Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

Liv-Randi Lersveen

The perception and production of nonnative English consonants in native Norwegian speakers

Master's thesis in Linguistics

Trondheim, May 2018

NTNU
Kunnskap for en bedre verden

Abstract

This study looked at the perception and production of the English unvoiced and voiced alveolar stops, alveolar fricatives, postalveolar fricatives and affricates. The voiced sounds in each of these pairs are nonnative to Norwegian speakers, except for the alveolar stops which are both present in the Norwegian sound system. Both a perception test and a production task were performed. Two groups of native Norwegian speakers participated. Group 1 consisted of people who had not spent more than a maximum of 6 weeks in an English-speaking country, and group 2 consisted of people who had lived in an English-speaking country for a period of time (4-10 months). A control group, consisting of native speakers of English also conducted the same experiments.

The perception task had an AXB design, where the listeners task was to identify which of the two words in a minimal pair word X was equal to; e.g. looking at the contrast /s/ and /z/ by using the minimal pair <ice> and <eyes> and playing <eyes> as word X. Different English-native speakers had produced the stimuli which consisted of 18 different minimal pairs, which contrasted in the target sounds. In the production task, the stimuli were presented both orthographically and by audio one by one, and participants were instructed to read the word out loud after hearing the stimulus. The recordings from the production task were then judged by two native speakers on a 5-point scale, where the sound in question was rated from 1-*wrong sound* to 5- *native-like*. The raters were blind to the hypothesis.

The study hypothesized that the frequency of minimal pairs containing the contrasting sounds would influence the results of the L2 groups in both the perception experiment and the production experiment. It was also hypothesized that time spent in English-speaking country would have an effect in both experiments. Contrary to the expectations, no effect was found on the influence of time spent in English-speaking country in either the perception or production results. The results showed an effect of the frequency of minimal pairs containing the contrasting sounds in the perception experiment, but not the production experiment.

Acknowledgements

I would like to thank both my advisers, without them this project would not have been possible. I would like to thank my adviser Anne Dahl for answering questions, for helpful feedback on drafts, and for always being full of optimism. I would like to thank my adviser Wim van Dommelen for giving feedback faster than what should be humanly possible, and for always wanting to discuss any questions I had concerning my thesis, big or small.

I would like to thank all the people who helped make this study possible. The project could not have been completed without the people who participated in the experiments, produced stimuli, or rated the recordings.

I would like to thank my parents, who have always had faith that I could do anything I put my mind to. I would especially like to thank my mom for always picking up the phone whenever I called to complain about my thesis.

I would like to extend a big thank you to all my fellow master's students for the five years leading up to this moment. A special thank you goes to Inger Anne who has helped keep my head straight during the process of writing this thesis, and Siri for sitting together with me the final days before submission. I would like to thank Sigrid for taking her time and helping me proof-read my thesis. You saved me many typos and double-spacings. Any errors that remain are totally my own.

Trondheim, May 14th 2018 Liv-Randi Lersveen

Content

1.0 Introduction

The limits of my language
mean the limits of my world".
Ludwig Wittgenstein (1889 – 1951)
(Wittgenstein and Ogden, 2014, p. 149, Satz 5.6)

The world is getting smaller and smaller. One reason for this is all the new technology, but another assumed important reason is the access of a world language. Today, anyone who can make themselves understood in English can travel and communicate with locals. There are several factors which affect the intelligibility of a second-language user, and one of these is the mastering of critical differences within the phonology of the second-language.

This thesis addresses second language acquisition of phonology. Little work has been done on this topic earlier with respect to first-language users of Norwegian and English as a second language, and my goal is to investigate further into the issue of acquiring nonnative phonemes in English.

The topic is especially interesting to look at in Norwegian speakers because most adult Norwegians are competent users of English, and they are fluent in most situations. My goal is to find out if there are any differences in the performance in high competent users of English as a second-language (L2) when you control for native input. The research will also situate itself among the models that exist for second-language acquisition of phonology.

This thesis therefore looks for differences between two groups of Norwegian speakers. One group of persons who had been on exchange to an English-speaking country at some point and for a minimum of four months were recruited. In contrast, the other group was made up by persons who had not been on exchange to an English-speaking country. A control group consisting of native speakers of English also completed the same experiments.

In chapter 2 I outline scientific works that are relevant to my research question and discuss second-language acquisition (SLA) models that have formed the basis for much of the current research on SLA concerning the acquisition of second-language phonology. Following, I present theory on the sound systems of Norwegian and English, and research that have been conducted on Norwegian speakers with concerning English a L2. Chapter 3 describes the methods used and gives reasoning for the methodological choices. Chapter 4 addresses the results of both a perception test and a production task. The results are discussed in chapter 5, and in chapter 6 I discuss possible implications of the results and give a conclusion.

2.0 Theory

2.1 Terms and definitions

The term "novel phoneme" refers to phonemes which are present in the second language (L2) and not in the first language (L1) of two languages. In this case, English is the L2 and Norwegian is L1. Any shared phonemes between those languages will be referred to as shared or common phonemes. The term *Phonetic inventory* refers to all the phonemes present in a language's phonology or sound system. The view on how humans store speech sounds can generally be divided in two;

1. Language consists of phonemes and each phone can be described by its features. A language has a set of phonemes which is then stored in categories based on its features (Flege, 1995, p. 239).
2. The sounds of Language consist of different *articulatory gestures,* and the speaker stores information about speech based on the *articulatory gestures*. Different *articulatory gestures* can be at play when the same phoneme is produced (Catherine and Louis, 1986, p. 225).

There is no standardized way of referring to speakers as either experienced or inexperienced in the literature about second-language acquisition, so this makes it hard to generalize across different studies where those terms have been used. But in short, the terms are often used to divide a group of L2 speakers based on native input or duration of stay in a country where the L2 is the native language.

2.2 Previous research on SLA phonology

Flege (1993) conducted a study on second-language learners of English, looking at novel phonemes. They selected participants who were native English-speakers, native Taiwanese-speakers who had learned English during childhood, and late learners from the following categories; Mandarin inexperienced, Taiwanese inexperienced and Taiwanese experienced. Experience was categorized based on duration of residency in the US. These groups were chosen because Taiwanese has word-final /p t k/, and not /b d g/, while Mandarin do not have stops in word-final position (Flege, 1993). The study looked at the production and perception of word-final /t/ and /d/ in a CVC environment.

There were four experiments in total and they consisted of both perception and production tasks. The production experiment (1) was made to test whether the L2 speakers made vowels longer before /d/ than /t/ and perception experiments (2 and 3) were made to check if they used vowel duration as a perceptual clue when separating /d/ from /t/. The final experiment (4) was a speech imitation task. The hypothesis was that production accuracy is limited by the adequacy of the speakers' perceptual representations for L2 sounds and contrasts. The prediction was that the results would differ whether as to if the speaker spoke a language which contained /t/ in word-final position, or if the contrast was not at all present in the speaker's native language. The study looked at the duration of the preceding vowel because it is easily measured and compared across participants (Flege, 1993), and it has been shown to be an essential clue in the interpretation of the voicing contrast in English (Flege and Hillenbrand, 1986).

The results from the production experiment showed that all the participants, both native and non-native, made vowels longer before /d/ than /t/. The results from an ANOVA comparing the average vowel duration differences, showed a significant effect of group. Further, the post-hoc test showed that the difference produced was significantly larger for the native speaker and the childhood learners, than the differences produced by Taiwanese late learners, the inexperienced Taiwanese and the inexperienced Mandarin late learners. The results from the perception experiment were similar to the production experiment. The native speakers and childhood learners showed the same pattern in the effect of vowel duration on perception of word-final /t/. There was a lack of between-group differences, and only inexperienced late learners showed a smaller effect of vowel duration in identifying word-final stops, when being compared to the native speakers.

In a test of *choice of best example*, the native speakers and childhood learners did not differ significantly in their choice of tokens. The comparison of the native speakers and childhood learners against the inexperienced groups, Mandarin and Taiwanese late learners showed that the native and childhood groups preferred a larger difference in vowel duration and the difference between the preferred durations were significant, when compared to the inexperienced groups. The experienced Taiwanese learners showed a larger preferred difference in vowel duration than the inexperienced groups, and a smaller difference than native speakers and childhood learners, but none of the comparisons were significant. In the speech imitation task, only native speakers and childhood learners showed discontinuities in their vowel production which could be attributed to the final stop contrast. The results of the grouped data supported the hypothesis that perception affects production; however, there were individual results that did not conform to this hypothesis.

Iverson *et al.* (2003) looked at how early language experience could influence the acquisition of non-native phonemes. Their participants were adult speakers of Japanese, German, and American English and they looked at the perception of the /r/-/l/ contrast in English. The stimuli used in this study varied across the second (F2) and third formant (F3). F2 is an important acoustic cue in Japanese, while F3 is important to distinguish /r/ from /l/ in in both German and English. The stimuli consisted of English /ra/ -/la/ tokens, and all the non-native speakers of English had studied English in school. The participants were asked to identify and rate the goodness of each stimulus according to their native language phoneme categories. The Japanese speakers identified the stimulus into their /r/ category, but stimulus with lower F2 frequencies were identified as /w/. The German speakers heard each stimulus as either a good exemplar of their /l/ or as a poor exemplar of their uvular fricative and the American English speakers identified each stimulus as either a good exemplar of their /l/ or as a good exemplar of their /r/. Both the German and the English speakers showed had a category boundary which was sensitive to F3. They also had their participants complete a discrimination task, where the stimuli varied in F3 frequencies, but were kept at the same F2 frequency. In this task, both the American English speaking and German speaking participants had the highest discrimination sensitivity when the stimuli being compared contained tokens with F3 frequencies from both sides of the centerline, and comparisons with F3 frequencies from the same side of the centerline had lower discrimination sensitivity. The Japanese speaking participants did not show an effect of this centerline, and their results were more linear, where only higher differences in frequencies of F3 had higher discrimination sensitivity.

