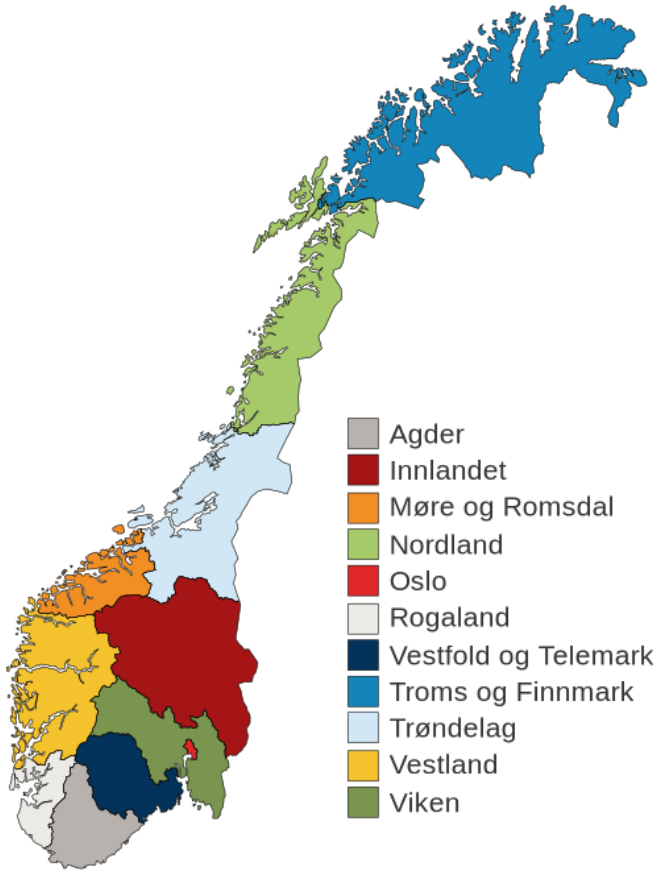


Figure 1. Map of the Norwegian counties. Approximately, UEN is spoken in the counties of Oslo, Vestfold og Telemark, Viken and Innlandet (source: Wikipedia)

Phonetically, the Norwegian language belongs to a small set of languages of the world employing a voiceless palatal fricative /ç/ phoneme, contrasting with a voiceless retroflex fricative /ʂ/. This opposition is very uncommon, as it was found in only 14 languages out of 504 containing the palatal fricative (LAPSyD, 2020). These two phonemes are at the core of the present research, due to their tendency to merge into a single retroflex fricative.

1.2. An ongoing merger

Linguistic change is part of the constant innovation operated by humankind on its own means of communication. There is an undeniable tendency to condemn it, pointing at it as the downfall of a standardised form of speech (a tendency of which every linguist is wary, openly criticised in Sandøy, 2013: 130). However, linguistic change and its observation are even more relevant today, as in Norway the "standardised language" takes the form of many decentralised dialects which are undertaking very swift transformations and losing their peculiar traits, often considered obsolete, to conform to the language of the biggest urban centres (Skjekkeland, 2005).

One of the most recorded changes is the merger between two distinct phonemes – the voiceless palatal fricative /ç/ and the voiceless retroflex fricative /ʂ/ – having a unified /ʂ/ pronunciation as its outcome. See the following example (1), showing the effects of the merger on a minimal pair:

(1) *kjekk* 'handsome' – *sjekk* 'cheque'
/ˈçɛkː/ 'handsome', "canonical" pronunciation
/ˈʂɛkː/ 'cheque', "canonical" pronunciation
/ˈʂɛkː/ merged pronunciation of *kjekk* – not contrastive anymore, entirely context-dependent

The status of such merger is hard to pinpoint, as most of the literature concerning UEN produced discrepant results. Gender is not always considered a significant factor (for a detailed discussion, see §2.4.) – and when it is, it tends to trump the Labovian expectations of female speakers as the driving force of change. The inference of geographical provenance across neighbourhoods when considering a single city was proven to be of little significance too (§2.4.) Ultimately, age was the only robust relevant factor across studies, with younger speakers more likely to merge phonemes, as expected by the general rules of language change. Further research outside the UEN area (Bergen, Molde[1], Tromsø, Trondheim; §2.4.) resulted in similar outcomes, with the only development being a differential in advancement of the merger, seeing the northernmost cities (specifically, Tromsø; Jacobsen, 2015) in an earlier stage of the merger acquisition.

Despite the mechanisms of linguistic change being very subtle, it is known from the accounts of Dalbakken (1996) and Nesse (2013) that the retroflex-palatal merger has been largely relegated as a remnant of childlike pronunciation or as a plain pronunciation mistake,

1.3. Scope of the thesis and research questions

As anticipated, the focus of this research lies in perception. Many studies have investigated the production of the Norwegian /ç/-/ʂ/ merger (§2.4.). This study builds upon their results – taking into account the most crucial factors thereby exposed – and sets up the experimental conditions to specifically test the inference of the age factor upon perception.

If the merged pronunciation is tied to the age of the speaker, how is perception tied to expectations, and how much can it be manipulated? If a listener is given an artificially manipulated stimulus, edited to sit in between the frequency values of the retroflex and palatal fricatives, along with information about the age of the speaker – how much does this information guides perception towards categorising the phoneme in one direction or the other?

The information about the circumstances of production presented throughout the literature review in § 2.4. leads to hypothesise that the participants to our experiment will be more likely to hear a merger – irrespective of its actual presence – in the speech of a younger subject.

The purpose of this thesis is to both answer the abovementioned research questions and confirm or disprove the hypothesis put forth, setting a new milestone in the research on linguistic change and thoroughly investigating its place in the faceted and ever-changing panorama of the Norwegian dialects.

1 This study (Rød, 2014) is very interesting, but geographically marginal and beyond the scope of this thesis.

1.4. Structure of the thesis

The present thesis is largely built upon 25 years of observation on the palatal-retroflex merger's evolution. The introductory chapter guided the reader through the topic of research and set the hypothesis and the research question.

The discussion is expanded in Chapter 2, which consists of a thorough literature review going from the seminal studies on the pre-existing conditions surrounding the production of such merger, to local studies having the city of Trondheim as their focus. The chapter is completed by the sociolinguistic parameters necessary to evaluate the causes and the social status of the merger. In addition, the experiment hereby conducted is based on two sociolinguistic studies, which shall be presented in §2.6. along with the indication of what has been kept and what has been modified in their experimental structure to fit our specific needs.

Chapter 3 contains all the sections inherent to the experiment: it gives indication of the material used, of the methods of manipulation, the experimental procedure, the statistical analysis, and its results – making them accessible with a condensed report and graphs to illustrate the findings.

Chapter 4 discusses the findings from a twofold point of view – that of statistical results and that of general observations encompassing the demographics of the experiment, the anecdotal data given by the participants' comments, and possible theoretical explanations for our findings. In this chapter, we also give indications of what can be changed or improved to obtain better and more conclusive results.

Chapter 5 concludes the thesis, giving a short and approachable summary of the results by linking them directly to the research questions and the hypothesis – thus giving a brief, conclusive answer to our study.

2. Theoretical setting

2.1. Introduction

This chapter shall provide the reader with a comprehensive overview on the relevant literature about the merger. It follows a micro- to macroscopic approach, starting off with the phonemes themselves, then widening its scope to the definition and implications of mergers, to the seminal literature about the Norwegian /ʂ/-/ç/ merger in the city of Trondheim and, lastly, to the experiments that the present thesis intends to use as a baseline.

As a word of warning, the reader is invited to keep in mind that all the given information in this chapter is relative to the city or area in which the mentioned studies had been conducted. This is due to the incredible range of variation within the Norwegian territory and to the fact that Norwegian cannot be considered a unitarian language, but rather a dialect continuum, influenced to various degrees by Danish. For more information about the history of the Danish rule and the dialectal features of Norwegian, see Jahr (1990), Skjekkeland (1997; 2005; 2010) and Mæhlum & Røyneland (2012).

2.2. A debate on postalveolar fricatives

The phonemic classification of the two aforementioned fricatives, /ç/ and /ʂ/, is not so clear-cut: whereas the voiceless palatal fricative is unanimously attested in every phonemic description of Norwegian (most notably: Vanvik, 1979; Kristoffersen, 2000), the nature of the contrastive sound – sometimes transcribed as /ʂ/ and sometimes as /ʃ/ – is much more disputed.

These two phonemes – /ʂ/ and /ʃ/ – have a high degree of similarity in terms of articulation. Through electropalatography (EPG) and electromagnetic articulography (EMA), Moen & Simonsen (2011) demonstrated how retroflex fricatives have a generally postalveolar place of articulation, but a lesser degree of backward bend of the tongue than their plosive counterparts (e.g. the Norwegian /ʈ/. Regarding the position of the tongue, Simonsen & Moen (2004: 618) confirm that: "the most common configuration is a straight, non-retroflexed tongue" – much as /ʃ/.

Given these degrees of similarity, how is it possible to distinguish /ʃ/ and /ʂ/ with precision? Moen & Simonsen (2011) come to help once more: the voiceless retroflex fricative

/ʂ/ shares with the Norwegian retroflex stops an apical articulation, whereas /ʃ/ is articulated with the tongue blade. And even given the varying – yet rather low – degrees of retroflexion of the tongue, Hamann (2003) argues that Norwegian apicals should be treated as retroflexes as they share resemblance and phonological behaviour with the retroflexes of other languages. Establishing four defining factors (apicality, posteriority, sublingual cavity and tongue retraction), Hamann (2003) proceeds to illustrate the flexibility of retroflex consonants in adapting to each category: the Norwegian voiceless retroflex fricative may not satisfy the posteriority condition, that is, the articulation of the tongue towards the velum, but it still complies to the other three.

Therefore, this thesis will treat what some other studies classify as /ʃ/ as a voiceless retroflex fricative /ʂ/.

2.3. Fundaments of sociolinguistics

Sociolinguistics is the discipline that deals with the intertwined factors of change between language and society in a geographically limited environment. Its tradition has been consolidated in the 1960s by the works of William Labov, whose method contributed to shape the field as the analysis of variation by the means of *charting* and *statistical analysis*. A small, but important part of the Labovian method whose incorporation in the present study will be later explained (§3.4.), is trying to raise feelings during the test: in fact, emotional involvement from the side of the test subject causes a "move[ment] without realizing it into a more casual style of speech" (Aitchinson, 2001: 46).

Although sociolinguistics deals with a synchronic view of time, a word must be spent about driving forces behind language changes, which are the results of diachronic processes. We intend to simplify the broader debate following two definitions, adapted from Labov (1972; as quoted in Sandøy, 2013) and Torp & Vikør (2012; as described in Jacobsen, 2015). The first driving force is set in motion by inner factors (or, as Labov, 1972 puts it, to "pressure from below"): it is the result of "pressures […] operating upon entire linguistic systems, in response to social motivations which are relatively obscure and yet have the greatest signification for the general evolution of the language" (Labov, 1972: 123), that is, forces within the environment of the language that lead alterations and structural changes. The second driving force is defined as "outer" by Torp & Vikør (2012) and as "pressure from above" by Labov (1972). "Social pressures from above […] represent the overt process of social correction applied to individual linguistics forms" (Labov, 1972: 123): in this case, social pressures can be triggered by contact, normalisation of the language or language politics (e.g. the process of suppression of the Sami language and the norwegianisation of the native Sami and Kven from the 1700s to the 1980s; Hansen & Olsen, 2004). Sandøy (2013: 130) also defines social correction as channelling the language into standardised variables transmitted through the teaching of the language from parents or instructors.

An extremely relevant factor from the array of sociolinguistic causes for change – among which mergers can be found – is *linguistic prestige*. Giving a single definition of prestige is difficult, although it can be described in broad strokes as the degree of social value and consideration given to the language itself and its dialects by a community that shares this language. Languages are devoid of superstructural value, having nothing that naturally determines their worth, but as Bonfiglio (2002: 23) points out: "it is the connection of the language in question to the phenomena of power that determines the value of that language and that contributes to the standardization process". When compared to the definitions of linguistic change given in the previous paragraph, it is clear how this interplay between power and language belongs to Torp & Vikør's (2012) "outer change" system. However, inner change in prestige can manifest itself in a further broadening of its definition, branching out to seek *covert* or *overt* prestige. These terms roughly correspond to an unconscious (covert) or conscious (overt) attitude of the speaker towards the change of their personal language patterns (Sandøy, 2013: 126). Attempts to conform to *covert* prestige entail the usage of social patterns belonging to nonstandard varieties of the language injected into the standard language itself; *overt* prestige is the reverse phenomenon, consciously raising the register of one's speech to conform to standard, non-vernacular speech, to the point that, in the pursuit of prestige, technical jargon becomes more important than a clearly conveyed message (Hudson, 2000).

This brief excursus shows how prestige is a major driving force in language change, and how it may indeed be linked to the main subject of interest of this study – i.e. mergers, the phenomena by which two phonemes become one; and it is known as a fact how any phonemic change often happens as a means to heighten one's prestige (cf. Hudson, 1981, and the case of British RP: as such variety is considered more prestigious than local British vernaculars, speakers who strive for prestige tend to change their phonemic inventory to match that of RP).

As far as mergers go, Labov (2010: 321) masterfully explains in a unified theory what others (Trudgill & Foxcroft, 1978; Harris, 1985; Herold, 1990; Shen, 1990) have observed throughout the years – giving us the opportunity to categorise with precision what is currently happening in the Norwegian language. Out of the three mechanisms described – merger by approximation, by transfer and by expansion, we are interested in the first two. These categories will be applied to better classify the findings in the literature (§ 2.4.).

The first merging mechanisms is that of *approximation*. A merger by approximation implies two phonemes overlapping until no distinction can be made anymore: this means the creation of a third realisation, a "mean value intermediate" (Labov, 2010: 321), sitting between the original two sounds.