Eckman, Iverson and Song (2013) and Eckman and Iverson (2013) have both investigated hypercorrection in SLA. Hypercorrection is when a contrast present in the L1 sound system gets used in the environment where a phoneme present in the L2 sound system would be correct. For hypercorrection to occur the L1 sound system has one phoneme equivalent of one of two contrasting phonemes in the L2 sound system. Eckman and Iverson (2013) states that hypercorrection errors will occur later rather than earlier in the acquisition process, and that transfer errors and hypercorrections can be present in the speaker's language at the same time. In their article Eckman, Iverson and Song (2013) report data which tested two hypotheses, namely that hypercorrection happens late in the language acquisition process and that in order for hypercorrection to occur, the target language contrast must have been acquired by the L2 speaker. They looked at native speakers of Korean, who lived in the US and were 'intermediate' or 'intermediate high' speakers of English. They looked at the acquisition of the following contrasting phonemes; /s/-/š/, /p/-/f/. These pairs of phonemes each contained one nonnative phoneme. They predicted two distinct patterns of errors to be present in the results for the phonemes; /s/-/š/, /p/-/f/. They also looked at /f/-/v/, which both were nonnative phonemes. They found a positive and significant correlation between the performance on the contrast, and the production of hypercorrection errors for the phonemes /s/-/š/. The data from the /p/-/f/ contrast patterned similarly to the data from /s/-/š/, but it was not statistically significant. The data from the production of the /f/-/v/ contrast did not show a pattern as the one found in /s/-/š/. Overall, their results supported their hypotheses.

2.3 Different models of SLA phonology

Many of the models of SLA phonology make similar predictions about how easy or hard it is to acquire speech sounds which are novel to the second language learner. They usually differ in whether they consider speech sounds to be stored as distinct members of phonetic categories in the mind of the language user (Flege, 1995), or whether they condsider the sounds to be saved as a memory of articulatory gestures (Best and Tyler, 2007). The models make many of the same predictions about which sounds the language learner might have trouble with and which error might show up in perception and production.

One of the most prominent second-language speech learning models have been suggested by James Flege. His model, the *speech-learning model* (SLM), is built on a set of postulates from which Flege (1995) draws different hypotheseses. SLM is ment to be applied to L2 users who have considerable experience. The theoy postulates that aspects connected to speech sounds are saved in long-term memory representations called *phonetic categories.* The ability to learn L1 speech sounds remain intact over the lifespan and the phonetic categories within a speaker continue to evolve over the life span so to reflect the properties of all L1 or L2 phones. The model also states that bilinguals strive to keep the phonetic categories of their L1 and L2 separate (p. 239). Flege draws different hypotheses from these postulates conserning second language speech acquisition.

The different hypotheses create testable predictions based on the postulates (Flege, 1995, p. 239). They make predictions about which phonemes will be difficult to acquire based on shared and separate phonetic inventory in L1 and L2. If there is a perceptual difference large enough between the L1 and L2 sound, a new slot will be added to the phonetic inventory to make room for the L2 sound keeping it distinct from the L1 sound. The model hypothesizes that age of learning will play into the probability of the L2 sound receiving its own phonetic category because L1 input will keep reinforcing the established phonetic categories, making their connections stronger. The model thus accounts for results showing age of arrival being related to perceived foreign accent (Flege, Munro and Mackay, 1995)

A competing model emerged from the Perceptual Assimilation model (Best and Tyler, 2007), which originally looked at how naïve listeners perceived non-native sounds (Best, 1995, 1994). The premise for this model was that mature listeners perceive nonnative phonemes based on their gestural similarities to native phonemes (Best, 1994, p. 14). This model assumes that the speaker stores the information about the speech segments based on the articulatory gestures, and not as phonetic categories. Any nonnative sounds are then either assimilated to a native phone, dissimilated and heard as a nonnative sound or it is not classified as a speech sound (Best, 1995, 1994). In Best and Tyler (2007) the model is extended from looking at just naïve listeners to including proficient L2 users. When PAM is applied to L2 perception, it differs from SLM when looking at the mental representation of language specific aspects. PAM states that there are no phonetic categories in the long-term memory; instead the language user becomes tuned to the relevant speech properties. This includes the phonological level, gestural level as well as the phonetic level.

In their predictions both models are quite similar. Both PAM and SLM suggest that L2 users will have more trouble acquiring a L2 phonological contrast when the frequency of minimal pairs which contain the contrast is low. Comparably, a high-frequency minimal pair containing a nonnative contrast will be easier to acquire because of the higher probability of exposure (Best and Tyler, 2007). Both models also suggest that orthography might help the L2 user in classifying sounds and ease the acquisition process. This is because the orthographic representation will make the L2 learner more conscious to the phonological differences in the L2 sounds (Best and Tyler, 2007).

2.4 The sound systems of Norwegian and English

The consonants present in both sound systems share a lot of similarities, which makes the contrasts that are present limited. Both languages make use of the voicing contrast, and they share most of the places and manners of articulation.

Most of the literature on Norwegian phonology agree that Norwegian contains the following unvoiced fricatives, labiodental /f/, alveolar /s/, retroflex /ʂ/, palatal /ç/, and glottal /h/. Some of the literature classifies the /ʂ/ as a postalveolar /ʃ/ (Vanvik, 1979). The different classifications came around because of the different origins of the sound, which is both historical and synchronic (Kristoffersen, 2000). In the historic case, the sound came out of /sj-/ and /skV$_{[+front]}$/ and is realized as [ʃ], and synchronically /rs/ is realized as [ʂ] (Kristoffersen, 2000). Since they do not create a different meaning, and their realization vary between speakers, I choose to transcribe both [ʂ] and [ʃ] as /ʃ/. The phoneme /h/ is classified as a fricative out of convenience, and Kristoffersen (2000) argues that the glottal fricative /h/ should instead be classified as an obstruent or approximant. However, the literature does not agree fully on whether the Norwegian phonetic inventory contains any voiced fricatives. Lundskær-Nielsen, Barnes and Lindskog (2005) states that the voiced fricatives; labio-dental [v], and velar/palatal [j], are a part of the Norwegian phonetic inventory. Vanvik (1979) and Kristoffersen (2000) does not include these, but instead they classify the sounds related to those phones to be an approximant /ʋ/ with no audible friction and a velar/palatal glide /j/. Based on this, I will assume that voiced fricatives are rare and might not even present in the phonetic inventory of Norwegian speakers.

Voiced fricatives are absent in many languages, and they are hard to produce because the vibrating vocal cords impede the flow of air through the vocal tract and high-velocity is needed to produce the turbulent noise characteristic of fricatives (Johnson, 2011, p. 156). In English however, all fricatives have both a voiced and an unvoiced counterpart. The English phonetic inventory contains the following fricatives, unvoiced and voiced labiodental /f/-/v/, unvoiced and voiced dental, /θ/-/ð/, alveolar, unvoiced, and voiced /s/-/z/, unvoiced and voiced postalveolar, /ʃ/-/ʒ/, and unvoiced glottal /h/ (all except /h/ have a voiced counterpart) (Davidsen-Nielsen, 1977, Hammond, 1999).

All stops in both languages occur in unvoiced/voiced pairs and both languages have the following stops, labial /p/-/b/, alveolar /t/-/d/, and velar/palatal /k/-/g/ (Davidsen-Nielsen, 1977, Vanvik, 1979, Kristoffersen, 2000, Hammond, 1999, Lundskær-Nielsen, Barnes and Lindskog, 2005). The Norwegian sound system also has unvoiced and voiced retroflex/apical alveolar /ʈ/-/ɖ/ (Vanvik, 1979, Kristoffersen, 2000, Lundskær-Nielsen, Barnes and Lindskog, 2005). All these stops are unaffricated and they are sometimes referred to as plosives in the literature (Vanvik, 1979, Davidsen-Nielsen, 1977, Hammond, 1999, Lundskær-Nielsen, Barnes and Lindskog, 2005).

Affricates are stops that are fricated, where the stop is released as a fricative (Johnson, 2011, p. 179). What separates an affricate from a two-phone sequence of a stop and a fricative is that the amplitude of frication noise rises quickly to full amplitude in affricates, and more slowly in fricatives (Johnson, 2011, p. 179). English has the two following affricates in its phonetic inventory /tʃ/- /dʒ/ (Davidsen-Nielsen, 1977, Hammond, 1999). Fromkin (1971, p. 33) looked at speech errors and found that these affricates are treated as one phoneme despite being made up by both a stop and a fricative. Looking at how the errors where the affricates were produced in the wrong location, he found that those sounds never split.

Some literature mention /ts/ and /tʃ/ as possible sequences of phonemes in Norwegian, but from what I have been able to find, this literature does not refer to the sequence as one single phoneme in Norwegian (Kristoffersen, 2000, Lundskær-Nielsen, Barnes and Lindskog, 2005, Vanvik, 1979). Skommer (2014) states that there are no affricates in Norwegian phonology, where the plosive and the fricative are realized in the same place of articulation. However, according to Vanvik (1979) the /t/ is realized as a post-alveolar when it occurs directly in front of /ʃ/. The phonetic inventory thus contains the possible phoneme combination that maps onto an unvoiced affricate /tʃ/ in English. To my knowledge, it is uncertain whether Norwegians, who learn English as a second language, treat affricates as separate phonemes or as a single unit. There are some dialectal differences, and you find the two affricates /c͡ç/ and /ɟ͡j/ in some west-coast dialects[1] (Kristoffersen, 2015).

2.5 The language situation in Norway

There are 5.3 million people in Norway and a large variety of different regional dialects are spoken. The country has two standardized written languages; one originating from the Danish written language (*Bokmål*, "book language") and one that has been constructed based on different dialects (*Nynorsk*, literally "new Norwegian"). Neither of the written languages are considered spoken languages. Both written languages have undergone major changes since they originated and the most used one today is *Bokmål* (Vikør, 2017, 17th october). None of the dialects have a formally higher position in the society (Vikør, 2017, 17th october), but Eastern Norwegian is often taught in Norwegian second language courses and it is the dialect that resemble *Bokmål* the most (Lundskær-Nielsen, Barnes and Lindskog, 2005).

In Norway, English is a mandatory subject every year starting first grade (6 years old), until and including first year of high school (16 years old) (Kunnskapsdepartementet, 2013, 2010, 2006a, 1997). This means that all young adults in Norway today have had 10 years of English teaching as a minimum, and there are also optional English subjects in secondary school (grade 8-10) and high school. The English subject has its own curricula, while German, Spanish,

[1] None of the Norwegian participants in this study were speakers of any of these dialects

French and other optional foreign languages have one shared curriculum (Kunnskapsdepartementet, 2013, 2006b).

The reasons for English having a stronger hold than any other foreign languages are diverse. English has had a high position in Norway for several years because Norway is reliant on trade with other nations, 50% of all food is imported (St. meld. nr. 9, 2011–2012, p. 17), and English is the most common lingua franca (Ku and Zussman, 2010). Most English or American movies and tv-shows for teens and adults are not dubbed, but rather they have subtitles in Norwegian. American pop-culture has a strong presence in Norwegian culture and social life. Today, the body of research is rapidly increasing, and most is written in English. Following from this, much of the curricula assigned to students at university level in Norway is written in English and many of the courses are also taught in English. It follows, that most Norwegians are intermediate or fluent users of English as a second-language.