The second movement pattern is the *merger by transfer*, defined as the complete superimposition of the phonemic values of one phoneme on the other, without intermediate forms. As a result, a form A and a form B in the course of being merged by transfer admit only one outcome: either A or B, unchanged and unscathed. Even more than mergers in general, the merger by transfer is influenced by prestige. Labov (2010: 321) observes how it is a stable sociolinguistic variable, due to the ascendancy of one phoneme as the prestige form, whereas the second one is relinquished due to a social stigma.

The next section introduces sociolinguistics to the linguistic environment of Norway: from the theoretical subtleties of this section, we shall move on to illustrate examples of the main object of this study, the /ç/-/ʂ/ merger, giving an overview of the past studies and trying to pinpoint a provisional explanation for sociolinguistic movements.

2.4. The Norwegian /ç/-/ʂ/ merger

2.4.1. The first studies

The earliest account of a merger between palatal and retroflex fricative dates back to the 1940s (Simonsen & Moen, 2006). However, this phenomenon has gained recognition and was deemed worthy of being researched only starting from the 1980s onwards.

Two seminal studies in the field (Johannessen, 1983 and Papazian, 1994) opened the rush to research the merger.

Johannessen's (1983) research took place in Bergen and used material collected in the 1970s; it considered gender, neighbourhood of provenance and age as relevant factors, but found significance only for age groups – with younger speakers merging more, as per expectations. However, when the merger took place, it was not consistent throughout the material, but it was limited to a small number of sparse instances (Jacobsen, 2015: 8).

Johannessen (1983: 27) himself admitted that the available material was insufficient to predict a development of the merger.

The latter study is Papazian (1994), which took place in the capital city, Oslo. The study was conducted through self-report question forms that the author handed out in schools[2]. The results showed that gender was the only variable that yielded significant results, weighing in favour of female subjects. Whereas the validity of reliance on self-report is doubtful, this study is still very important: the feedback was given directly by the participants, but one's perception can be skewed both by a lack of linguistic awareness or by a matter of perceived prestige. If the subjects were aware of a higher prestige for the palatal pronunciation, they were also more likely to report themselves as using that variety; consequently, people who reported the merger stood out. Given the results, it can be inferred that at the time of the study the /ç/ realisation was viewed as the more prestigious one.

The last seminal study covered here, despite its recency that sets it apart from the previous two, is van Dommelen (2019): to our knowledge, this is the only research covering the entirety of the nation – setting it as the most complete one. In fact, van Dommelen (2019) proceeded to analyse material from a large nation-wide database (NB Tale, property of the National Library of Norway, containing both read and spontaneous speech) from an acoustic perspective, in the attempt to frame the degree of completeness of the merger using gender, age and the type of speech (read/spontaneous) as variables. The NB Tale database consisted of 240 speakers in 12 groups containing 20 speakers with a shared dialectal background, instructed to use their everyday dialect for the recordings. Gender was noted down, and age was accounted for by separating speakers in two groups (age 18-40 and age 40-80). Out of the 240 speakers, 15 were excluded due to the absence of underlying /ç/ occurrences in the content of their speech, and only 32[3] merged pronunciations (that is, [ç] > /ʂ/) were found; van Dommelen (2019: 3) reports how only 6.4% of the read speech and 6.7% of spontaneous speech contained the merger, a rather small percentage. Additionally, it is reported that females tended to have the edge over males in change (18 vs. 14, not significant enough in regard to their Euclidean distances; van Dommelen, 2019: 4), and that no geographical clustering was found, besides 12 speakers of a south-western dialect and 6 speakers of an eastern (cf. UEN) dialect. van Dommelen (2019: 4) notes that for those who separated the two pronunciations, acoustic-phonetic contrasts were unsurprisingly starker. However, younger speakers tended to separate the two realisations more than the older group. In the light of the

2 Primary school, 5th grade (11-year-old); primary school, 9th grade (15-year-old); secondary school, 3rd grade (18-year-old) – Papazian (1994: 71).
3 Of which only 3 were above 40 years of age.

data presented, van Dommelen (2019) concludes that the merging phenomenon between /ç/ and /ʂ/ is still relatively rare, and no evidence points towards the endangerment and the eventual disappearance of the voiceless palatal fricative.

The next section deals with studies centred around the capital of the Trøndelag *fylke* and the focus of the present research, the city of Trondheim.

2.4.2. Trondheim

Trondheim was chosen as the fulcrum of this thesis for two reasons: the richness of studies on its dialect, and the extensive local research on the /ʂ/-/ç/ merger. A comprehensive overview on the production of such merger as presented here, leads us to a better understanding of what to expect in terms of perception. The hypothesis formulated in § 1.3. has its roots in these studies: higher merged pronunciation rates in the local youngest population – as expected by the mechanisms of language change – may mirror a higher response to the perception of the merger when the speakers used for the experiment come from the very same region.

The first major study was conducted by Dalbakken in 1996. The author tested 85 participants[4] through a four-part investigation, involving free-speech trials, texts read aloud and the production of two series of minimal pairs (in isolation and in carrier sentences). The variables taken into account by Dalbakken (1996) were gender, age, and neighbourhood of provenance. Results varied greatly from task to task: from a 13% of merging in the reading part, the numbers went up to 15.4% for the free speech test. Dalbakken (1996: 72-94) qualified this as a *tendency* to merge /ç/ into /ʂ/, without a significant distinction between sexes, but with a slight predominance of males as the driving force of change (1996: 125). No significant difference had been found for the neighbourhood of provenance, and the tests yielded balanced results across age groups.

The second study was conducted by Hårstad (2010) and later summarised in Hårstad & Opsahl (2013). The data was collected from 31 informants from Trondheim, aged 16-18 (Hårstad, 2010: 130). The test subjects were slightly unbalanced in favour of male participants (18, against 13 females). Hårstad's results yielded a positive match for the merger in 14 of the 31 informants, with an unbalance in favour of males (Hårstad & Opsahl, 2013: 59). What is more interesting, is the evolution of the sociolinguistic reading key behind this phenomenon. Hårstad (2010: 172) explains that: "[s]elv om "feil" /ʃ/-bruk [sic] er en

---

4 40 pupils from two 7 th grades of primary school (13-year-old) and 45 pupils from two 1st grades of secondary school (16-year-old).

novasjon, blir ikke dette trekket nødvendigvis oppfattet som noe ”moderne” […]” [even if the “erroneous” use of /ʃ/ is an innovation, this is not necessarily perceived as something “modern” – *personal translation*], and Jacobsen (2015: 9) reports that three years later, in Hårstad & Opsahl (2013: 82), the authors claim that it is still unknown whether the phonemic merger is a sign of a higher social status or not. The present study hopes to give a definite answer to this dilemma. Another very interesting factor is that Hårstad & Opsahl (2013: 82) reported that some speakers utilised middle forms between /ç/ and /ʂ/: this is particularly noteworthy because the picture that emerged from every other study is that of a merger by transfer (Labov, 2010: 321) whereas this is the first reported case of a merger by approximation (Labov, 2010: 321), in which two phonemes move to converge to a middle point.

The last and most recent study is Jacobsen (2015). This thesis is a sociolinguistic picture of the merger in Trondheim and Tromsø, the second biggest city in the northernmost part of Norway. Jacobsen (2015: 1) points out from the very beginning the biggest stigma on the /ç/-/ʂ/ merger: even though it represents an example of linguistic innovation, it is often perceived and condemned as linguistic *sabotage* – in line with the observations of Sandøy (2013: 130). The author also raises two valid questions about the interplay between the merger and the education system, which are here reported out of general interest, but whose scope lie beyond that of the present thesis: “[e]r det opp til skolen å stoppe eller fremme språkendringsprosesser? Er det i det hele tatt skolens sak å ha ei mening?” [is it up to the schools to stop or correct linguistic change processes? Is it a matter schools should have a say in at all? – *personal translation*] (Jacobsen, 2015: 1). To carry on her research, Jacobsen (2015: 39) interviewed 88 informants (24 girls and 16 boys from Tromsø, 29 girls and 19 boys from Trondheim) aged 14-15 (8th and 9th grade of the primary school): the test required the participants to count from 1 to 30 (with the numbers between 20 and 29 as critical targets, whose first part is invariably pronounced *tjue* /ˈçʉːə/; e.g. *tjueen*, *tjueto*, *tjuetre*…), to read a text (namely, a diary entry) out loud, to read minimal pairs out loud and to have a less formal part of the test consisting of a free speech interview to disclose the most vernacular parts of their everyday speech. Jacobsen’s (2015) general results showed how the merger involves 9.2% and 39.6% of the speakers of Tromsø and Trondheim, respectively. Out of the total tokens, the results showed how 21.7% of the females and 31.5% of the males employed the merger, confirming what already noted before – that is, the male speakers as major innovators in the context of the /ʂ/-/ç/ merger. Only one speaker (Jacobsen, 2015: 66) consistently merged every palatal fricative into a retroflex one. Jacobsen (2015: 92) also observed a strange reverse effect (that is, /ç/ in place of /ʂ/) for 5.5% (Tromsø) and 3.4% (Trondheim) of the speakers, concluding that the higher percentage of /ç/ for /ʂ/ in Tromsø can be interpreted as a blurring of lines typical of the geographical areas in which the *proper* merger (that is, /ʂ/ for /ç/) is a relative novelty. The difference in percentage between the /ʂ/-/ç/ merger and the opposite development do not suggest the existence of two different mergers in competition, but rather different stages of linguistic change. To better explain: in the beginning, when the occurrences of the merger are few, developments seem to be going in both directions. As the frequency of merging increases, the direction of development becomes clearer – unequivocally, from /ç/ to/ʂ/. This means that we are witnessing a starting phase in Tromsø, which in a few years would possibly lead to a similar situation as the one we see now in Trondheim. Conclusively, despite the presence of one informant having a complete merger and the high rates of exchange between the two phonemes, the results do not subvert the expectations of a collective tendency to still distinguish the palatal from the retroflex fricative.

The present study intends to consider what these studies claim unanimously – that is, the young age of the merging speakers – as the baseline for the experiment we intend to run. Furthermore, as the social status of the merger is still dubious, we hope to give a definitive answer to this dilemma.

2.5. From production to perception

Putting together what was reported in the last two sections, the picture emerging from the circumstances of the /ç/-/ʂ/ merger suggests that Norwegian is primarily dealing with a merger by transfer, exceptions notwithstanding (cf. merger by approximation in Hårstad & Opsahl, 2013). The factors to watch are reduced to age alone, as gender and neighbourhood provenance seem to have little effect on the results, not being consistently significant throughout the studies[5]. Unluckily, most of the studies have considered only urban areas[6]: it would be fascinating to follow the development of such merger in rural areas. Perhaps, a broadening of the geographical areas of interest would generate a broadening of the defining factors of the merger as well – with a higher inference of gender or a particular dialectal group. However, most of the academic attention seems to be focused on UEN and other urban varieties – and not without a reason: cities are more likely to host universities, which in turn

5 Males seem to be more proactive in linguistic change (§ 2.4.2), but the section is entirely centred on Trondheim. On a nation-wide scale, van Dommelen (2019) observed the reverse effect – females bringing linguistic change more than males.
6 This is not true for van Dommelen (2019), as the corpus he used was supposedly nation-wide: we cannot imply it was collected in cities only.

are aggregation points for young people and, possibly, willing test subjects. Gathering data in rural Norway would be extremely difficult, given the geographical extension of the country, the scarce population density outside of urban areas, the migration flows towards the cities and the lack of aforementioned gathering points in small towns.

The previously illustrated studies also have a shared baseline: they were built upon observations on production and articulation. This leaves a gap in what concerns the *perception* of such merger: we do know the extent of this change within the language and its speakers is considered a tendency but, if that is something that has come to be considered an integral part of everyday language, we do not know how likely people would actively *perceive* the merger. As noted by Hårstad & Opsahl (2013), we still do not precisely know how Norwegians perceive the *prestige* of merged and unmerged pronunciations; Sandøy (2013: 130) hints at social correction from parents and institutions, as linguistic change is seen as a source of outrage rather than as a natural phenomenon – giving a prominent clue to posit that a retroflex realisation in place of a palatal one must be regarded to with open hostility.

This study is set to fill this gap as much as possible but, to do so, it needs a robust experimental design that shall allow us to draw precise observations. To put it in context, let us proceed to illustrate the two studies that inspired the experiment in this thesis.

2.6. Theoretical baselines: Hay & Drager (2010) and Drager (2011)

Sociolinguistic studies need to be tailored to properly account for delicate interplays of elements concerning the speakers – age groups, geographical provenance, gender distribution, educational background and more. Using what has been established in earlier studies is common procedure: the inspiration for the present experiment comes from Hay & Drager (2010), and Drager (2011), fine-tuning their procedure to better fit our research topic. Whereas the first study's aim was to check the influence on perception by the exposure to an extralinguistic concept (specifically, the region of provenance of the speakers), the second one applies the perception manipulation techniques to an ongoing merger in New Zealand English, making it even more pertinent to the object of investigation in this thesis.

The procedure followed by Hay & Drager (2010) was designed to check the inference of geographical provenance on the perception of a synthesised vowel continuum, ranging from raised and fronted Australian tokens to lowered and centralised New Zealander tokens, in the absence of overt reasons to associate the voice to a precise region. The participants were divided in two groups subject to different experimental conditions: the test tokens were the same for the two conditions, but the researchers conducting the experiment made sure to draw the participants' attention to a stuffed animal – a kiwi, representing New Zealand, or a kangaroo, representing Australia – sitting in a cupboard containing the answer sheets for the experiment. After a brief moment of feigned surprise, the sheets were distributed and the participants were prompted to evaluate the adequacy of the synthesised voice, matching the target sounds (which were highlighted in the answer sheet) to a 6-step vocalic continuum. The results of Hay & Drager's (2010) study indicated that the perception of vowels could indeed be guided by which stuffed animal the participants were exposed to. It is inferred that linguistic and non-linguistic factors are closely intertwined, to the point that any extralinguistic factor, invoking preconceptions in the listeners' mind, can bias linguistic perception – primarily, the recognition of phonemic details. Hay & Drager (2010: 889) are cautious about their own results: as they explain, if a stuffed toy can shift perception, then the testing environment must be as neutral and as controlled as possible, as every detail could possibly trigger inferences on the participants' perception.

Drager (2011) establishes a further link with the present thesis, focusing on the relation between speaker age and phonemic perception. Her study aimed to demonstrate a link between the perceived age of speakers and phonemic variance in the context of a vowel chain shift in New Zealander English (NZE). To do so, she extrapolated a NZE minimal pair from a corpus (*bad/had*; Canterbury Corpus) and manipulated the vowels so as to create two continua ranging from *bad* to *bed* and from *had* to *head*. Test subjects were presented with two series of stimuli, both consisting of a picture of a speaker (out of four: old/young female, old/young male) followed by the auditory tokens slightly delayed: for each series, they had to choose the closest word to what they heard, thus being forced to a binary, categorical choice. Results showed that the relation between perception and age is duplicitous: if the information about the speakers is not disclosed and it needs to be inferred by extralinguistic clues, then it is not only their age to be of relevance, but the age of the participants as well. In fact, it plays a role in attributing an age to the speakers, when presented with visual cues about their appearance. The divergent age evaluations lead to different categorisation of vowels, bolstering the idea of a higher saliency of social cues than of acoustic ones. In addition to the relevance of age, gender was found to trigger significant differences too, in an ambivalent way similar to what discussed above: phonemic categorisation depends on the interplay between the gender of the speaker *and* that of the test subject. To conclude, Drager (2011) points out that sociolinguistic factors – often underplayed by formal theories – are known to create predictable patterns in linguistic analysis, as social factors are strongly tied to the storage and the categorisation of phonetic details.

We have decided to extrapolate three crucial factors for our experimental design from the past experience of Hay & Drager (2010) and Drager (2011). The first is that the attention of the participants must always be drawn to the misdirection element relevant to the experiment – the nationality through the stuffed animal in the former, the age through the pictures in the latter. Secondly, those misguiding devices can be applied to what we considered the ideal proving grounds for our theory, that is an online experiment. This represents the most neutral possible testing environment, as the participants are subject to a controlled amount of stimuli and can be reminded of the misdirection element through the means of a constant header identifying the speaker at all times. Thirdly, we decided to present the experiment as a series of binary choices rather than a choosing task from a continuum of possible answers. This complies with the concept of phonemic perception as categorical (Liberman et al., 1957). The application of such factors is explained in §3.

3. Experimental procedure

This chapter introduces the reader to the notion of salient features of the fricatives, applying them to the editing process employed to create the tokens used for the experiment (§3.3.). It also covers the experiment itself, listing the parameters used to determine the results (§3.4.) and the statistical method that led to the results (§3.5.).

3.1. Acoustics of fricatives

The objects of the present study are the voiceless retroflex fricative /ʂ/ and the voiceless palatal fricative /ç/. Every fricative has a shared property in the form of friction noise that the air in the oral cavity produces, but speakers are subject to substantial individual variations in pronunciation that make it hard to consistently describe the precise spectral characteristics and place of articulation of their fricatives. The two key properties described here are the centre of gravity and the spectral peaks, as presented in Scholtz (2009) in her study on Trøndersk Norwegian.

The centre of gravity (CoG) is the most commonly accepted method for the analysis of fricatives (Gordon et al., 2002; Żygis & Hamann, 2003; Hamann & Avelino, 2007). It refers to a measure, expressed in Hertz, encompassing the description of the weighted average of the energy scattered on all the frequencies of the spectrum – in which frequencies that appear more frequently play a major role in calculating the value itself (Kwakkel, 2008). In a nutshell, CoG condenses the spectral shape of a fricative in a single value. This measure was found to be reliable in distinguishing fricatives by their place of articulation (Gordon et al., 2002): for instance, if the fricative is produced with forward constriction (i.e. alveolar), the cavity formed between the bottom side of the tongue and the teeth would be smaller, resulting in higher frequencies leading to a higher CoG. Nonetheless, it appears that CoG measures are to be used with caution, as language-specific studies (Toda; Gordon et al., 2002) or interpersonal variation (Żygis & Hamann, 2003) could undermine the statistical significance of such value in the discrimination of fricatives.

Spectral peak analysis represents another description method for fricatives. Spectral peaks are recognisable as the highest amplitude peaks in the spectral analysis, and their position in the spectral envelope (i.e. the shape of the spectrum) is relative to the highest intensity frequency bands: for instance, posterior fricatives generate more noise in the lower spectral bands, resulting in visible peaks in the spectrographic view (Gordon et al., 2002). The most prominent of those (that is, the one with the highest amplitude) is considered the main peak, and its position in the frequency range determines the frontness or backness of the articulation of the fricative. However, across a number of the same phonemic tokens, the frequency of the main peak may fluctuate: in this case, fricatives are better distinguished by taking into account two or more peaks (as demonstrated in the analysis of the Polish voiceless retroflex fricative by Nowak, 2006), measuring both their intensity and the distance from one another.

As discussed before, CoG seems to be the most reliable measurement used for the analysis of fricatives. However, as we have seen, it is not always totally reliable in discriminating acoustically similar fricatives – a claim that is bolstered by Scholtz (2009), who proposes to still use CoG values, but to fine-tune the analysis of fricatives with the aid of spectral peak measurements. In fact, in her study, Scholtz (2009) took into the consideration *four* spectral peaks to discriminate /ç/, /ʃ/, and /ʂ/.

This study needs to be able to employ such measurements for manipulation purposes. CoG values are indicative of a sum of frequencies, but they cannot be manipulated as such, if not through the editing of selected frequency bands: as a consequence, changing those bands means to change the entire CoG value of a fricative as well. On the other hand, spectral peaks can be manipulated both by changing their position in the spectral envelope and by editing their intensity to make them more or less salient.

The next two sections shall present the speakers who lent their voices for the experiment, and the editing process used on fricatives. Visual examples shall exemplify aforementioned features.

3.2. Speakers

The speakers employed in the experiment had to comply to a very specific set of requests: they needed to be aware of their manner of speech and to be able to modify it at will, so to adopt features typical of two different age ranges, young and old (e.g. higher presence of dialectalisms for the “old” voice, the usage of UEN-like personal pronouns and interrogative words for the “young” voice), to be employed each in two experimental conditions (see §3.4. for in-depth details). Additionally, they needed to distinguish the two fricatives – palatal and retroflex – in their own pronunciation. These requirements made people working in Linguistics – students or professionals – the perfect candidates for this endeavour.

The recordings came[7] in the form of two speakers, each giving two short presentations of the speakers under pseudonyms (see §3.4.), declaring 24 and 40 years of age each, along with two repetitions of a minimal pair.

The minimal pair, *kjære* 'dear' and *skjære* 'to cut', was chosen among all the other /ç/ - /ʂ/ minimal pairs listed in Palombella (2018) because it was the only one to have the most similar frequency in the NoWaC corpus[8] (4676 entries for the former, 10295 for the latter).

The speakers were provided to us by Professor emeritus Wim van Dommelen (NTNU – Norges Teknisk-Naturvitenskaplige Universitet), and they are:

- F1, A.L.F.: 23 years old, born in Grong, living in Trondheim
- M1, E.H.: 32 years old, born in Lundamo, living in Trondheim

The recordings were made at the NTNU of Trondheim. The speakers were recorded on disk in the sound-treated studio of the Department of Language and Literature, using a Shure KSM44A microphone (44100 Hz sampling frequency; 16-bit quantisation), with no further post-processing apart from the deletion of irrelevant parts.

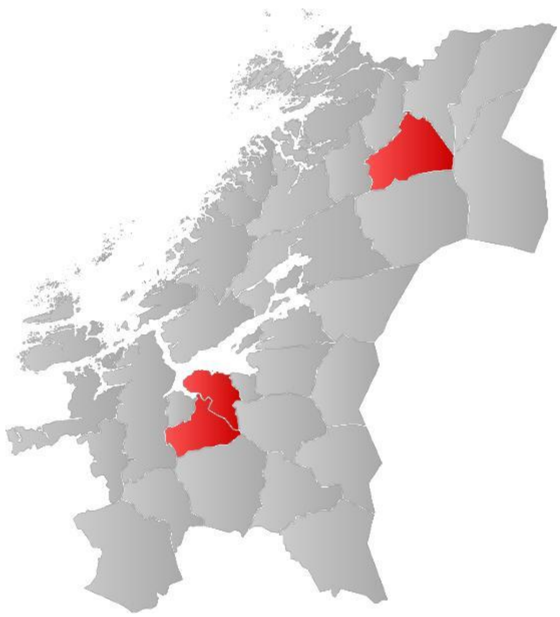


Figure 2. Map of the Trøndelag county. The Grong municipality is the north-eastern red portion of the map; the red cluster in the centre of the map represents Lundamo – in the Melhus municipality – in the South, and the city of Trondheim in the North (source: Wikipedia, adapted by Iep Bergsma).

[7] The data collection should have taken place on the field: due to the 2020 COVID-19 travel bans, it had to be deputised to external helpers – namely Professor van Dommelen, to whom we are extremely grateful for the collection and the theoretical inputs, who gathered samples for this study among his students and collaborators, sending us the recordings of two speakers utilised here.

[8] Norwegian Web as Corpus v1.0 (2010), 700 million tokens, used under advice of the Text Laboratory Centre of the University of Oslo.

3.3. Data manipulation

Every step of the manipulation employed the software *Praat* (Boersma & Weenink, 2020).

The first step required the extraction of the fricatives from the minimal pairs, cutting the fricatives themselves from the vowels following them at their zero crossing. This resulted in four different tokens for each speaker, two voiceless retroflex fricatives and two voiceless palatal fricatives. Their spectral contour is shown below.

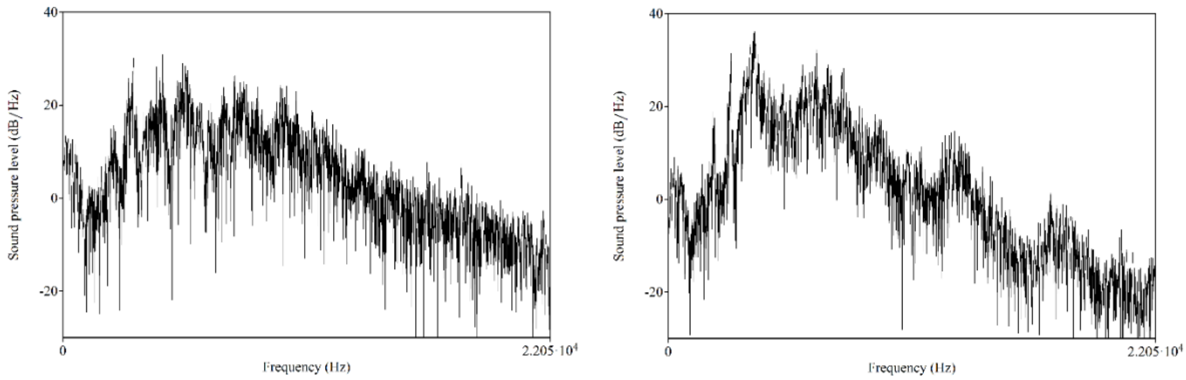


Figure 3. Female speaker, 1st minimal pair. L: palatal, R: retroflex.

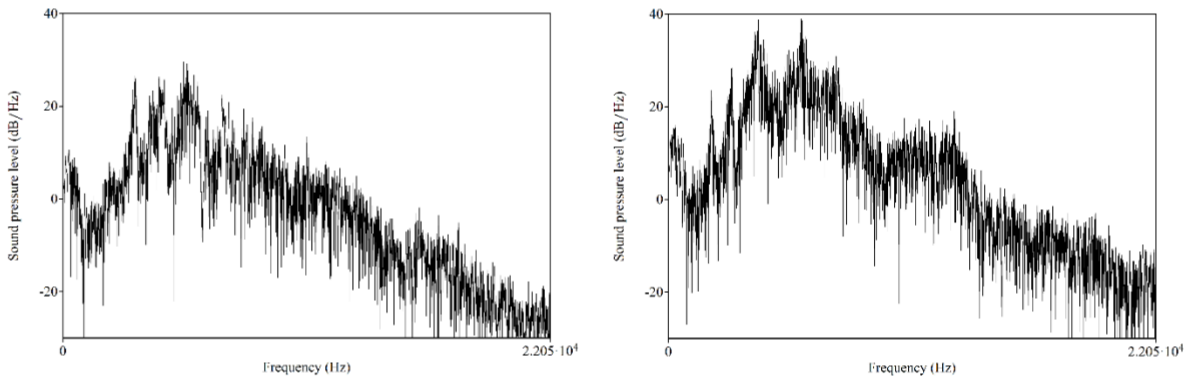


Figure 4. Female speaker, 2nd minimal pair. L: palatal, R: retroflex.

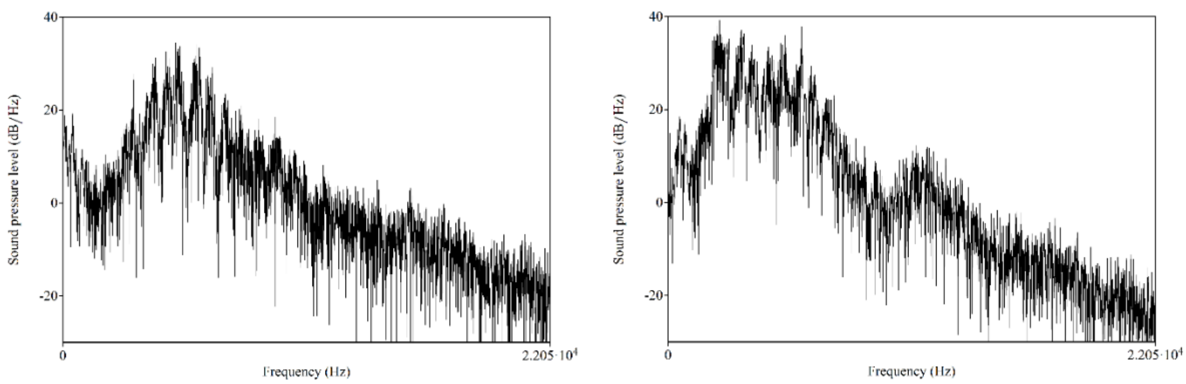


Figure 5. Male speaker, 1st minimal pair. L: palatal, R: retroflex.