2.6 Research on English as an L2 on Norwegian speakers

The aim of van Dommelen and Hazan (2010) was to investigate factors that influence second language speech perception in different noise conditions. They conducted research on the perception of English consonants in both native English and Norwegian listeners. The Norwegian listeners were all high proficiency users of English as L2. They looked at several consonants (24 in total) and the following English consonants which were novel to Norwegians; /ʒ/, /dʒ/, /w/, /z/, /θ/, /ð/. They did both an identification test in quiet and different noise conditions, and an assimilation test to see what categories the English consonants assimilated to in Norwegian and their similarity to the mapped Norwegian category. The hypotheses were that there would be a greater disadvantage in the noise condition for the L2 listeners and that the perception of the L2 listener would be affected by whether the consonant was shared or novel. The results of the identification test did not support their hypotheses. Their results showed a poorer identification of the novel category consonants in the Norwegian listeners, but the English listeners also showed a poorer identification of the consonants belonging to this category. They found this likely to be related to the acoustic-phonetic properties of the consonants within the novel category. Within the novel category, they found that /ð/ and /θ/ had particularly low identification scores, and this was true for both English and Norwegian listeners. They also found no evidence that there was a greater non-native disadvantage in the noise condition (van Dommelen and Hazan, 2010).

Aleksander Morland (2010) investigated Second language acquisition of English in Norwegians speakers in his master thesis. He looked at the production and perception of the phonemes /t/-/d/ and /s/-/z/ in L1 and Norwegian L2 speakers of English. The phoneme /z/ is novel to Norwegian L2 speakers of English. The participants were native speakers of English, and native speakers of Norwegian with a high L2 proficiency in English. There was conducted both a perception and a production experiment to investigate whether the L1 and L2 speakers used the same cues in perception and production of the four phonemes. He was particularly interested in which cues were used to determine voicing and he looked at vowel/consonant ratio, duration of the vowels and consonants, and the vibration of the vocal folds. These factors were chosen because they have been shown to play into the perception and production of voicing both in English (Flege and Hillenbrand, 1986) and in Norwegian (Fintoft, 1961).

In the production results the average vowel/consonant ratio across both voicing conditions showed that the L2 speakers used the ratio but to a much smaller degree than the L1 speaker. Both L1 and L2 used some phonetic voicing to separate the /s/-/z/ contrast. The percentage of the length of the consonant differed, but the absolute length of the voicing itself was almost identical for both groups. The ratio was more similar in the /t/-/d/ condition, with the ratio of L2 speakers being slightly lower than for L1 speakers. In the case of the /s/-/z/ contrast, the difference between the two groups was bigger. The L1 speakers had the same ratio as in the /t/-/d/ contrast, but the ratio fell significantly for the L2 speakers. He also briefly addressed whether there were any differences between the L2 speakers with respect to native input measured in time spent in an English-speaking country, but he found no effect of native input in his results.

The results from the perception experiment showed that L1 and L2 speakers appear to be relying on the same acoustic features to separate the two pairs of phonemes. Morland (2010) states that the results showed that the Norwegian L2 speakers of English were not very good at producing the difference between /s/-/z/, even though they relied on the same as L1 speakers when separating the two in perception.

2.7 My hypotheses

My hypotheses are primarily based on SLM and suggestions from PAM-L2. Factors such as, the amount of native input a L2 user receives, and whether this input is from a L1 speaker and how much the L2 is used compared to L1 contribute to how well nonnative phonemes are acquired (Best and Tyler, 2007, Flege, 1995). A key factor that might influence the likelihood of detecting significant differences in the L2 phonology is whether the nonnative sound is crucial in discriminating between high frequency words. High frequent minimal pairs would apply pressure to perceptually learn the distinction (Best and Tyler, 2007, p. 30). Both PAM and SLM take the amount of relevant input on native and nonnative sounds into consideration when making its hypotheses. It is also suggested that any differences in the orthographic-mapping could potentially contribute to greater phonological awareness (Flege 1995, Best and Tyler, 2007, p. 32). The speech sounds that were used in the experiments can be compared to each other based on these suggested factors for SLA of phonology. If hypercorrection was to occur in production, this would be expected to happen only if the results of the perception experiment shows that the nonnative phoneme has been acquired by the L2 speaker (Eckman, Iverson and Song, 2013). I also predict that participants who do use hypercorrection or has a low number of native transfer errors will have an accuracy above chance in the perception task.

The unvoiced and voiced pair of alveolar stops /t/ -/d/ were included since this contrast is present in both sound systems. The Norwegian speaking participants (L2) were expected to perform close to the same level as the Native English-speaking participants (L1). This was also the hypothesis for the production experiment. Morland (2010) also found that L2 perception and production of these phonemes resembled the perception and production of L1.

For the unvoiced and voiced alveolar fricatives, I hypothesize that the L2 speakers are likely to have formed a L2 phonological category for the voiced alveolar fricative /z/, because this contrast appears in several high-frequent minimal pairs, three out of four words with this contrast in my list of minimal pairs for the perception experiment were among the 5000 most frequent words (COCA, 2017). For the unvoiced alveolar fricative /s/ the same phonetic category present in the Norwegian sound system should be applied. The alveolar fricatives can also sometimes be separated based on orthography, e.g. <bus> vs. <buzz>. The hypothesis is that the speakers will perform well above chance in the perception experiment, and they are predicted to show some mastery of the contrast in their production. This leads to the prediction that the L2 participants will perform at a level close to the alveolar stops in both the perception and production experiment.

In the case of the pair of unvoiced and voiced post-alveolar fricatives /ʃ/-/ʒ/, the frequency of words that are minimal pairs is low, for example only 1 out of 4 words in the minimal pairs that I found were among the 5000 most frequent words (COCA, 2017). Another factor that might contribute to this contrast being harder to acquire is the fact that there is no way of telling the difference based on the orthography, for example <mission> (/ʃ/) and <vision> (/ʒ/). Following, this contrast should be more difficult to acquire in comparison to the alveolar fricatives /s/-/z/.

For the L2 participants, if both phonemes are perceptually assimilated to the unvoiced postalveolar fricative (which is present in the Norwegian sound system), then the percentage of correct responses will be equal to chance. If any participant has cases of hypercorrection in their production, their results in the perception experiment should be above chance. Following, if there is no evidence of the phoneme being acquired in the perception experiment, then the production of the voiced postalveolar fricative should be poor or wrong for the L2 participants.

The pair of unvoiced and voiced affricates /tʃ/-/dʒ/ are predicted to be treated by the L2 participants as new phonemes. The L2 participants are assumed to have established two L2 phonetic categories associated to those sounds. The frequency of minimal pairs which have this contrast is low, for example only 1 out of 4 words in the minimal pairs that I found were among the 5000 most frequent words (COCA, 2017). However, the sounds are represented differently in the orthography, but since both phonemes are nonnative the low volume of contrasting input should not hinder the acquisition of these phonemes. The acquisition of these two nonnative phonemes should be easier than the acquisition of the postalveolar /ʃ/-/ʒ/ contrast. A perceptual assimilation effect is also not expected since the hypothesis is that both phonemes will be treated as new. This leads to the prediction that the L2 participants will perform at a level close to the alveolar stops in both the perception and production experiment. There is no expectation of any hypercorrection in the production.

The L1 speakers are predicted to have results reaching ceiling-effect in both the perception test and the production task.

3.0 Method

All the data in this thesis were collected by using a questionnaire, a perception test, and a production task. I used two different questionnaires for the two groups of participants. The questionnaire given to the control group contained questions asking about their language background and any potential hearing or speech difficulties. The second questionnaire made for the Norwegian L1 speakers contained the same questions, but it also asked about school background, language exposure, attitude towards learning and speaking English, evaluation of their own level, and their final grade in the English subject in high school. All potential participants were asked to fill out the questionnaire first to be able to filter out any candidates that did not fulfill the criteria of the study. The questionnaires are included in appendices 4 and 5.

The questionnaire distributed to the Norwegian speaking participants contained close-end questions about their motivation and attitude. The close-end questions asked for an answer on a scale from 1-6, with 1 being the most negative score and 6 being the most positive score. The questionnaire also contained questions about time spent listening to and participation in English conversation give estimates of the participants' average week, and these were open-ended. The final question was an open-ended question about other possible factors that could have affected their English.

The participants were informed that the experiment was voluntary and that they could withdraw at any time. The study was registered with the NSD - Norwegian Centre for Research Data.

3.1 Participants

Two different groups of Norwegian speakers were recruited, where the participants in one group had all been on exchange while the other had not. The study included 16 Norwegians aging between 22-25(13 female, 3 male). A control group consisting of 6 native-speakers of English aging between 20-38 (5 female, 1 male) was also recruited. All the participants were recruited by convenience and snowball sampling. The Norwegian L1 participants were all students at NTNU and the English L1 control group was also recruited at NTNU. The informants received a gift card worth 125 NOK at Trondheim Kino as a token of gratitude for their participation.

The Norwegian speaking participants[2] all aged between 22-25 years (mean=23.7). The age range was not predetermined, but the goal in the sampling of the participants was to keep the age-range small. There was a total of 16 Norwegian speaking participants, and out of this group, 8 had been on at least one exchange (min. 4 months) to an English-speaking country. None of the participants had studied English at university level. I chose to sample the group from a university student population to be able to control for the age factor (partially) and higher education, and thus be able to examine the impact of variables such as attitude and language exposure when looking at their performance. These are variables that have been suggested to carry more weight than age of acquisition when looking at SLA (Marinova-Todd, Marshall and Snow, 2000).

The control group consisted of 6 native English speakers (5 females, 1 male). The participants were between 20-38 years (mean = 29). One of the participants in the control group also spoke Norwegian at a fluent level but indicated in the questionnaire that English was the stronger language. The rest of the control group did not indicate having more than English as their native language, and only basic knowledge of any other languages. Two of the participants in the control group spoke American English and the rest of the group spoke British English.

None of the participants in either group had any hearing or speech impairments.

3.2 The experiments

The experiments were created to target the unvoiced-voiced pairs of alveolar stops /d/-/t/, alveolar fricatives /s/-/z/, postalveolar fricatives /ʒ/-/ʃ/ and affricates /dʒ/-/tʃ/. The Norwegian L1 informants were not expected to be familiar with all the words used in either the perception or the production task. This was not considered to be a problem, because the hypotheses of the experiments were assuming that the frequency of minimal pairs might play a role in the perception and production of the different words. The task did not require the listener to be familiar with the words, but rather it focused on the perception of the words, and the sounds in them. The control group completed the same experiment as the Norwegian L1 speakers. The perception experiment was always completed before the production experiment.

The software OpenSesame (Mathôt, Schreij and Theeuwes, 2012) was used to create both the perception and the production experiment. The recordings happened in the phonetics lab at

[2] One participant also spoke Tamil as a first language, but the participant was included in the study because no sources was found that Tamil has voiced fricatives or affricates. The participant was also asked about this after completing the experiment and the participant did think that Tamil contains voiced fricatives.

Dragvoll, using a Shure condenser microphone KSM-44 in a soundproof recording room. The signal was amplified and digitally converted using a Focusrite PRO-40 analog-to-digital converter (DAC). The software Adobe audition 2.0 was used for the recording. The signal was sampled at 44,1 kHz at a 16-bit amplitude resolution and was high-pass filtered at 60 Hz.

3.2.1 Perception

The perception experiment consisted of minimal pairs where the contrasting sounds were the unvoiced-voiced pairs of alveolar stops /d/-/t/, alveolar fricatives /s/-/z/, postalveolar fricatives /ʒ/-/ʃ/ and affricates /dʒ/-/tʃ/. An example of such a pair is <batch> and <badge>. Some near-minimal pairs were also included, but these were only used during a training-phase of the test. The wordlist consisted of a total of 11 minimal pairs, and out of these, two pairs were near-minimal pairs. The wordlist is included in appendix 2. Mainly words with a vowel preceding the target sound (VCV or VC) was chosen, such as <confusion> /kənˈfjuːʒən/. Some words containing nasal sound preceding the target sound (nasalC) were included in the perception test, such as <lunch> /lʌn(t)ʃ/. The chosen words were mostly high-frequency words, but because cases of minimal pairs with the postalveolar fricatives and affricates are rare, the set also include some low-frequency words. 15 (out of 22) of the words were among the 5000 most frequent words in English (COCA, 2017).

The speech material for the perception test was produced by 6 native speakers of English (3 female, 3 male). Their age range 20-29 and the average age was 23.2 years. None of them had any speech or hearing impairments. Two of the speakers spoke British English, while the others spoke American English.

The words were produced in two different carrier phrases; “Say the word __ more than once” and “Say the word___ again”. The words were produced in carrier phrases to give the speakers some variation. The software Praat (Boersma and Weenink, 2018) was used when working with the sound files. The words were then cut at the positive zero crossing out of the sentences and saved to separate sound files. The separate sound files containing separate words were auditorily assessed to choose the best exemplar of each speaker; e.g. determining whether the final stop was realized.

The perception test was an identification task which had an AXB design. The informants were asked to listen to three words and indicate if the second word was the same as the first (word A) or the third (word B). The stimuli were only presented auditorily and thus the focus was only whether the informants could identify the target word and match it with the correct word.

In the test, the words being compared was produced by different L1 users of English to avoid the listener getting to sensitive to the voice of the speaker and avoid a ceiling effect. Different speakers were used for each word in the compilation of the three words. Word A and word B came out of the minimal pairs and one of the two was played as word X. The stimuli were played in a block-randomized order, and the three voices were the same through each block. The experiment had a balanced comparison across gender and the different varieties of English. The experiment had both a training phase and an experiment phase. The first 10 tokens in the experiment were used as a training phase and they were not included in the data analysis. The experiment phase consisted of 5 blocks, with each block containing a randomized presentation of 36 compilations of the minimal pairs. Each word was presented twice as word X, and each word was presented both as word A and word B for each block. This gave a total of 180 tokens in the experiment-phase.

The participants were informed about the training phase and instructed to use it to become familiar with the task and adjust the volume of the sound. All the participants did the perception experiment in the phonetics lab, using the same laptop and a Jabara MOVE headset. The experiment took approximately 15 minutes to complete.

3.2.2 Production

The production experiment consisted of minimal pairs selected in the same fashion as the perception experiment. The list of words was a revised version of the perception list and consisted of a total of 9 minimal pairs, and out of these, two pairs were near-minimal pairs. Only words with a vowel preceding the target sound (VCV or VC) was included. The words where the target sound was preceded by a nasal sound was excluded because they cannot be analyzed in the same manner the data in the study by Morland (2010). An attempt was made to restrict the selection to high-frequency words, but the set also include some low-frequency words. 12 (out of 18) of the words were among the 5000 most frequent words in English (COCA, 2017). The wordlist is included in appendix 3.

The stimuli were presented in a block-randomized order and each word was repeated a total of 8 times. This resulted in a total of 18*8= 144 tokens for each participant. The participant was asked to read the word that appeared on the screen. The word was presented both orthographically and by audio at the same time and the participant was asked to read the word once they had finished hearing the auditory representation. The experiment progressed as the participant read each word out loud. The orthographical representation stayed on the screen while the participant read the word.

The choice of presenting the stimuli in this manner was made because English is a language where the relationship between the orthography and pronunciation is not always evident. This kind of presentation also helped avoid a listing-effect in the production, without using a carrier-phrase. The speech used was produced by one of the male voices from the perception experiment. His voice was chosen because he had a comfortable voice to listen to and he clearly differentiated between the voiced/unvoiced consonants (assessed by my own listening).

All the participants who passed the criteria for the study completed both experiments and the experiments ran with no problems for all the participants.

3.3 Analysis

The recordings were rated by two independent raters, both were native speaker who had linguistic competence. Their mean age was 31.5(SD= 0.71) and both listed English as their only first-language. Neither had any hearing or speech impairments. They were asked to rate the participants' production of the consonants in question. The rating was on a 5-point scale from 5, *native like*, to 1, which was labeled *wrong*. The 4^{th}, 5^{th} and 6^{th} production of each word was included.

R, a programming language, (R Core Team, 2017) with the use of the RStudio interface (RStudio Team, 2015) was used when conducting the statistical analysis of both the perception results and the results from the results given by the independent raters on the production task. The correlation tests, stepwise regression, ANOVA, t-test and Tukey's HSD test came out of the "stats" and "base" library (R Core Team, 2017). The plots were made using the "ggplot2" library (Wickham, 2009).

4.0 Results

The answers from the questionnaires distributed to the Norwegian-speaking participants were entered in a spreadsheet. The questionnaire asked the L2 participants several questions about their attitude, competence and exposure to English. The questionnaire asked the participants to give their final grade in English in high school. The average final grade in the English subject in High school in Norway was 4.3 for both previous school years (2015-2016/2016-2017) (Utdanningsdirektoratet). The mean of all the L2 speakers was 4.75 (SD= 0.77). The grading scale in Norwegian high schools goes from 1 (fail) to 6 (high achievement). Open-ended questions were estimations of weekly exposure, and the close-end questions went on a scale from 1 (low/negative) to 6 (high/positive).

Table 1: The different answers of each L2 participant in the questionnaire

Participant number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Average correct response in percent (perception task)	81.1	78.9	82.8	76.7	87.2	82.8	81.7	80.6	85.6	88.3	86.7	80.0	84.4	68.3	72.2	62.0
No. of months living in English-speaking country	0	0	0	0	0	0	0	0	9	10	10	5	10	4	6	6
Been on exchange to English-speaking country	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Age	25	23	25	22	24	23	23	24	24	23	25	23	24	25	22	24
Hours of exposure to oral English per week	30	25	37.5	3	1.5	6	12	20	21	20	35	14	30	3	8	14
Hours of oral conversation in English per week	0	0	0.5	1	0.5	7	2	0	1	0	20	8	2	0	0.5	6
Final grade in English in High School	6	5	5	6	5	5	5	4	4	5	5	5	5	3	4	4
Self-evaluation of oral competence in English	6	3	3	3	4	4	4	4	4	5	5	4	4	3	3	4
Attitude towards English subject in high school	6	5	5	5	5	4	4	1	4	5	5	3	4	2	3	4
Attitude towards speaking English	6	3	4	3	3	5	5	2	4	5	6	4	5	4	4	5
Number of L1s	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
Other comments												[3]	[4]		[5]	[6]

[3] Attended an English immersion program (International Baccalaureate) 2nd and 3rd year of high school.
[4] Lived in England between the age of 0-3 years old.
[5] Was not feeling well on the day of the test.
[6] Speaks Tamil with family members.

4.1 Perception

4.1.1 Percentage of correct responses on perception

The data from the perception test resulted in binary values indicating whether the participant responded correctly. The average percentage of correct responses was then computed for each participant. A list of means is presented in table 2.

Table 2: Percentage of the correct responses on the perception experiment for the participants divided into the groups L2A (never lived in English speaking country), L2B (have spent at least 4 months in English-speaking country), and L1 (native English speakers)

Participant number	L2A	Participant number	L2B	Participant number	L1
1	81.1	9	85.6	17	97.8
2	78.9	10	88.3	18	97.8
3	82.8	11	86.7	19	93.3
4	76.7	12	80.0	20	99.4
5	87.2	13	84.4	21	98.9
6	82.8	14	68.3	22	99.4
7	81.7	15	72.2		Mean= 98%, SD= 2%
8	80.6	16	62.2		
Group	Mean= 81%, SD= 3%		Mean= 78%, SD= 10%		

According to a two-sided t-test of the means scores of L2A and L2B showed that the difference between them was not significant ($t = -0.830$, $df = 8.415$, $p = 0.430$). Both L2 groups showed a true difference in means when compared to the L1 control group when using a two-sided t-test; L2A ($t = 11.3$, $df = 11.997$, $p < .001$) and L2B ($t = 5.4294$, $df = 8.029$, $p < 0.001$).

Table 2 shows that three of the participants in L2B, had a lower rate of correct responses, compared to all the L2 speakers. Combining the two groups of Norwegians gave a mean of 80% and SD =7%. Comparing the means of subjects 14, 15, and 16 to the mean of L2 data combined showed that their score fell outside of 1SD from the mean. No other participants scores were this far from the mean. The lower SD of the L1 control group (98.0%, SD=2.0%) shows that the means of the L1 participants had a smaller variability than the means of the L2 groups (80%, SD= 7.0%).