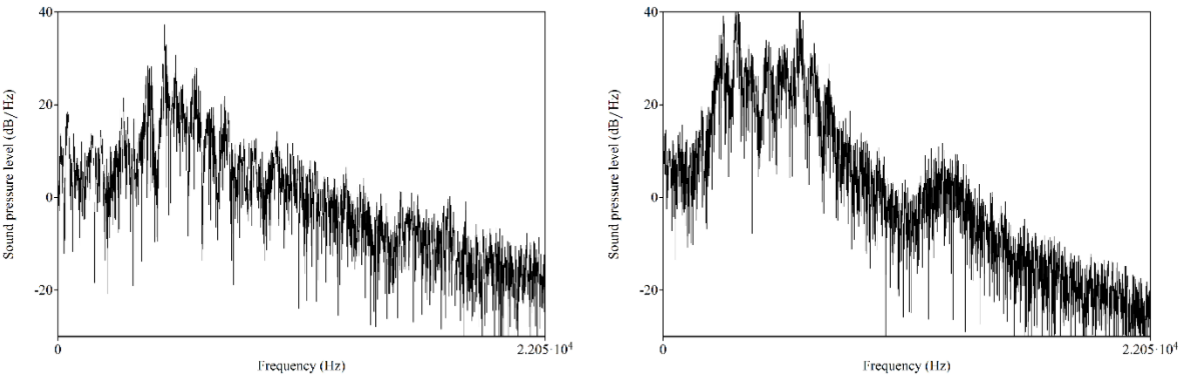


Figure 6. Male speaker, 2nd minimal pair. L: palatal, R: retroflex.

The comparison by juxtaposition of the spectra was essential to pinpoint salient differences in the spectral envelope of retroflexes and palatals. As seen in the figures above, voiceless palatal fricatives tend to have three frequency bands with a higher amplitude in the lower part of the spectrum, creating three spectral peaks. The rightmost part of the spectrum – i.e. the higher frequencies – constitutes a regular slope. On the other hand, voiceless retroflex fricatives have only two high intensity peaks in the lower part of the spectrum, followed by a plateau in the spectral region adjacent to the right of the second peak, followed by a downward slope. Peaks were numbered starting from the highest one in amplitude, from the lowest frequencies (left) to the highest ones (right) – cf. figures 3-6. Their position in the spectral band and their intensity are reported in Table 1.

		P1		P2		P3	
		Hz	dB	Hz	dB	Hz	dB
F1	ç_1	3203	30.0	4501	30.7	5638	29.8
	ç_2	3262	26.8	4363	26.1	5454	29.1
	ʂ_1	3921	35.9	6733	32.0		
	ʂ_2	4070	27	6001	38.8		
M1	ç_1	4665	32.1	5105	33.5	6030	33.8
	ç_1	4825	36.8	5340	30.7	6268	28.3
	ʂ_1	2334	38.8	6040	37.7		
	ʂ_2	3407	41.2	6197	40.8		

Table 1. Frequency and intensity of the peaks.

The CoGs of the tokens is as it follows:

		Palatal /ç/	Retroflex /ʂ/	Difference
F1	MP1	6341	5115	1226
	MP2	5388	5450	62
		Retroflex /ʂ/	Palatal /ç/	Difference
M1	MP1	3742	5409	-1667
	MP2	4407	5146	-739
		Avg. /ç/	Avg. /ʂ/	Avg. difference (absolute value)
		5571	4678.5	892.5

Table 2. Centre of Gravity of the unedited tokens. MP: Minimal Pair. Values expressed in Hz.

As anticipated in §3.1., CoG alone is not a sufficient measure to discriminate between fricatives. This is bolstered by the data in Table 2, from which two observations can be drawn: the first minimal pair always has a higher distance, while the gap is significatively smaller in the second minimal pair, with the singular case of the second minimal pair produced by F1 in which the two tokens have a reversal – in fact, the CoG of the retroflex fricative is higher in frequency than the palatal's. Causes and consequences of such differences will be analysed in the discussion section.

Additionally, referring back to figures 3-6, it can be noticed how the spectral shapes of the second minimal pair produced by F1 (figure 4) are much more regular (i.e. with more precisely identifiable features). For this reason, we have chosen to use the voiceless palatal and retroflex fricatives belonging to this specific minimal pair to build our set of tokens. Using exactly the same friction noise across the two speakers eliminates undesirable differences created by the difference in the gender of speakers[9].

In an experiment such as this, it would have been desirable to have one single continuum ranging from one phoneme to the other; however, in our case, that would have been exceedingly challenging. In fact, the voiceless retroflex fricative belongs to the class of *sibilant* fricatives. The sibilance cannot be deducted by the CoG, but fricatives belonging to this category, upon closer scrutiny, generally entail a more vigorous energy distribution in the highest frequencies (Ladefoged & Maddieson, 1996) and, while we cannot directly witness this as true by our CoG

[9] In a cross-linguistic study on fricatives, Gordon et al. (2002) report statistical significance between genders only in one language out of seven. Therefore, we can assume that a female fricative pasted onto a male carrier word does not create hindrances for the experiment.

values alone (palatals tend to be higher, cf. Table 2), it can still be seen how the second peaks of the extracted retroflexes have a generally higher frequency than the third peaks of the palatals. Additionally, the presence of the plateau is a direct testimony of a higher intensity scattering of energy in the highest part of the spectrum, which palatals lack in full. Due to this fundamental difference in the nature of the two fricatives, one cannot artificially create a token that stands in the middle of the two fricatives, as if it were the product of a merger by approximation. Nonetheless, what could be done was the creation of *two* continua, each having one of the fricatives at one end and trying to approximate the shape of the opposing phoneme in five steps.

T1 - /ç/	T2 - /ç/	T3 - /ç/	T4 - /ç/	T5 - /ç/
T1 - /ȿ/	T2 - /ȿ/	T3 - /ȿ/	T4 - /ȿ/	T5 - /ȿ/

Table 3. Representation of the two continua. T1 is the original phoneme, T2-4 are edited token.

The creation of the two continua was accomplished by the manipulation of three features for every token: *frequency movement* (the editing of a frequency band so to move it in the lower or higher part of the spectral envelope), *number of peaks*, *height of the plateaux.*

The first tokens (henceforth referred to as T1s) of each continuum have been left nearly untouched: the only form of manipulation they have undergone was trimming away frequencies with a similar spectral shape across tokens. To do this, a pass Hann band filter has been applied with the following values:

- T1 - /ç/ – lower bound: 1200 Hz; higher bound: 14450 Hz; smoothing value: 100 Hz.
- T1 - /ȿ/ – lower bound: 920 Hz; higher bound: 14600 Hz; smoothing value: 100 Hz.

Starting from these initial tokens, the continua have been built as follows.

3.3.1. Palatal continuum

Each token (T2-5) started as a copy of the palatal T1. To approximate the spectral shape of their retroflex counterparts, we had to move the spectrum higher in frequency (i.e. to the right), lower the intensity of the first peak (to approximate the two peaks of the retroflexes by reducing one of the three peaks as much as possible) and raise the plateau proper of a retroflex articulation.

This approximation needed to have four values extracted from the contrastive phoneme, the retroflex P1: its position in the spectrum (gathered by the position of its P1, used as an “anchoring point”), the average intensity value of the frequency band corresponding to the palatal P1 across all retroflex tokens (so to know to what extent the P1 had to be lowered), the position and intensity of the retroflex plateau (so to be able to isolate and raise the intensity of the spectral band to create a plateau from scratch). Additionally, the frequencies of the palatal P1 were needed to understand which frequency band needed to have its intensity manipulated. These values are as follows:

- Retroflex P1 frequency: 4070 Hz
- Retroflex P1 average intensity: 23.25 dB
- Palatal P1 position in T1: 3070-3390 Hz
- Retroflex plateau frequency: 8870-13610 Hz
- Retroflex plateau intensity: 13.3 dB

The first peak of the palatal T1 measures 3262 Hz. Its value ought to be subtracted from 4070 to obtain the difference in the distance of the peaks, and the result divided by four – the number of edited tokens following T1.

$$(4070-3262)/4=202$$

The first peak also measures 27 dB in intensity. The difference of barely 3.75 dB with the average intensity of 23.35 dB in the 3070-3390 Hz band measured among the retroflexes was deemed too small: thus, in order to accentuate the volume decrease and make it more audibly noticeable, it was decided to decrease the intensity by 10 dB to reach 17 dB in the final token (T5). The difference of 10 dB was divided by four, thus dictating a decrease of intensity of 2.5 dB per token.

The average intensity of the plateau in the palatal token was 5.4 dB, rising to 13.3 dB. Their difference was divided by four in order to obtain the values of organic increment of intensity to transition from the former intensity to the latter.

$$(13.3-5.4)/4=1.975$$

The values used for each token are summarised in the table below.

	Frequency gain	P1 position	P1 intensity	Plateau position	Plateau intensity	CoG (Hz)
T1	0 Hz	3070-3390 Hz	27 dB	8870-13610 Hz	5.4 dB	5421.519
T2	+202 Hz	3272-3592 Hz	24.5 dB	9072-13812 Hz	7.375 dB	5826.575
T3	+404 Hz	3474-3794 Hz	22 dB	9274-14014 Hz	9.35 dB	6257.240
T4	+606 Hz	3676-3992 Hz	19.5 dB	9476-14216 Hz	11.325 dB	6719.237
T5	+808 Hz	3878-4198 Hz	17 dB	9678-14418 Hz	13.3 dB	7189.454

Table 4. Palatal continuum, detailed values per token.

See Appendix B for the spectra of each token.

3.3.2. Retroflex continuum

The editing process to create the retroflex continuum was much like the one used for the palatal tokens. T1 was used as the base for all the other tokens. However, in this case, the manipulation called for the P1s to be raised and the plateaux to be flattened out to fit a descending slope shape as much as possible. To approximate the spectral contour of the opposing phoneme, the palatal fricative, the same measurements as the previous section have been taken into account:

- Palatal P1 frequency: 3262
- Palatal P1 average intensity: 26.8
- Retroflex P1 position in T1[10]: 2630-3090 Hz
- Retroflex plateau position: 8870-13610 Hz
- Palatal plateau intensity: 5.4 dB

As it was for the process described in the previous section, the difference in peaks between the palatal and the retroflex T1 yields a result of 202 Hz. This time, the spectral envelope of the fricative must be moved down by four steps (in T2-5) to approximate the position of the palatal P1. Additionally, the intensity of the retroflex P1 is 26.8 dB, whereas the intensity measured in the 2630-3090 frequency band across the palatal tokens was of 37.4 dB. The two intensities

[10] Although here this is labelled as a P1, we know from the spectral envelope of the retroflex that it is actually a "phantom" peak, i.e. a portion of the spectrum that needs to be raised to create a peak. In that sense, we are measuring the intensity floor to understand by how much it needs to be multiplied to reach the target ceiling; the same goes for "palatal plateau intensity", as there are no plateaux in the palatals – this is just needed to have a reference point for the target frequency band and intensity used to lower the existing plateau in the retroflex token.

have been averaged as 27 and 37 dB respectively, and their difference was divided by four, resulting in four 2.5 dB increase steps.

For the plateau, refer back to the measurements in the previous section. The intensity of 13.3 dB needed to be lowered to 5.4 dB; thus, using the same formula presented above, the former value has been decreased to the latter by steps of 1.975 dB.

The values used for each token are summarised once again in the table below.

	Frequency gain	P1 position	P1 intensity	Plateau position	Plateau intensity	CoG (Hz)
T1	0 Hz	2630-3090 Hz	27 dB	8870-13610 Hz	13.3 dB	5464.576
T2	-202 Hz	2428-2888 Hz	29.5 dB	8668-13408 Hz	11.325 dB	5249.425
T3	-404 Hz	2226-2686 Hz	32 dB	8466-13206 Hz	9.35 dB	5044.587
T4	-606 Hz	2024-2484 Hz	34.5 dB	8264-13004 Hz	7.375 dB	4841.442
T5	-808 Hz	1822-2282 Hz	37 dB	8062-12802 Hz	5.4 dB	4642.376

Table 5. Retroflex continuum, detailed values per token.

See Appendix B for the spectra of each token.

3.3.3. Building the continua

The modified fricatives were labelled as follows: PAL/RET_T1-5, with the first part signalling the type of fricative and the second part carrying the numeration used in the steps above.

The second minimal pair uttered by the male and female speakers was then imported into *Praat*; its fricatives were cut off at the zero crossing with the vowel and replaced by the tokens, thus altering the original words into the appropriate carrier words. The labelling for each continuum was changed to: M1/F1_PAL/RET_T1-5, following the conventions established for above, but adding the speaker code at the beginning.

The manipulated word tokens were saved as separate .wav files, (44100 Hz sampling frequency; 16-bit quantisation).

3.4. Experimental structure

The experiment was hosted online, on the Pavlovia platform, after having been built on PsychoPy and converted to JavaScript[11]. Participants have been recruited online – especially using the online social news aggregator and discussion forum *Reddit*[12] – and the final count was that of 140 subjects gathered.

Participants were greeted by a landing page informing them of the technical details of the experiment – such as privacy disclaimers and storage consent – but keeping the aim of the experiment confidential. Upon acceptance of such terms and conditions, they were sorted by entrance order into two separate conditions, so to have the odd participants partaking in Condition A, and the even ones in Condition B.