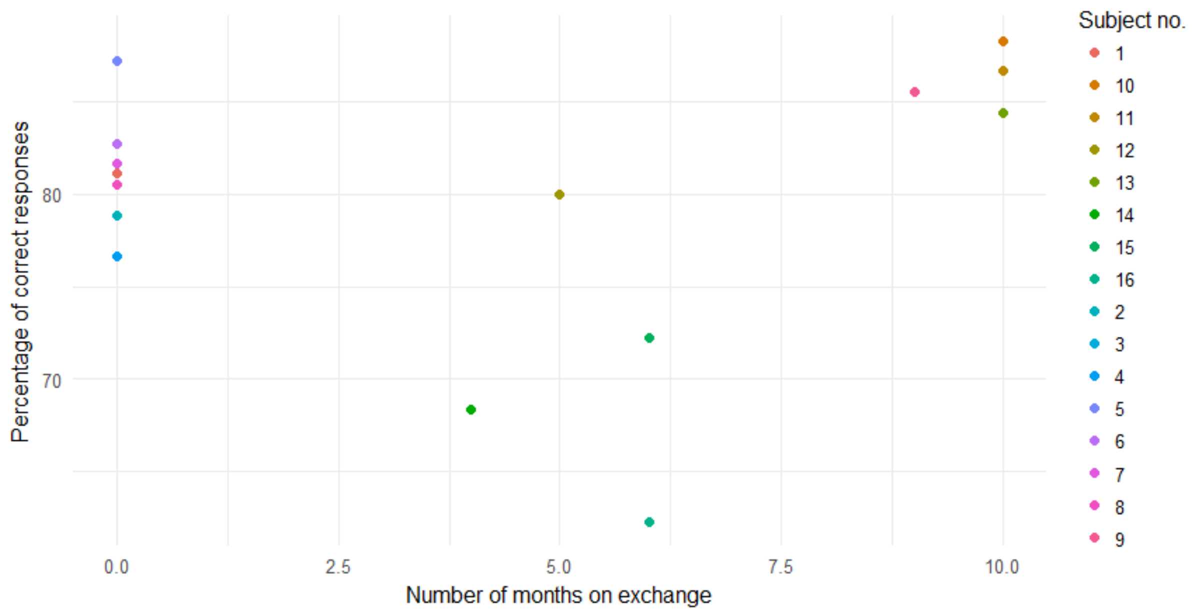

Figure 1: The mean percentage of correct responses for the L2 subjects plotted against the numbers of months spent in a English-speaking country

Figure 1 shows the plot of the mean percentage of correct response of all the participants who had English as L2(y-axis), and the length of stay in an English-speaking country(x-axis). The correlation between *percentage of correct responses* and *number of months on exchange* was positive, weak and not statistically significant, r = 0.103, df = 14, $P_{\text{one tailed}} = 0.352$. This shows that the number of months spent in an English-speaking country was a poor estimator of the results on the perception test.

Because L2A and L2B showed no true difference in means, I chose to pool them together for further data analysis. A two-sided t-test comparing means of the L1 and the pooled L2 gave a significant p-value for the difference in mean scores (t = -8.8432, df= 19.892, p< 0.001).

4.1.2 Effect of different explanatory variables on perception

All the means and correlation tests below are taken from table 1.

A two-sided Spearman's rank correlation coefficient test on the variables *percentage of correct responses* and *participants final grade in English subject in High school* was positive, weak but not statistically significant (rho = 0.314, S = 466.16, p-value = 0.236). The same test on the correlation of the *percentage of correct responses* and the participants' *self evaluation of oral competence* was positive, weak and statistically significant (rho= 0.572, S = 290.94, p-value = 0.0207).

The participants were asked about their attitude towards English. Specifically, they were asked about their attitude towards the English subject in high school and their attitude towards speaking English. The correlation of between *percentage of correct responses* and *attitude towards the english subject in high school* was positive and weak, but not statistically significant (rho = 0.428, S = 389.06, p-value = 0.0983). The correlation of the *percentage of correct responses* and *attitude towards speaking English* was positive and very weak, but not statistically significant (rho = 0.267, S = 498.77, p-value = 0.318).

The Pearson Product-Moment Correlation of the *percentage of correct responses* and *exposure to English speech* was positive, weak and not statistically significant (r = 0.377, df= 14, t = 1.525, $P_{\text{two tailed}}$ = 0.1495). The correlation of the *percentage of correct responses* and *participation in conversations in English* was positive and very weak, but not statistically significant (r = 0.113, df= 14, t = 0.426, $P_{\text{two tailed}}$ = 0.677).

The sample is small with a n= 16, so a multiple linear regression will not be able to say anything about a larger population; but it will show which of the explanatory variables show the most effect on the responses. The available variables included in the initial multiple linear regression were, *self-evaluation of oral competence, final grade in high school, number of months living in an English-speaking country, hours of exposure, hours of oral conversation, attitude towards the english subject in high school* and *attitude towards speaking english.*

None of the variables came back with a significant p-value, except the intercept. A step-wise regression, performed in both directions only included the variable *final grade in english subject in high school* (see table 3). The multiple linear model was marginally significant.

Table 3: Print-out of multiple linear regression model after performing stepwise regression on the results from the perception experiment

Variable	**Estimate**	**Std. Error**	**t-value**	**Pr(|t|)**
(intercept)	58.63	0.10316	5.684	<0.001
Final grade in HS	4.49	0.02145	2.093	0.055
Multiple R-squared: 0.328 Adjusted R-squared: 0.184 *F-statistic: 4.382 on 1 and 14 DF , p = 0.055*				

4.1.3 Effect of sound pair on perception

The data were then grouped with respect to sound pair (*sound*) and the voiced and unvoiced phonemes were pooled together. The L2 participants had the lowest standard deviation in the alveolar stops (7%), and the mean was also the highest for this sound (88.1%). The L2 postalveolar fricatives had the lowest mean score and the mean was 60.0% (SD =10%), which is just above chance.

The L1 control group differed little in mean score and standard deviation was low across all the different sounds. The highest standard deviation in the L1 group was 5% for the affricates, which had a mean score of 97.2%. The lowest mean score of L1 was found in the alveolar stops, which had a mean score of 97.2% and standard deviation of 3.0%.

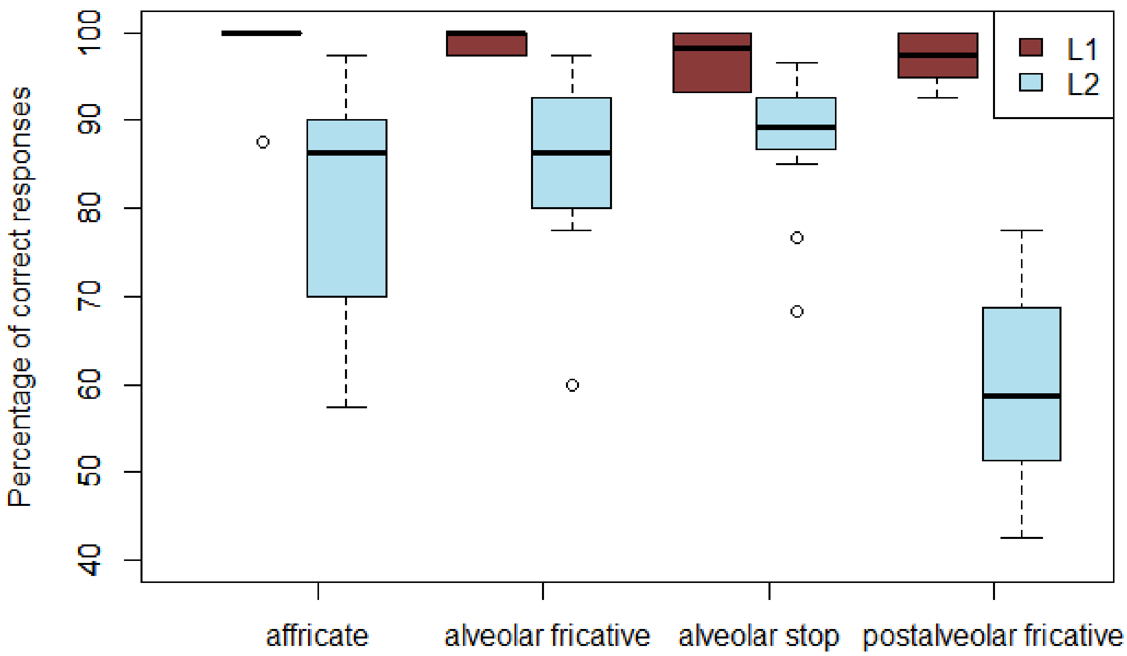

Figure 2: boxplot of the percentage of correct responses on the perception experiment by the L1 and the L2 group seperated by sound pair

To test the differences illustrated in figure 2, an ANOVA and a Tukey's HSD test was conducted on *L1/L2* and *sound*. The ANOVA gave p-values < 0.001 on the means of the L1 and L2 participants compared (F (1, 80) =83.17), for the variable *sound* (F (3, 80) = 26.01), and the interaction between these variables (F (3, 80) = 8.87). A Tukey's HSD test was then performed to see which of the different means in the boxplot was significantly different from each other.

The Tukey HSD test showed that the responses in the L1 group did not differ significantly from each other with respect to target sound.

Looking at the L2 group, the alveolar fricatives, affricates and alveolar stop-values did not differ significantly from each other. The postalveolar fricatives had a significantly lower percentage of correct responses compared to the other sound pairs, with p-values < 0.001. Table 4 lists all differences between means and p-values reported in the Tukey HSD test.

Table 4: difference and p-values of a Tukey HSD test across sound pair for the L2 group

	affricate	*alveolar fricative*	*alveolar stop*
alveolar fricative	diff= 3.4% p= 0.949		
alveolar stop	diff= 6.1%, p= 0.495	diff= 2.7%, p= 0.988	
postalveolar fricative	diff= -21.9%, p < 0.001	diff= -25.3%, p < 0.001	diff= -28.0%, p < 0.001

When comparing L1 and L2 with respect to sound, the Tukey HSD test reported a significant difference ($p = 0.029$, diff = -13.7%), for the means of the alveolar fricatives. The means of the affricates were also significantly different from each other ($p = 0.0059$, diff = -15.9%). It is observable from the boxplot that the postalveolar fricatives were different, and the Tuckey HSD test gave a $p < 0.001$, diff= -36.9% when comparing L1 and L2 perception of the postalveolar fricatives. The mean of correct responses for L1 and L2 was not significantly different for the alveolar stops, which reported a $p= 0.367$, diff = -9.1%.

4.1.4 Effect of voicing feature on perception

A boxplot of the results shows the results split by whether participants correctly identified the unvoiced and voiced sounds, respectively. It shows that the percentage of correct responses to word X overlap within the L2 group when the result is split by whether participants correctly identified the voiced and the unvoiced sounds, respectively.

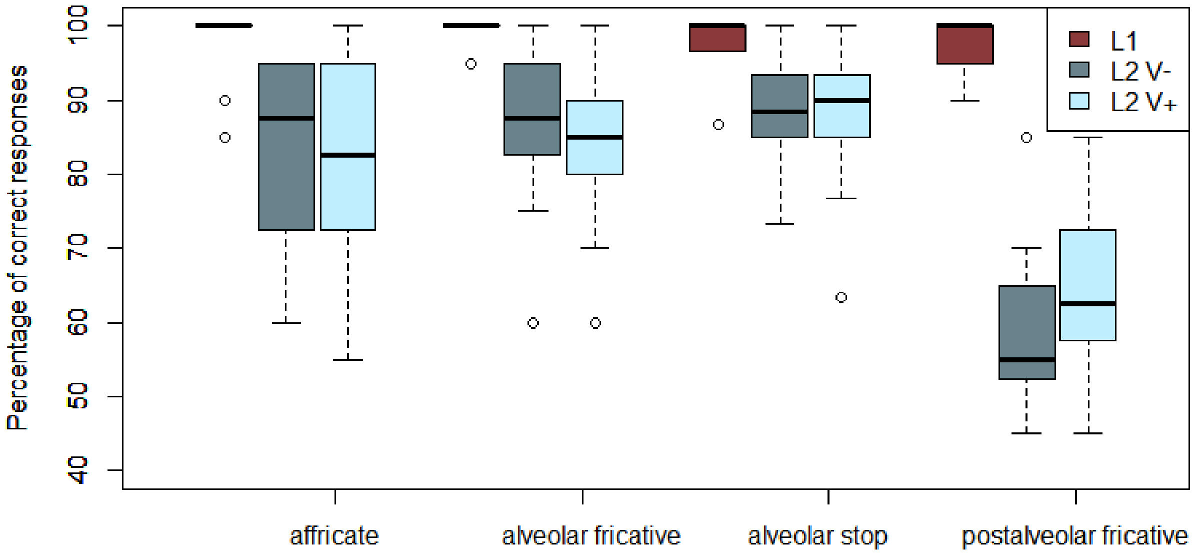

Figure 3: boxplot of the percentage of correct responses on the perception experiment by the L2 group seperated by sound and voicing feature [V+(voiced), V-(unvoiced)] of the target phoneme in word X (L1 plotted for reference)

The results from both the ANOVA on variables sound, voicing, L1/L2 showed that there was no difference in how well participants identified the voiced vs. the unvoiced counterparts of each phoneme.

The effect of voicing in the ANOVA was not significant, $F(1,163) = 0.070$, $p = 0.791$. The effect of L1/L2 was significant, $F(1,163) = 125.990$. $p < 0.001$. The effect of the interaction between L1/L2 and sound was significant, $F(6,163) = 26.420$. $p < 0.001$. The effect of the interaction between the voicing feature and sound was not significant, $F(3,163) = 0.009$, $p < 0.448$. The effect of the interaction between the voicing feature and L1/L2 was not significant, $F(1,163) = 0.048$, $p < 0.827$.

4.2 Production

For the production data, ratings from each of the two native English-speaking informants were z-score transformed prior to analysis to eliminate any potential bias that might arise with rating tasks (Schütze and Sprouse, 2014). The means of the z-scores given by both raters were used in the analysis of the data. The z scores ranged from min = -2.34 to max =1.77 (range = 4.11).

4.2.1 The average z-scored rating on production

Table 5: The mean z scored ratings of the production data for each participant divided in groups L2A (never lived in English speaking country), L2B (have spent at least 4 months in English-speaking country), and L1 (native English speakers)

Participant number	L2A	Participant number	L2B	Participant number	L1
1	-0.144	9	-0.561	17	1.051
2	-0.579	10	0.541	18	0.930
3	-0.727	11	0.625	19	0.680
4	-0.681	12	-0.181	20	0.458
5	-0.987	13	-0.292	21	1.458
6	-0.542	14	-0.061	22	1.134
7	0.254	15	-0.598		Mean= 0.952, SD= 0.35
8	-0.357	16	-1.422		
Group	Mean= -0.470, SD= 0.39		Mean= -0.243, SD= 0.66		

A two-sided t-test comparing L1 and the pooled L2 means gave a significant p-value ($t = 6.682$, $df = 13.913$, $p < 0.001$) showing that the ratings received by L2 participants were significantly lower than the L1 participants.

The average z scored rating of the pooled L2 was -0.357 (SD= 0.53). Comparing the mean of subject 16 to the mean of L2 data showed that this participant was scored more than 1SD away from the mean (range = -0.89 - 0.17). No other participant fell outside the 1SD range.

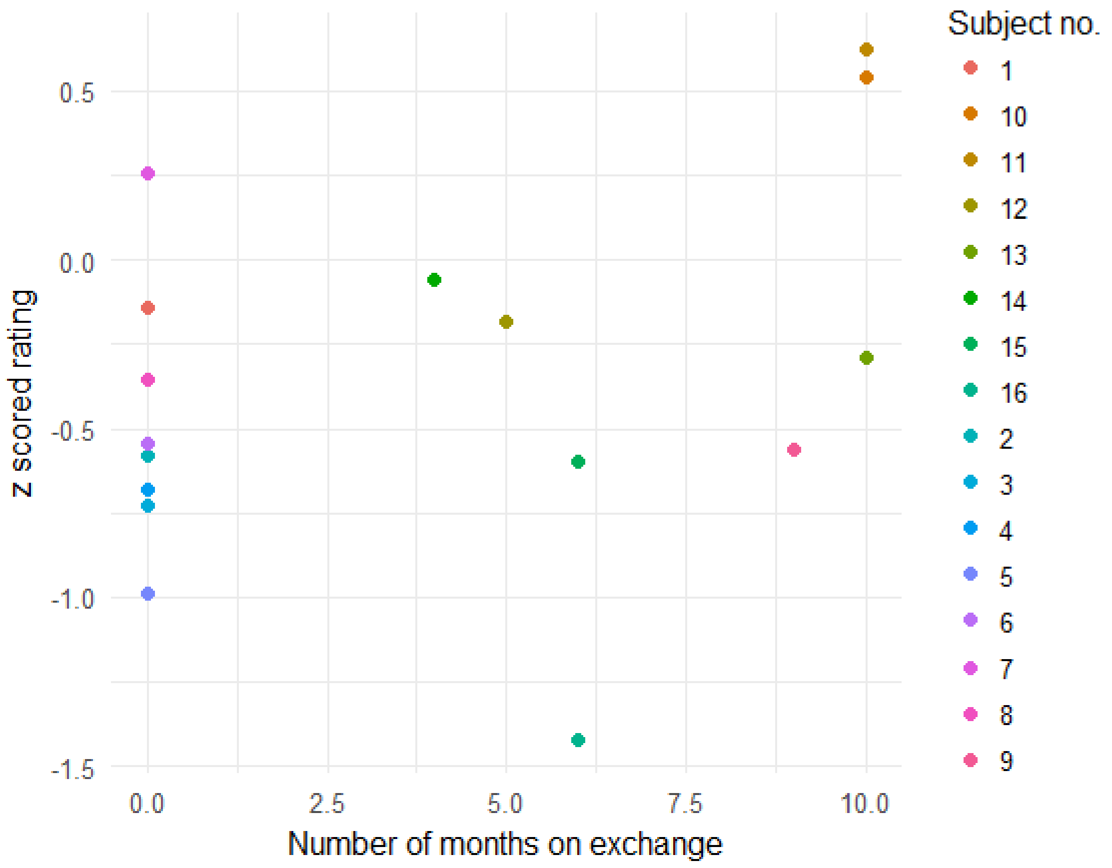

Figure 4: The mean z-scored rating of production of the target consonants by the L2 subjects plotted against the numbers of months spent in a English-speaking country

A two-sided t-test of L2A and L2B showed that the difference between scores was not significant ($t = -0.842$, $df = 11.3$, $p = 0.417$). A two-sided t-test comparing both L2A and L2B to L1 showed significant differences in judgement scores when compared to the L1 control group; L2A ($t = -7.191$, $df = 11.437$, $p < .001$) and L2B ($t = -4.377$, $df = 11.096$, $p = 0.001$). The mean score of the participants from the L2B group was slightly higher (-0.243) than the score of the L2A group (-0.470).

4.2.2 Effect of different explanatory variables on production

All the data on the explanatory variables have been taken from table 1.

A Spearman's rank correlation test showed that the *ratings of the recordings* did not significantly correlate with the *final grade* (rho = 0.105, S= 608.77, $p_{two\text{-}sided}$=0.699), and there was also no significant correlation with the *attitude to the English subject in high school* (rho= -0.055, S=717.58, $p_{two\text{-}sided}$=0.839). A Spearman's rank correlation test showed that the *ratings of the recordings* had a medium, positive and significant correlation with the participants *self-evaluation of oral competence* (rho= 0.526, S=315.69, $p_{two\text{-}sided}$=0.032). The correlation with *attitude towards speaking English* was also medium, positive and significant (rho= 0.524, S=324.02, $p_{two\text{-}sided}$=0.037).

A Pearson's product-moment correlation test showed that the *rating of the recordings* did not significantly correlate with the *number of months the participant had been on exchange* (r = 0.350, t=1.396, $p_{two\text{-}sided}$= 0.185). The same test also did not find any significant correlation with either *hours of weekly exposure* (r=0.296, t= 1.161, df=14, $p_{two\text{-}sided}$= 0.265), or *hours of weekly conversation in English* (r=0.323, t=1.277, df=14, $p_{two\text{-}sided}$= 0.222).

A linear regression model created, in the same fashion as the linear model of the perception data. As before, the sample is small with a n= 16. The available variables included in the initial multiple linear regression were, *self-evaluation of oral competence, final grade in high school, number of months living in an English-speaking country, hours of exposure, hours of oral conversation, attitude towards the english subject in high school* and *attitude towards speaking english* (can be found in table 1).

None of the variables came back with a significant p-value, not even the intercept. Also, the model did not receive a significant p-value (0.6624).

A step-wise regression, performed in both directions, kept the variable *self-evaluation of oral competence in English.* The F-stats and the coefficients of the revised linear model are reported in table 6. The p-value of the rewised multiple regression model also did not receive a signinficant p-value and the estimate of *self-evaluation of oral competence in English* was not significant (table 6).

Table 6: Print-out of final multiple linear regression of production after performing stepwise regression

Variable	**Estimate**	**Std. Error**	**t-value**	**Pr(|t|)**
(intercept)	-1.2268	0.5012	-2.448	0.0282 *
Self-evaluation of oral competence	0.2047	0.1142	1.792	0.095
Multiple R-squared: 0.187 Adjusted R-squared: 0.128 *F-statistic: 3.21 on 1 and 14 DF , p = 0.095*				

4.2.3 Effect of sound pair on perception

The ratings of the participants' production were split by whether the speaker belonged to the L1 or L2 group, and sound. A boxplot of the results shows how the z-scored ratings distributed across L1 and L2 participants and the different sound pairs that were investigated. It shows that the z-scored ratings of L2 speakers and L1 speakers overlap across all sound pairs.