As already stated, these two conditions were designed following the precise requirement to check the inference of speaker age in phonemic perception and categorisation. As reported in §3.2., each condition contained both speakers presenting themselves under pseudonyms ("Berit Skogen" for F1 and "Bjørn Sørgård" for M1): Condition A consisted of "Berit Skogen" presenting herself as 24 years old, followed by "Bjørn Sørgård" presenting himself as 40 years old and conversely, Condition B saw "Bjørn Sørgård" as the first speaker, declaring 24 years of age, followed by "Berit Skogen" presenting herself as a 40-year old. This choice was made to put emphasis on the age difference between the speakers, which is thought to exert pressure on the choices of the participants: as mentioned in §2.3., the Labovian method makes use of trying to raise feelings in the subject, and this is exactly the aim here.

After the first speaker's introduction, the participants were presented with tokens in a semi-randomised order (cf. Appendix C). The tokens were presented on separate slides, one at a time, and they were only playable once; the slides also contained the name, age, and provenance of the speakers as their headers, as a constant reminder to the participants – with the headers as a covert device of conditioning, designed to invoke an age bias into our participants. Since the tokens consisted of a single word whose fricative was edited, the discrimination process was made easy for the participants, who were confronted with a binary choice based on two orthographic forms – *kjære* and *skjære*[13] – in the shape of two clickable buttons under the audio clip of the token. The choice was made available only upon listening to the audio clip of the token, and moving forward was not possible until a choice was made. As the first set of tokens

---

[11]Courtesy of Dirk Jan Vet, Speech Lab, Laboratory of Phonetic Sciences, University of Amsterdam.

[12]The forums ("SubReddits") the experiment was posted in were: r/norge, r/norsk, r/Norway, r/ntnu, r/SampleSize.

[13]With small differences between Condition A and Condition B. In the former, *kjære* is presented on the left and *skjære* on the right; in the latter, *skjære* is on the left and *kjære* is on the right.

uttered by the first speaker finished, the second speaker introduced themself, followed by a second series of token presented in the same way as the first round.

At the end of the two parts, we decided to implement a final page with three optional closing questions, whose results are not factored in the statistical analysis, but considered useful to have a picture of the participants' demographics and thoughts. The first question was: "What did you think the experiment was about?". This was followed by a blank textbox in which participants could write their hypotheses. Two more questions required to pick from a drop-down menu. The first one asked about the age range (possible choices: *below 18, 18-25, 26-30, 31-35, 36-40. 41-45, 46-50, above 50*) of the participants: categories have been implemented because the exact age would have marked a more fine-grained distinction, but at the cost of cluttering the data. The provenance of the participants was the subject of the second question, whose possible choices were: *Oslo, Rogaland, Møre og Romsdal, Nordland, Viken, Innlandet, Vestfold og Telemark, Agder, Vestland, Trøndelag, Troms og Finnmark* or *I live abroad*. Those questions marked the end of the test.

Of all the participants partaking in the experiment[14], not all data points have been saved. This discrimination was based on the file size of the result files. Those with a size smaller than 7 KB were found to be incomplete – perhaps because of the abandonment of the experiment before its end. Eventually, the present study could count on 140 complete data points.

The deletion of incomplete files inherently created an imbalance in conditions, as people who abandoned the experiment page were still sorted according to the condition system. Fortunately, this imbalance is almost negligible, as we have gathered 68 data points for Condition A and 72 data points for Condition B.

3.5. Statistical analysis

The master data sheet (see Appendix A) was systematised as follows: each participant was represented twice (once per speaker age): for each appearance, they were assigned two scores, respectively related to the palatal tokens and the retroflex tokens. The scores represented how many times the participants matched what was heard to the corresponding orthographic form. Since participants were exposed to five palatal and five retroflex tokens per speaker, the score was first calculated as matching instances out of five, then converted to a percentile score. Thus, an excerpt of the data file would look as follows:

[14] The exact number has not been noted down. Cf. below for the rationale.

participant	condition	age	palscore	retscore
2	A	Y	60	80
2	A	O	80	80

Table 6. Excerpt of the data file. Y equals 24 years of age, and O corresponds to 40.

The two different outcomes for the palatal and the retroflex word tokens called for two separate models, identical to one another but with the score factor changed to fit the two fricatives. We have chosen to analyse the data through a linear mixed-effects model, having the score predicted by the interaction of the age and the condition (to account for possible variation in the data: certain speakers might trigger a bigger age effect than other ones), and the speaker as a random slope. Age is a random slope too, but it cannot be used as a random slope per participant, as there is only one value per speaker.

Ultimately, the model was built as follows:

```
(pal/ret)score ~ age * condition (1 | participant)
```

3.6. Results

The most important average values are those of the average percentile recognition scores for palatal and retroflex tokens for age label and for condition. These are as follows:

Condition	Age	Fricative	Score
A	Y	PAL	96.47%
A	Y	RET	92.35%
A	O	PAL	93.52%
A	O	RET	86.76%
B	Y	PAL	87.50%
B	Y	RET	87.22%
B	O	PAL	90.27%
B	O	RET	87.50%

Table 7. Percentile answer scores for palatal and retroflex tokens between conditions and between speakers.

3.6.1. Statistical reports

The reports of the models read as follows.

For the palatal model, the percentage of palatal answers given by the participants (weighed equally heavily over the four subpopulations) is 91.00% (95% confidence interval: 88.28…93.73). The palatal answers for the speakers labelled as "young" are higher than those for the speakers labelled as "old" by 0.23% (95% confidence interval: -3.90…4.37), but not statistically significant (t = 0.11; p = 0.91). The palatal answers given by the participants sorted in Condition A are significantly higher those sorted in Condition B (t = -2.02; p = 0.04). Therefore, we conclude that participants in Condition A were better at discerning palatal fricatives than the participants sorted in Condition B (estimated difference: 5.62%; 95% confidence interval: 0.18…11.07). Finally, the estimated amount by which speakers labelled as "young" get worse discerning results than speakers labelled as "old" is higher in Condition A than in Condition B (estimated difference: 6.58%; 95% confidence interval: -1.68…14.85), but this interaction is not significant (t = -1.55; p = 0.12).

For the retroflex model, the percentage of retroflex answers given by the participants (weighed equally heavily over the four subpopulations) is 90.40% (95% confidence interval: 87.50…93.30). The retroflex answers for the speakers who were presented with the label "young" are higher than those for speakers labelled as "old" by 1.92% (90% confidence interval: -2.16…6.00), but not statistically significant (t = 0.92; p = 0.35). Participants sorted in Condition A scored higher (estimated difference: 3.30%; 95% confidence interval: -2.49…9.10) than those in Condition B, but the effect of Condition is not significant (t = -1.11 p = 0.24). Lastly, the estimated amount by which speakers labelled as "young" get worse discerning results than speakers labelled as "old" is higher in Condition A than in Condition B (estimated difference: 4.39%; 95% confidence interval: -3.78…12.57), but this interaction is also not significant (t = -1.05; p = 0.29).

3.6.2. Results per token

Despite the lack of statistical significance across models, it is still interesting to visualise the scores for the same token uttered by the young and the old speaker, compared to each other.

In order to do this, four graphs have been created. Two of them (Fig. 7-8) contain the results of Condition A, two of them (Fig. 9-10) those of Condition B. The X-axis considers the declared age of speakers across tokens; the Y-axis, on the other hand, shows how many matches the participants have scored on said conditions – that is, on each age parameter per token. These graphs will be discussed in §4.

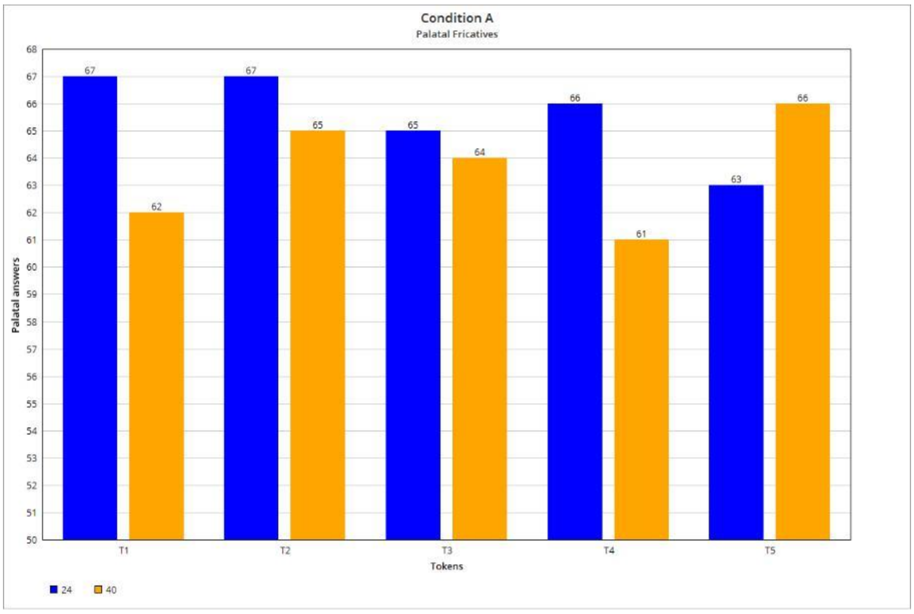


Figure 7. Palatal fricatives in Condition A: score comparison across young (F24) and old (M40) speaker.

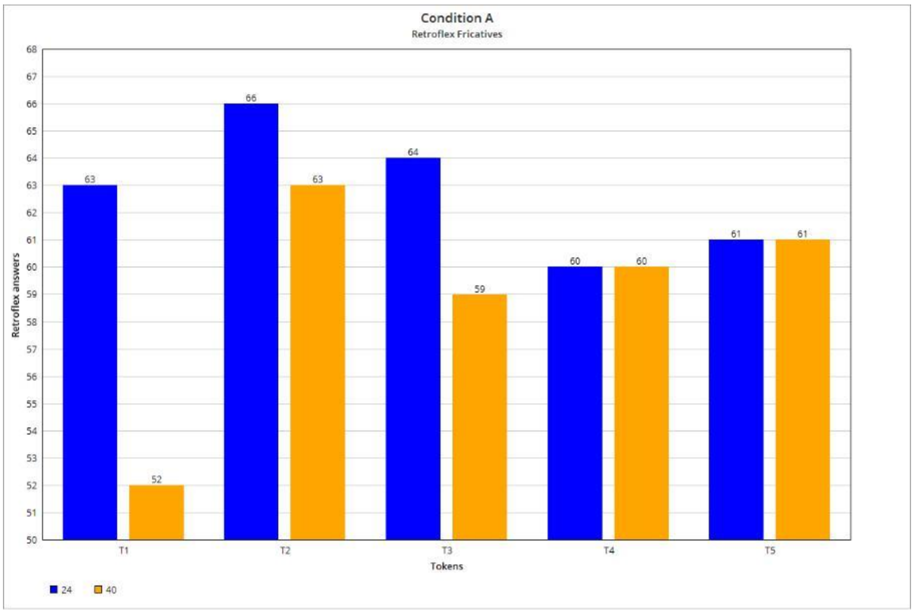


Figure 8. Retroflex fricatives in Condition A: score comparison across young (F24) and old (M40) speaker.

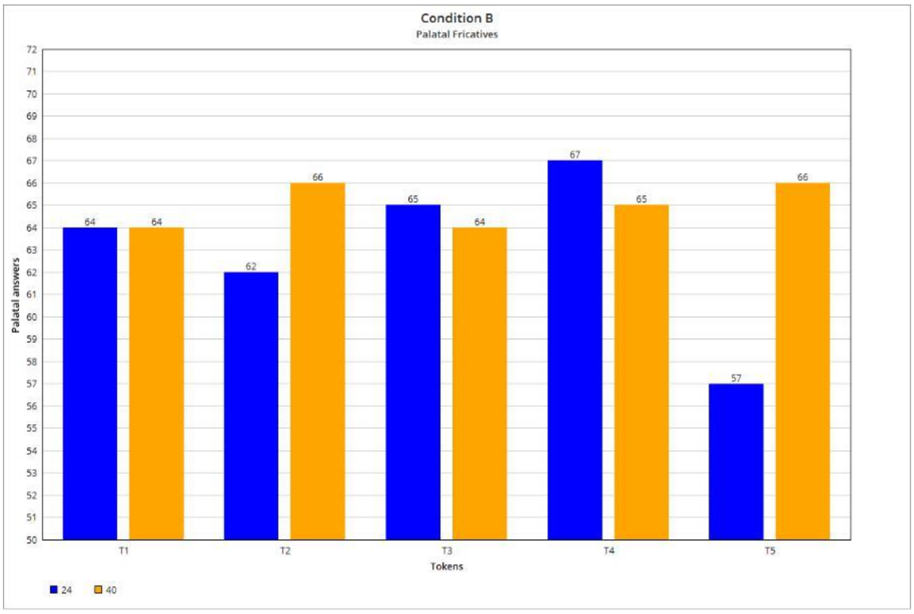


Figure 9. Palatal fricatives in Condition B: score comparison across young (M24) and old (F40) speaker.