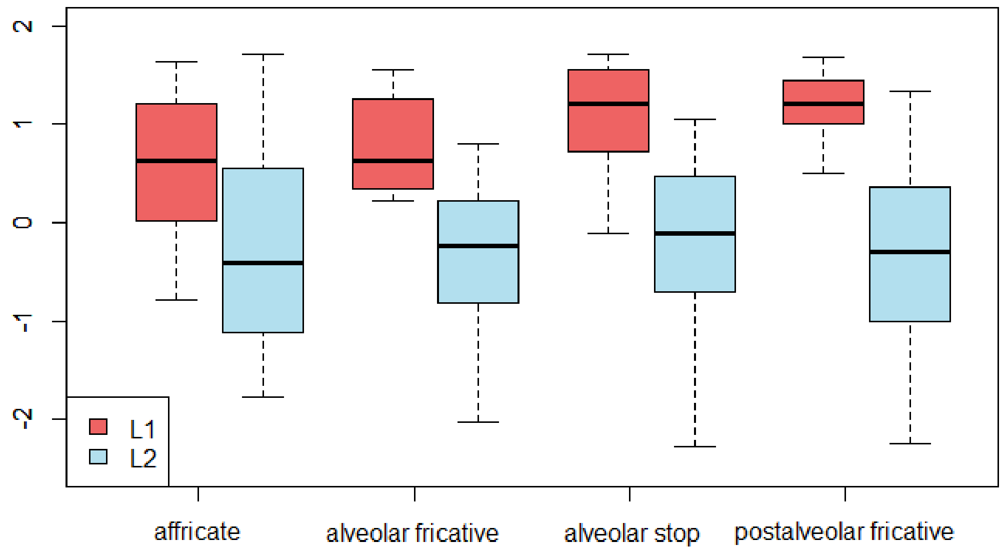

Figure 5: Boxplot of the z scored ratings of the production data by L1 and L2 groups across the different sound pairs

An ANOVA and a Tukey's HSD test were conducted to test the differences illustrated in the boxplot.

The ANOVA was performed on the variables, *L1/L2* and *sound.* The only significant p-value was on the variable L1/L2 ($F (1, 168) = 85.149$, $p < 0.001$). The variable *sound* was not significant ($F (3, 168) = 0.759$, $p = 0.519$), and the interaction between variables *L1/L2* and *sound* was not significant ($F (3, 168) = 1.146$, $p = 0.332$).

A Tukey's HSD test was then performed to see which of the groups in the boxplot were significantly different from each other. When comparing L2 and L1 with respect to place and manner of articulation, the test reported all p-values were significant. All the means of L1 and L2 were significantly different from each other when looking at the target sound pairs. The differences and p-values are listed in table 7.

Table 7: P-values and differences comparing L2 and L1, across the different sound pairs

L1 | L2	affricate	alveolar fricative	alveolar stop	postalveolar fricative
difference and p	d =-1.328 $p < 0.001$	d = -1.214 $p < 0.001$	d = -0.873 $p = 0.031$	d = -1.569 $p < 0.001$

When comparing L2 across place and manner of articulation, the Tukey's HSD test reported no significant p-values, and they were all quite high (none lower than 0.969). When comparing L1 across place and manner of articulation, the Tukey's HSD test reported no significant p-values, and they were all quite high (none lower than 0.584).

In summary, the ANOVA and the Tukey's HSD showed that the L1 received a significantly higher score than the L2 participants, but the tests located no significant differences with respect to the sound pairs for either the L1 or the L2 group individually.

4.2.4 Effect of voicing feature on production

The ratings of the participants' production were split by whether the speaker belonged to the L1 or L2 group, sound and the voicing feature of the intended sound. A boxplot of the results shows how the z-scored ratings distributed across L1 and L2 participants, voicing feature and the different sound pairs that were investigated. It shows that overall, the z-scored ratings of L2 participants were higher in the unvoiced sounds.

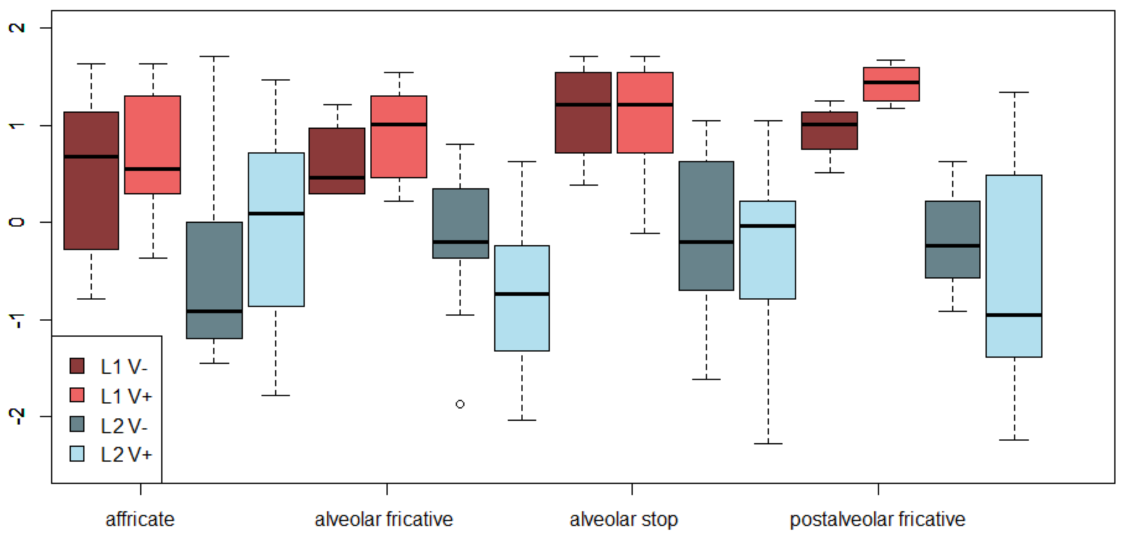

Figure 6: Boxplot of the z scored ratings of the production data by L1 and L2 groups across voicing[V+,V-]

In order to test the differences illustrated in the boxplot, ANOVA and Tukey's HSD test was conducted.

The ANOVA, looking at the interaction between L1/L2, voicing feature, and sound pair, was significant (F (15, 160) = 6.966, p < 0.001).

A Tukey's HSD test was then performed to see which of the groups in the boxplot were significantly different from each other, and the results are presented in table 8.

When comparing L1 and L2 individually across the voicing feature and sound pair, the Tukey's HSD test reported no significant p-values of the within group comparisons, and they were all quite high (none lower than 0.732).

When looking at the comparisons across the L2 and L1 groups, the Tukey HSD test showed that when comparing the two groups with respect to the unvoiced affricate, the unvoiced alveolar fricative, the unvoiced alveolar stop, and the unvoiced postalveolar, there were no true differences (the lowest p = 0.076). Both the voiced and the unvoiced alveolar stop showed no significant difference in means (see table 8).

Comparably, when looking at the voiced sounds, the affricate, alveolar fricative and postalveolar fricative received significant p-values. The differences and p-values of all the comparisons are listed in table 8.

Table 8: Difference and p-values comparing L1 and L2 with respect to voicing feature and sound pair

V+|V-	**affricate**	**alveolar fricative**	**alveolar stop**	**postalveolar fricative**
V- : L2|L1	Diff = - 1.260 P = 0.076	Diff = -0.762 P = 0.812	Diff = -1.021 P = 0.339	Diff = -1.122 P = 0.194
V+ : L2|L1	Diff = -1.396 P = 0.026	Diff = -1.667 P = 0.002	Diff = -0.726 P = 0.863	Diff = -2.015 P < 0.001

4.3 Relationship between perception and production

The data from both the perception experiment and the production data were merged into one dataset where the available variables were *Participant number, L1 /L2, voicing feature, sound, z scored rating, ratio of correct responses*. This resulted in a dataset of 176 observations (sound [alveolar stop, alveolar fricative, affricate, postalveolar frivative]* voicing feature [voiced, unvoiced]* 22 participants = 176 observations). Figure 7 illustrates the relationship between the two variables *z scored rating*(=production) and *ratio of correct responses*(=perception).

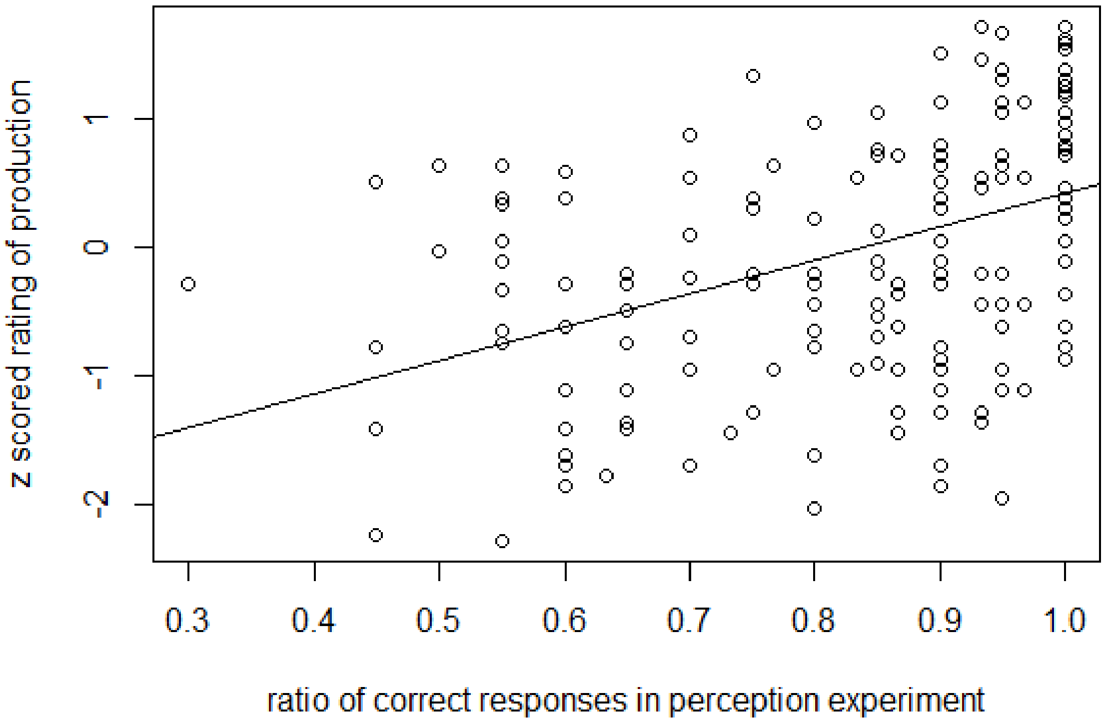

Figure 7: scaterplot of z scored ratings of the recordings and the ratio of correct responses in the perception test, with the estimated linear model

A Pearson product-moment correlation coefficient was computed to assess the relationship between the two variables *production* and *perception*. It showed a significant, and positive correlation between the two variables *rating* and *response*, r = 0.4305, df = 174, t= 6.291, p < 0.001, n= 176.