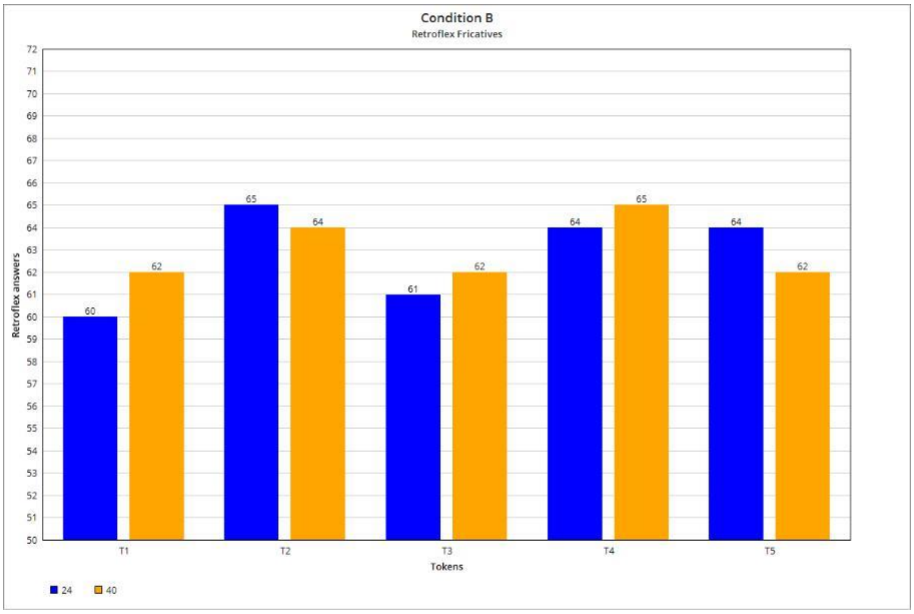


Figure 10. Retroflex fricatives in Condition B: score comparison across young (M24) and old (F40) speaker.

3.6.3. Demographic results

Results about the participants' demographics have been left outside of the statistical analysis. Nonetheless, they are necessary to comment on the results of the experiment, as both the age and the provenance of the participants can add up to the observations necessary to unravel the (mostly) null results.

The first graph visualises the age range of the participants, organised in groupings of 5 years each; participants below 18 years of age and above 50 have been categorised independently and are shown on either side of the graph. The results are as follows:

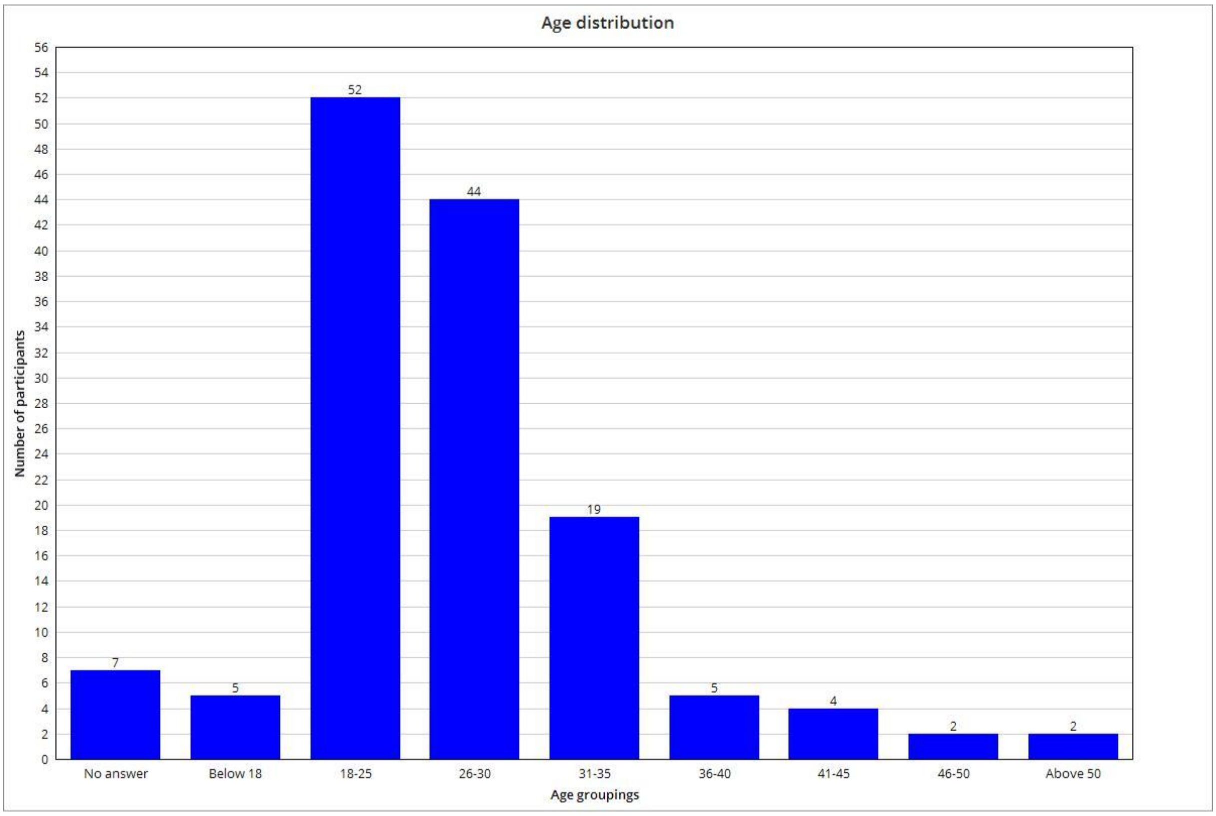


Figure 11. Visualisation of the age groupings of the 140 participants.

The demographics fit the means used to gather participants, as we can assume that later age groupings entail less familiarity with the platform used (*Reddit*).

The geographical distribution – better accompanied by a visual comparison with Figure 1 – is as follows:

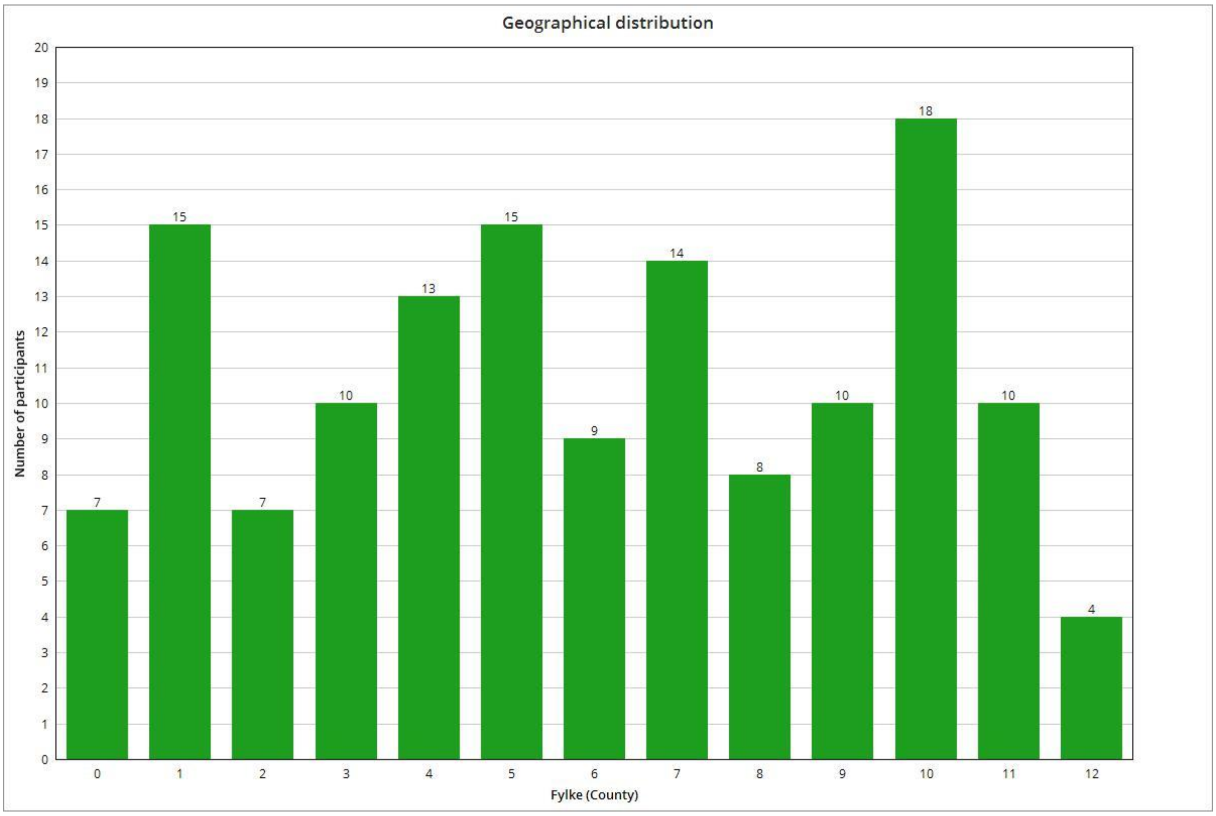


Figure 12. Visualisation of the geographical distribution of the 140 participants by county. *Reading key*: 0. No answer; 1. Oslo; 2. Rogaland; 3. Møre og Romsdal; 4. Nordland; 5. Viken; 6. Innlandet; 7. Vestfold og Telemark; 8. Agder; 9. Vestland; 10. Trøndelag; 11. Troms og Finnmark; 12. The participant lives abroad.

4. Discussion

4.1. General discussion

At the beginning of the present study, we hypothesised that a higher occurrence of merged utterances in the younger speakers would mirror higher rates of perception of the merger if a participant listens to a speaker labelled as "young". The results of our experiment proved us wrong: participants exposed to two speakers labelled as "young" and "old" showed more difficulties in discerning the pronunciation of the two fricatives – the voiceless palatal /ç/ and the voiceless retroflex /ʂ/ – in older speakers. The following section discusses these findings, giving tentative explanations to such phenomenon; additionally, we discuss further observations about the imbalance between fricatives in the answers, the statistical difference in answers across conditions, and anecdotal observations stemming from the participants' comments.

We start out by discussing the graphs reported in Figures 7-8. These graphs were built to check a decline in palatal and retroflex answers for the appropriate continua from T1 to T5. Figure 7 shows a promising behaviour in the response to palatal tokens that, with the exception of T3, seem to degrade the further the tokens stray from the original, unedited phoneme. However, this is a one-off instance: in the same figure, retroflexes seem not to follow any pattern at all – which is even more accentuated in Figure 8, with the first token of the retroflex continuum getting a very low score. The lack of patterns is persistent throughout figures, and the sole observation emerging from this is: the two continua do not behave as such. We do not get under any circumstance a linear degrading of perception departing from T1, which leads to think of a fault in the token editing process. Suggestions for possible corrections are given in §4.2.

It is hard to discuss statistical relevance when the results are largely null, especially if statistical significance is not equally spread across the two fricatives. In fact, significance is only found in the statistical analysis of the palatal fricatives, and it accounts for a higher percentage of palatal answers in condition A – which does not happen for the retroflex fricatives. Age and interaction between condition and age are not significant in either of the two models. The less-than-significant results for the predictors that we hypothesised to be crucial make it difficult to gauge any influence of fricative perception, and challenging to interpret the results in a univocal way.

At the current state, we cannot report anything about the status of prestige. Our statistical results have not found a significant predominance of palatal or retroflex answers, and the fact that there *is* significance showing a higher rate of palatal answers in Condition A is not mirrored by any significant result in the retroflex model.

Of course, the report of the full palatal model (see Appendix A) shows how participants gave 0.23% more palatal answers for speakers labelled as “young”, and one could hypothesise a tendency from the subjects to apply the standards of social correction (as described in Sandøy, 2013) to what they heard – and consequently to perceive palatals more often than not, even without the benefit of a frame sentence to be contextualised in; however, we simply do not have statistical significance to back up this theory.

Another hypothesis to consider about what led to null results could be that of a *fatigue/training effect* over the course of the experiment. What is meant by this, is that participants could have got either worse (fatigue effect) or better (training effect) as they heard tokens in sequence.

To obviate to this and to be able to establish patterns, one must check the global results of the experiment in the same order as the one followed by the participants:

Condition A											
		T1_P	T1_R	T5_R	T3_P	T2_R	T4_R	T5_P	T3_R	T2_P	T4_P
	F24	67	63	61	65	66	60	63	64	67	66
		T3_R	T1_R	T3_P	T4_P	T2_R	T4_R	T5_P	T5_R	T2_P	T1_P
	M40	59	52	64	61	63	60	66	61	65	62
Condition B											
		T3_P	T1_R	T1_P	T2_R	T3_R	T4_P	T5_R	T4_R	T2_P	T5_P
	M24	65	60	64	65	61	67	64	64	63	57
		T5_R	T4_R	T5_P	T4_P	T1_R	T3_P	T2_P	T2_R	T3_R	T1_P
	F40	62	65	66	65	62	64	66	64	62	64

Table 8. Answers per token, according to the actual token order. Palatal and retroflex have been shortened to P and R for editing purposes.

As seen in Table 8, there is no clear pattern to ascribe to token recognition. The abovementioned statistical analysis reports as significant only a worsening in discerning palatals across conditions, with better recognition scores in Condition A. This too, despite having statistical significance, is hard to spot as a pattern in the table above.

Additionally, it can be seen how the results are stable (that is, with scores fluctuating, but consistently over 60 points) for both speakers in Condition B, with the sole exception of the last token heard from the first speaker. The reason of this stand-out has two possible explanations: the fatigue effect of which above, or a lower recognition rate as it is the last token of the palatal series, that is, the one whose editing diverges the most from the original palatal token (T1). However, both hypotheses are debatable. For the former, participants had a break after the last token from the first speaker (that is to say, the presentation of the second speaker) leading to no mirroring of this effect in the second half of the experiment, with the conclusive token of Speaker 2 scoring impressively higher (64). For the latter, if one takes a look at the T5s across speakers and conditions, it is clear to see how each and every token this kind scored above 60 points – which in turn leads to believe that the editing of the token has a relatively small effect on the recognition task. Conclusively, we rule out the effect of training or fatigue in our experiment.

A second discussion prompt on the null results comes from van Dommelen (2019: 4), as mentioned in §2.4.1. According to said study, younger speakers who distinguish the two fricatives have a tendency to hyperarticulate them in order to obtain a starker contrast, compared to a closer pronunciation of the two fricatives by older speakers. Palatal answers across young speakers are generally high (above 60 points), which leads to believe that subjects could have had a subconscious awareness of sorts towards this phenomenon, thus giving the younger speakers in our experiment the benefit of doubt, even under the suspicion that younger speakers might not produce "canonical" (that is, palatal /ç/) realisations at all. One reason for this could be an adaptation of the Ganong effect, presented in Ganong (1980): in a lexically biased continuum presenting the choice between a word and a non-word, even in the presence of an auditory non-word stimulus, participants would feel compelled to categorise what is heard as a meaningful word. Ganong (1980) makes use of the VOT values to explain this phenomenon for stop consonants. As mentioned above, we prefer to postulate an *adaptation* of such effect for two reasons: the absence of a non-word in our continuum – making it a non-lexical phenomenon – and the absence of stop consonants in the onset. Thus, to explain the phenomenon, one can posit that the distinguishing feature used as a cue between the two fricatives is *sibilance,* present in the retroflex and absent in the palatal. It is entirely possible that subjects frequently exposed to the /ç/-/ʂ/ merger developed cognitive mechanisms geared towards "removing" sibilance in perception – when context calls for it – perceiving a word like *kjøtt* 'meat', whose merged pronunciation would be /ˈʂœtː/, as if it had the canonical pronunciation /ˈçœtː/ even in the presence of the merger.

In this case, one would observe a division between *phonemic perception* and what can be called *social perception*, in which the latter overrides the former. In fact, it can be posited that the mechanisms of phonemic perception *do* register a degree of ambiguity in the sounds heard, or even a mismatch between what is heard and the listener's underlying representation of the word: however, even though it would be categorised "correctly" (i.e. matching the actual phoneme), the subjects experience a correction in the course of their perception because of a significative unconscious (but prolonged) training in discerning phonemes and categorising them according to their canonical pronunciation in real life situations – even when the words are given without a frame context, such as in our experiment.

The last factor to cover is the comment section left by our participants as a result of the questionnaire presented at the end of the experiment. Even though it represents nothing more than anecdotal evidence, it can still shed some light on the thoughts and opinions of the participants. When asked about the aim of the experiment, most participants opted for a nondescript answer such as: "distinction between the kj- and the skj- sound", worded in several different ways. Among them, some reported an alleged inability to differentiate between the two, while others had a clear attitude of either resignation or open disapproval of the "disappearance" of the voiceless palatal fricative /ç/ – which once again cements the hypothesis of it as the highest prestige phoneme when compared to a retroflex /ʂ/ merged pronunciation (cf. §2.5.). Another good number of participants seemed to think the experiment to revolve around the perception of the two phonemes in question by speakers of different dialects – impression that was possibly reinforced by the question about provenance at the end. Age as a predictor is mentioned one time (Participant 117), but its use was not interpreted as intended: according to the participant, the goal was: "[e]xamining whether one can differentiate between the "kj" and "skj" sounds in Norwegian, possibly relating to age and dialect of sender [that is, the participant]". One participant (45) poignantly answered: "[h]onestly it sounds like you're mocking the Trondish and I can stand behind that", which seems to comply with what described in Scholtz (2009: 14), i.e. a negative bias towards the Trøndersk dialect, often perceived as ugly or vulgar.

4.2. Suggestions for further research

As it happens for research with null results, there are a number of factors which could alter the result. The following section contains suggestion on how to improve the outcome of future studies.

The first and most important suggestion would be that of recruiting more speakers. Having more people per condition, covering a wider range of age and opening the possibility of having 15-16 year old speakers would be ideal, as more speakers represent a better chance to pinpoint patterns in perception that otherwise would be hard to spot with two speakers only – as it has been for the present study. Results might be further shaped if future studies took into account both the age of the speaker and that of the participants, and built a statistical model showing the interaction between the two.

Secondly, even though most participants seemed unaware of the true aim of the experiment, it would be advisable to add some distractors to the word sequences. It might be hard to find a suitable distractor to fit the experiment, but Scholtz (2009) is once again useful: her study revolved not only around the palatal-retroflex fricative merger, but it also took into consideration other ongoing mergers in the Norwegian language, e.g. the merger between the voiced retroflex lateral approximant /ɭ/ and the voiced coronal[15] lateral approximant /l/, which would represent a fitting choice.

The extralinguistic information about the speakers was given in a written form, constantly reminding the participants by the means of a header containing the name, age and provenance of the speaker they were hearing. Perhaps this external tentative inference was not influential enough, so we would recommend to find other means to remind the participants about the identity of the speaker, such a visual means (pictures). This could trigger a bigger reaction in categorising phonemes according to a much more invasive identity suggestion.

Lastly, we would suggest a more fine-grained token manipulation process. This study made use of ample frequency bands, with the results that tokens could only edited to a certain extent. If such bands were divided further in smaller portions of the spectral contour, their manipulation in terms of spectral properties and the change in dB values would be more precise. Furthermore, creating artificial fricatives could lead to having more regular spectral shapes, which in turn

---

[15] Disputed: we are following Kristoffersen (2000: 38).

would make the editing process (to bridge the gap from palatal to retroflex and vice-versa) easier.

5. Conclusions

The aim of this thesis was to check the inference of age in the perception mechanisms surrounding the merger between the voiceless palatal fricative /ç/ and the voiceless retroflex fricative /ʂ/. In §1.3. we asked how much social expectations can manipulate perception, and how much can extralinguistic given information guide phonemic categorisation in one way or another.

We proceeded to hypothesise that, given the circumstances of production involving the two fricatives, a subject hearing a younger speaker would be more likely to hear a merged pronunciation – that is, a word canonically containing a voiceless palatal /ç/ being realised with a voiceless retroflex /ʂ/ – irrespective of its actual presence.

The results were surprising. The answer to the research question seems to suggest that, however many extralinguistic information are given to the participants, they had little inference on the end results. Comments seems to indicate that the social expectations on the participants' side were focused largely on the speakers' dialect compared to their own rather than on the age difference (and logically, pronunciation idiosyncrasies) between speakers.

The hypothesis was subverted too, as participants seemed to discriminate the pronunciation of young speakers better than that of the older ones. We posit that higher recognition rates have to do with the implicit knowledge that younger speakers distinguishing the two fricatives tend to hyperarticulate them to mark a stronger distinction.

In the light of the null results, we have given advice to improve the experimental design by adding more speakers, more distractors, and a different way to socially condition the participants.

In a collective effort to shed light on the details of the Norwegian linguistic panorama, we hope for future studies to take ours as a baseline and to provide conclusive results on the perception of the palatal-retroflex merger.

References

Aitchinson, J. (2001). *Language Change. Progress or decay?* New York (NY): Cambridge University Press.

Boersma, P. & Weenink, D. (2020). *Praat: doing phonetics by computer [Computer program].* Version 6.1.16, retrieved 6 June 2020 from http://www.praat.org/.

Bonfiglio, T.P. (2002). *Race and the Rise of Standard American.* Berlin (DE): Walter de Gruyter.

Dalbakken, L.O. (1996). *Distinksjonen kje/sje i lydendringsperspektiv. En empirisk undersøkelse av barns og unges beherskelse av distinksjonen kje/sje i Trondheim.* Master's Thesis. Oslo (NO): Universitetet i Oslo.

van Dommelen, W.A. (2019). Is the voiceless palatal fricative disappearing from spoken Norwegian? In: Calhoun, S., Escudero, P., Tabain, M. & Warren, P. (eds.). *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra (AU): Australasian Speech Science and Technology Association Inc.

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech 54/1, pp. 99–121*. Thousand Oaks (CA): SAGE Publishing.

Ganong, W.F. III (1980). Phonetic Categorization in Auditory Word Perception. *Journal of Experimental Psychology: Human Perception and Performance 6/1, pp. 110-125*. Washington (DC): American Psychological Association.

Gordon, M., Barthmaier, P., Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association 32/2, pp. 141-174.* Cambridge (UK): Cambridge University Press.

Hamann, S.R. (2003). *The Phonetics and Phonology of Retroflexes.* PhD Thesis. Utrecht (NL): LOT Publishing.

Hamann, S.R. & Avelino, H. (2007). An acoustic study of plain and palatalized sibilants in Ocotepec Mixe. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany 2007.* Saarbrücken (DE): Universität des Saarlandes.

Hansen, L.I., Olsen, B. (2004). *Samenes historie fram til 1750.* Oslo (NO): Cappelen Damm Akademisk.

Harris, J. (1985). *Phonological Variation and Change: Studies in Hiberno-Irish.* Cambridge (UK): Cambridge University Press.

Hay, J., Drager, K. (2010). Stuffed toys and speech perception. *Linguistics 48/4, pp. 865-892.* Berlin (DE): Walter de Gruyter.

Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania.* PhD Thesis. Philadelphia (PA): University of Pennsylvania.

Hudson, G. (2000). *Essential Introductory Linguistics*. Hoboken (NJ): Blackwell Publishers.

Hudson, R. (1981). Some Issues on Which Linguists Can Agree. *Journal of Linguistics 17/2, pp. 333–343*. Cambridge (UK): Cambridge University Press.

Hårstad, S. (2010). *Unge språkbrukere i gammel by. En sosiolingvistisk studie av ungdoms talemål i Trondheim*. PhD Thesis. Trondheim (NO): NTNU.

Hårstad, S., Opsahl, T. (2013). *Språk i byen. Utviklingslinjer i urbane språkmiljøer i Norge.* Bergen (NO): Fagbokforlaget Vigmostad & Bjørke AS.

Jahr, E.H. (1990). Dialekter og dialektbruk. In Jahr, E.H. (ed.). *Den store dialektboka*. Oslo (NO): Novus.

Jacobsen, A.S. (2015). *"Vi ungdomma sir jo sånn /ʃ/ i stedet for /ç/". Ei sosiolingvistisk sammenligning av fonemutviklinga av /ç/ og /ʃ/ i Tromsø og Trondheim.* Master's Thesis. Tromsø (NO): UIT, Norges arktiske universitet.

Johannessen, S.H. (1983). Om "skjendisar" og "chipsreiarar". Bruken av sje-lyd of kje-lyd i bergensmålet. In Sandøy, H. (ed.). *Talemål i Bergen*. Oslo (NO): Novus.

Kristoffersen, G. (2000). *The Phonology of Norwegian.* Oxford (UK): Oxford University Press.

Kwakkel, V. (2008). *Sibilants in Dutch and English: typologies, learning tasks and language change*. Master's Thesis. Utrecht (NL): University of Utrecht.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia (PA): University of Pennsylvania Press.

Labov, W. (2010). *Principles of Linguistic Change, Volume I: Internal Factors.* Oxford (UK): Blackwell.

Ladefoged P., Maddieson, I. (1996). *The Sounds of the World's Languages.* Oxford (UK): Blackwell.

LAPSyD, Lyon-Albuquerque Phonological Systems Database. Retrieved from: http://www.lapsyd.ddl.cnrs.fr/lapsyd/index.php. Date of access: 23/07/2020.

Liberman. A.M., Harris, K.S. Hoffman, H.S., Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology 54, pp. 358-368*. Washington (DC): American Psychological Association.

Moen, I., Simonsen, A.G. (2011). The Norwegian Retroflex Fricative. *Proceedings of the 16th International Congress of Phonetic Sciences, Hong Kong 2011*. Hong Kong (HK): City University of Hong Kong.

Mæhlum, B., Røyneland, U. (2012). *Det norske dialektlandskapet.* Oslo (NO): Cappelen Damm Akademisk.

Nesse, A. (2013). *Innføring i norsk språkhistorie.* Oslo (NO): Cappelen Damm Akademisk.

Nowak, P.M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics 34, pp. 139-152.* Amsterdam (NL): Elsevier.

Palombella, E. (2018). The Norwegian Voiceless Palatal Fricative: A Demographical Study. *Bollettino del Laboratorio di Fonetica Sperimentale "Arturo Genre", pp. 21-26.* Turin (IT): Università degli Studi di Torino.

Papazian, E. (1994). Om sje-lyden i norsk, og ombyttinga av den med kje-lyden. *Norskrift: Avdeling for nordisk språk og litteratur 83*. Oslo (NO): Universitetet i Oslo.

Rød, H.M. (2014). *Molde bymål. Ein kvantitativ korrelasjonsstudie i verkeleg tid.* Master's Thesis. Bergen (NO): Universitetet i Bergen.

Sandøy, H. (2013). Driving forces in language change – in the Norwegian perspective. In Kristiansen, T., Grondelaers, S. (eds.). *Language (De)standardisation in Late Modern Europe: Experimental Studies, pp. 125-151*. Oslo (NO): Novus.

Scholtz, A. (2009). *A Phonetic Study of the Status of Three Mergers in the Trøndersk Dialect of Norwegian*. Bachelor's Thesis. Williamstown (MA): Williams College.

Shen, Z. (1990). Lexical diffusion: A population perspective and a numerical model. *Journal of Chinese Linguistics 18, pp. 159-200.* Berkeley (CA): University of California.

Simonsen, H.G., Moen, I. (2004). On the distinction between Norwegian /ʃ/ and /ç/ from a phonetic perspective. *Clinical Linguistics and Phonetics 18/6-8, pp. 605-620.* Milton Park (UK): Taylor & Francis.

Simonsen, H.G., Moen, I. (2006). Fonetisk perspektiv på sammenfallet av sj-lyden og kj-lyden i norsk. *Spraknytt 1, pp. 11-13, 28*. Oslo (NO): Språkrådet.

Skjekkeland, M. (1997). *Dei norske dialektane.* Kristiansand (NO): Høyskoleforlaget.

Skjekkeland, M. (2005). *Dialektar i Noreg: Tradisjon og fornying.* Kristiansand (NO): Høyskoleforlaget.

Skjekkeland, M. (2010). *Dialektlandet.* Oslo (NO): Cappelen Damm Portal Forlag.

Torp, A., Vikør, L.S. (2010). *Hovuddrag i norsk språkhistorie.* Oslo (NO): Gyldendal Akademisk Forlag AS.

Trudgill, P., Foxcroft, N. (1978). On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. In Trudgill, P. (ed.). *Sociolinguistic Patterns in British English, pp. 69-79*. London (UK): Edwin Arnold.

Vanvik, A. (1979). *Norsk Fonetikk: Lydlæren i standard østnorsk supplert med materiale fra dialektene.* Oslo (NO): Fonetisk Institutt, Universitetet i Oslo.

Venås, K., Skjekkeland, M. (2019). *Dialekter i Norge*. Retrieved from the Store norske leksikon. Last update: 13/05/2019. Date of access: 23/07/2020.

Żygis, M., Hamann, S.R. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In Solé, M.J., Recasens, D., Romero, J. (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 2003, pp. 395-398.* Barcelona (ES): Universitat Autónoma de Barcelona.

Appendix A

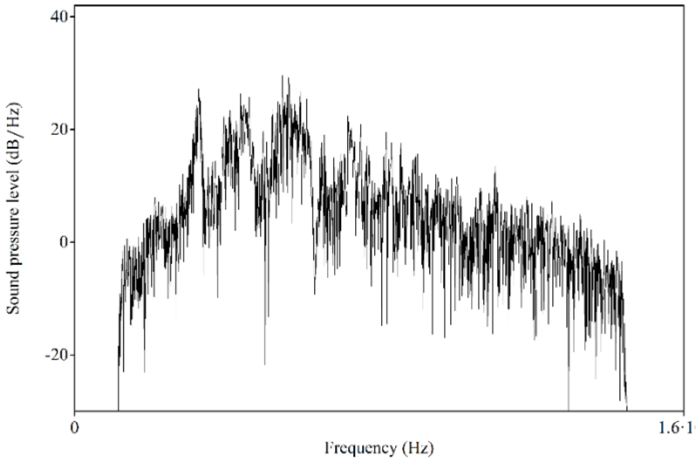
In order to guarantee transparency and facilitate reproducibility, we have decided to create a permanent, publicly accessible repository containing the original audio files, the manipulated tokens, the instructions given to build the experiment, the result files, the master spreadsheet containing all the answer, the condensed spreadsheet and the text file used to run the statistical models and the R markdown containing the model (both in .Rmd and .html format).

Please, note that while the repository will remain accessible, the result files will be deleted after 5 years from the publishing of this thesis – that is, the 25th of June, 2025 – in compliance with the agreement presented to the participants.

Click here to open the Archive of the Institute of Phonetic Sciences (IFA) of the University of Amsterdam.

Appendix B

This appendix contains the spectra of the edited token. Palatals and fricatives are juxtaposed, while the rows of images go from Token 1 (T1) to Token 5 (T5).



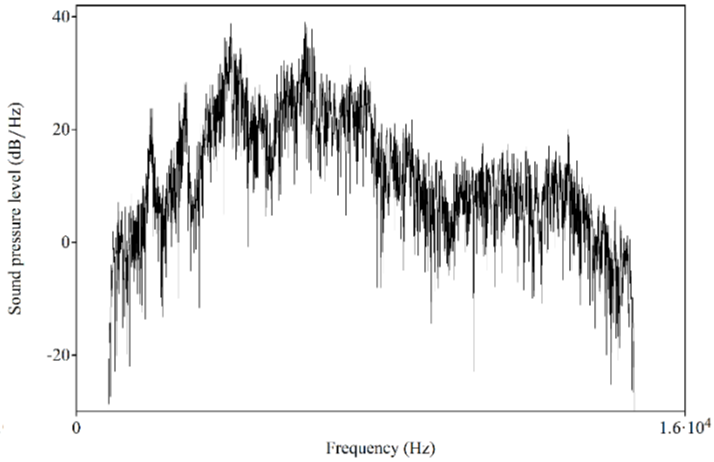
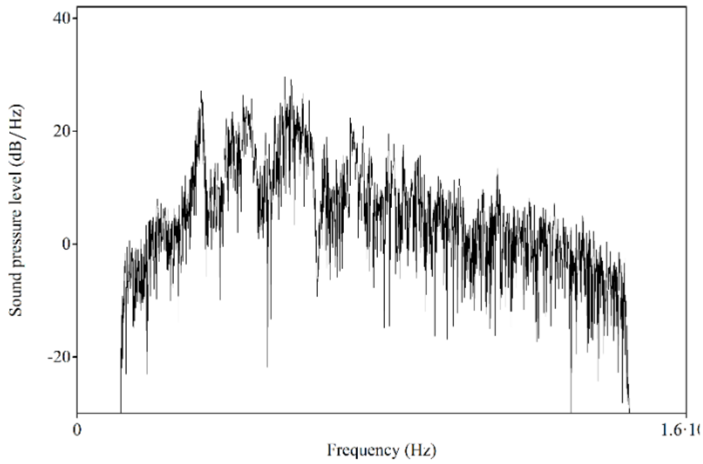


Figure 13-14. Palatal T1 (left); retroflex T1 (right).



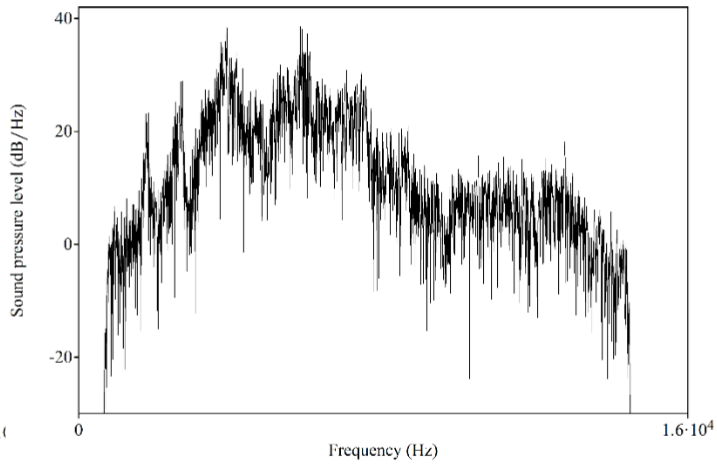
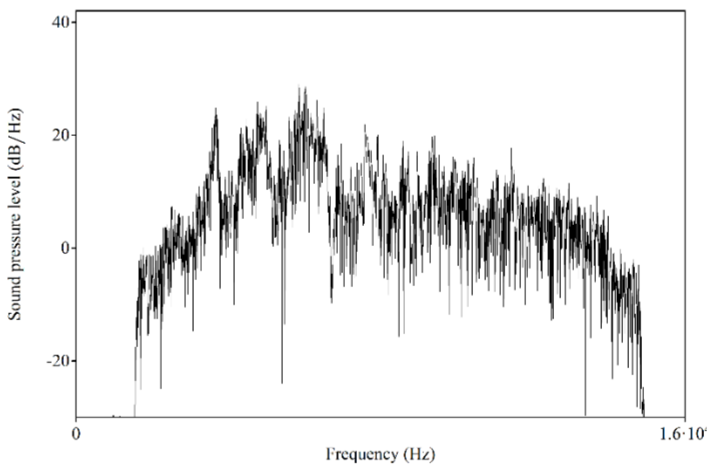


Figure 15-16. Palatal T2 (left); retroflex T2 (right).



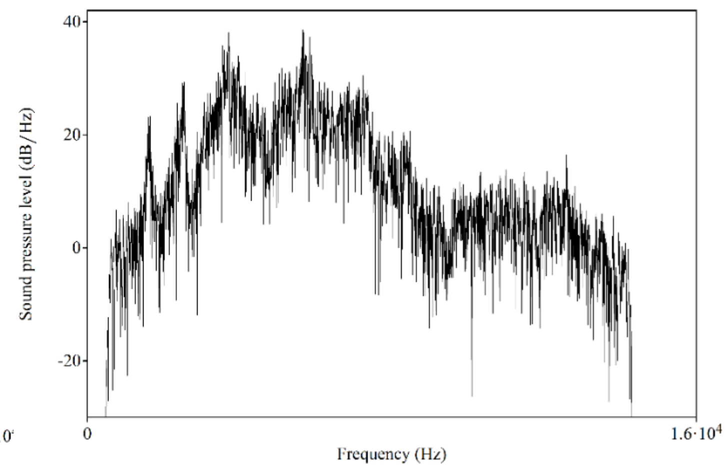
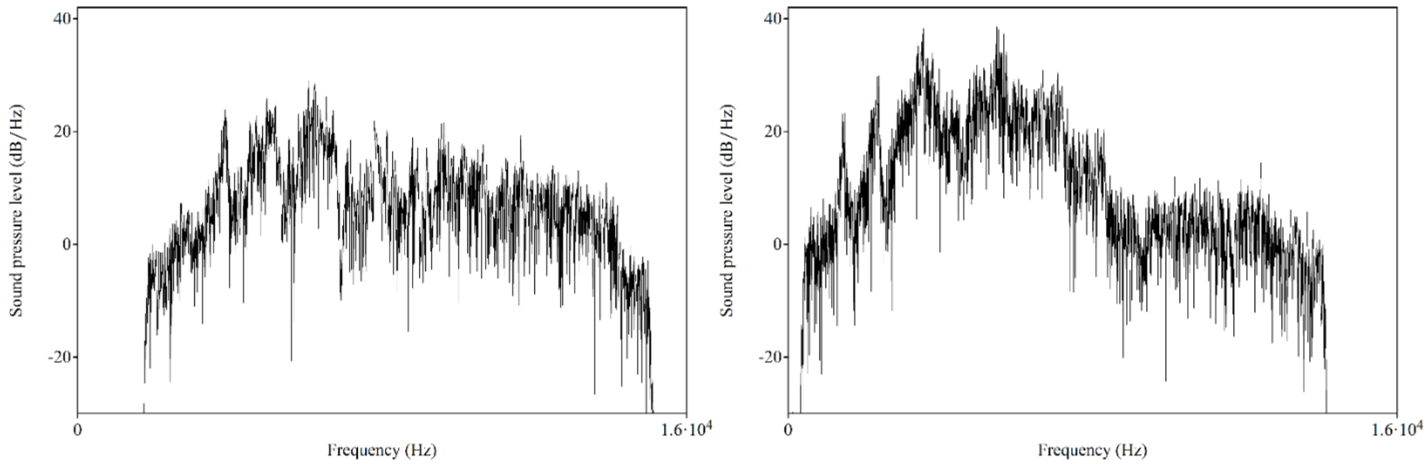
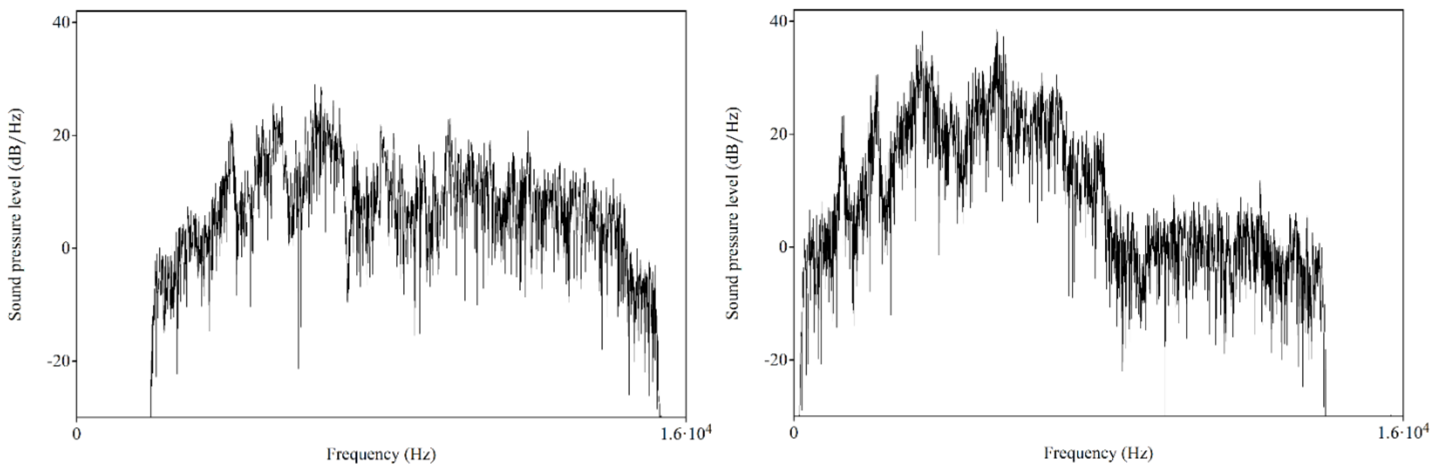


Figure 17-18. Palatal T3 (left); retroflex T3 (right).



Figures 19-20. Palatal T4 (left); retroflex T4 (right).



Figures 21-22. Palatal T5 (left); retroflex T5 (right).

Appendix C

In the following appendix we give the semi-randomisation order of the tokens as it appeared in the experiment.

SLIDE	CONDITION I		CONDITION II	
	F1_24	M1_40	M1_24	F1_40
1	F1_T1_PAL	M1_T3_RET	M1_T3_PAL	F1_T5_RET
2	F1_T1_RET	M1_T1_RET	M1_T1_RET	F1_T4_RET
3	F1_T5_RET	M1_T3_PAL	M1_T1_PAL	F1_T5_PAL
4	F1_T3_PAL	M1_T4_PAL	M1_T2_RET	F1_T4_PAL
5	F1_T2_RET	M1_T2_RET	M1_T3_RET	F1_T1_RET
6	F1_T4_RET	M1_T4_RET	M1_T4_PAL	F1_T3_PAL
7	F1_T5_PAL	M1_T5_PAL	M1_T5_RET	F1_T2_PAL
8	F1_T3_RET	M1_T5_RET	M1_T4_RET	F1_T2_RET
9	F1_T2_PAL	M1_T2_PAL	M1_T2_PAL	F1_T3_RET
10	F1_T4_PAL	M1_T1_PAL	M1_T5_PAL	F1_T1_PAL

Table 9. Order of the semi-randomised tokens[16].

16 The randomisation was achieved through the list randomised retrieved from this website.