4.3.1 Instances of both raters giving a score of 1 on the production

The results from the production test were examined for instances where both raters scored the production of a sound wrong. These results were visualized in a table and analyzed for possible correlations. It is observable from table 9 that any L2 participant that had their production of an unvoiced fricative labeled wrong had an accuracy above chance in the perception experiment. The full table including all observations and both L1 and L2 participants can be found in appendix 6).

Table 9: A list of L2 participants, giving the percentage of instances rated 1 (wrong), indicated with the intended voicing (v-, v+), and the results of the percentage of correct responses in the perception experiment

Participant no.	sound	wrong v- %	wrong v+ %	perception %
13	postalveolar fricative	33.33	0.00	72.5
7	postalveolar fricative	25.00	0.00	57.5
16	postalveolar fricative	25.00	25.00	60
11	postalveolar fricative	16.67	0.00	65
9	postalveolar fricative	16.67	8.33	67.5
10	postalveolar fricative	8.33	0.00	77.5
6	postalveolar fricative	8.33	16.67	62.5
1	postalveolar fricative	0.00	16.67	50
8	postalveolar fricative	0.00	16.67	50
3	postalveolar fricative	0.00	16.67	57.5
2	postalveolar fricative	0.00	16.67	72.5
12	postalveolar fricative	0.00	25.00	52.5
5	postalveolar fricative	0.00	25.00	70
15	postalveolar fricative	0.00	91.67	47.5
16	alveolar fricative	50.00	50.00	60
11	alveolar fricative	16.67	16.67	95
13	alveolar fricative	0.00	16.67	82.5
3	alveolar fricative	0.00	16.67	90
7	alveolar fricative	0.00	16.67	92.5
2	alveolar fricative	0.00	33.33	77.5
4	alveolar fricative	0.00	33.33	77.5
5	alveolar fricative	0.00	50.00	95
15	alveolar fricative	0.00	66.67	82.5
8	alveolar fricative	0.00	66.67	92.5
6	affricate	0.00	33.33	80
5	affricate	0.00	33.33	95
16	affricate	0.00	100.00	57.5
16	alveolar stop	0.00	16.67	68.3
5	alveolar stop	0.00	16.67	88.3
2	alveolar stop	0.00	33.33	90

4.4 Summary of the results

In summary, neither the perception or the production results showed any effects of whether the L2 participant had lived in an English-speaking country. In the perception experiment, the L1 participant scored significantly higher for the affricates, alveolar fricatives and postalveolar fricatives. The L2 participants scored significantly better for the affricates, alveolar fricatives and alveolar stops than the postalveolar fricatives. The mean score on the postalveolar fricatives for the L2 participants was 60.2% (SD = 10%). The performance in the perception test did not differ across the voicing contrast, so the likelihood of responding correctly was not dependent upon whether “word X” contained a voiced or unvoiced target consonant.

In the production data, the L1 and performed better than the L2 group across all manner and places of articulation when the voiced and unvoiced data was grouped. When dividing the data according to the voicing feature, the L1 no longer performed significantly better than the L2 group in the voiced and unvoiced alveolar stop. There was also no longer a significant difference in the scores given to the unvoiced affricates, unvoiced alveolar fricatives and the unvoiced postalveolar fricatives, but difference between the voiced counterparts remained significant.

The correlation between the results in the perception and production data was positive and significant. Any L2 participant that had their production of an unvoiced fricative labeled wrong had an accuracy above chance in the perception experiment. There were more instances of unvoiced postalveolar fricatives which were produced wrong, compared to alveolar fricatives.

Participant 16 in the L2 group spoke two first-languages and had the lowest score on both the perception and production experiment, and this person was also in the L2B group. Including this participant might have affected the results and might be the reason why some of the tests returned non-significant p-values.

5.0 Discussion

5.1 On the issue of native input

This thesis started with the assumption that native input might show an effect in both production and perception of English consonants which were nonnative to the native Norwegian speakers. The amount of native input a L2 user receives, and how much the L2 is used compared to the L1 were two of the factors that SLM (Flege, 1995) and PAM-L2 (Best and Tyler, 2007) predicted would influence the acquisition of nonnative phonemes, and the aim was to investigate this hypothesis in the present study. However, none of the tests were able to detect a significant difference between the L2 participants who had lived in an English-speaking country, and the ones who had not. The tests of both the perception results and the production showed no significant difference between the means of the two L2 groups. These results were in line with earlier findings on effect of native input in Norwegian L2 speakers of English (Morland, 2010). The lack of a significant effect of the time spent in an English-speaking country might be because all Norwegian speakers are exposed to so much native input already, that a stay for (up to) 10 months does not significantly improve their perception of nonnative English consonants. These findings predict that L2 users who are highly competent and have had substantial input over time will not show a significant improvement of their perception non-native phonemes in the L2 if their exposure to native input increases significantly for a limited period of time.

5.2 The effect of different explanatory variables

Since the numbers of months in an English-speaking country did not have a significant influence on neither L2 perception nor production, other explanatory variables were investigated. When attempting to create multiple linear models based on the available explanatory variables, the process of stepwise regression was unsuccessful at creating models with significant results. For the results of the perception task, the multiple linear model kept the explanatory variable *final grade in English subject in high school* and the p-value was marginally significant. The correlation test found that the correlation between the variables *percentage of correct responses* on the perception experiment and *final grade in English subject in High school* was positive and significant. Under normal circumstances this grade is given in the first or second year of high school, and this would have been prior to when the L2B participants would have been living in an English-speaking country, so this grade should not be affected by an increase in English competence caused by a potential stay in an English-speaking country. The variable *final grade in English in high school* had a positive correlation with the variable *attitude towards the English subject in high school*. The correlation between attitude and final grade might stem from two different reasons, either the participants who did well in the English subject remember the it as being enjoyable, or their attitude to learing influenced their motivation for learning English, and thus influenced their learning in a positive direction.

When looking at the results of the production data, the correlation test found that the correlation between the variable *rating on production* and the variable *self-evaluation of oral competence* was positive and significant. This was also true for the correlation between the variable *rating on production* and the variable *attitude towards speaking English*. A multiple linear model of the results from the production data, which underwent stepwise regression, kept the explanatory variable *self-evaluation of oral competence*, but this model was not significant. The participants self-evaluation could potentially be influenced by whether the L2 participant had been living in an English-speaking country, however the correlation test found no significant interaction between the two explanatory variables.

It is possible that the linear models were affected by the low accuracy in the perception and low scores on the recordings of participant no. 16. One explanation for why this participant had a lower accuracy and received lower scores than everyone could be because this person had two first languages. It might be that if a speaker has two first languages, he or she will require more input in order to acquire nonnative phonemes, if the nonnative phonemes in question is not present in either of the two first languages. The inclusion of this outlier might be why the models are not significant, because the only explanatory variable that separated participant no. 16 from the other participants where that person's response on question "Did you grow up speaking any other languages than Norwegian?". Since this response only had one observation in the data, the variable was not included as an explanatory variable in the initial linear models.

5.3 Differences with respect to the sound pairs

The results from the perception experiment showed that the L2 participants were as good as the L1 participants at distinguishing the minimal pairs containing the alveolar stops /t/-/d/. This conforms to the hypothesis of SLM, which states that if there are similar phonemes in the native language, this would lead to there being established phonetic categories in the L2 speakers phonetic inventory because the phonetic categories of the L1 would be adapted and used in the L2 as well (Flege, 1995). The production data of the L2 participants on the alveolar stops were scored as poorer by the independent raters than the production data by L1 participants, and when the results were compared without taking the voicing feature into account, the tests showed a significant difference in the results. However, when the results of the alveolar stops were split into voiced and unvoiced the differences between the production by L1 and L2 participants were no longer significant (table 15). Based on this I argue that the difference which was present was probably due to small articulatory realizations of the two phonemes in the two languages. The raters scores on the consonants could possibly have been influenced by any potential nonnative sounding surrounding vowels in the productions by the L2 participants. By and large, the results support the hypothesis that the voiced and unvoiced alveolar stops are likely to be a part of the phonetic inventory of the L2 participants because the sounds both are present in the phonetic inventory of Norwegian, and their performance should resemble the one of native speakers. The results of both the perception and production task supported this. The results are in line with previous research, conducted by Morland (2010), who found that the perception and production of English of the alveolar stops of Norwegian L2 speakers in English resembled the perception and production of L1 speaker of English.

The unvoiced and voiced affricates /tʃ/-/dʒ/ were predicted to be treated by the L2 participants as phonemes with no link to the phonetic inventory of their L1. The results of the perception experiment showed that the L2 participants' ability to separate of the two affricates did not significantly differ from their ability to separate the two alveolar stops. However, when comparing the results of the L1 and L2 participants ability to separate the unvoiced and voiced affricate, the difference was significant. The results suggest that the L2 participants had established new phonetic categories for the two nonnative phonemes, but that their ability to separate the two was not fully native-like. In the production data, the L1 participants performed better than the L2 participants in the affricates when the ratings of voiced and unvoiced productions were grouped, and the difference was significant. When dividing the data according to the voicing feature, the L1 participants no longer performed significantly better than the L2 participants on the unvoiced affricates, but the difference between the voiced counterpart remained significant (see table 15). However, the sample was small, containing only three observations of each participant (only one minimal pair containing the affricate was produced by each speaker) The small sample could have made the results of these tests highly sensitive to possible outliers in the data.

Looking at the L2 participants' results in the perception experiment, they showed a high mastery of the alveolar fricatives /s/-/z/. The comparison of the mean of correct responses for the L2 participants of the alveolar fricatives and the alveolar stops was not significant. However, the comparison of the L2 and L1 participants showed that the L2 participants did not perform as well as the L1 participants in the alveolar fricatives. The results suggest that the L2 participants had established a new phonetic category for the nonnative phoneme /z/, but that the participants' ability to separate the two was not fully native-like. In the production data, the L2 participants scored significantly lower than the L1 participants when the voiced and unvoiced data were grouped. When dividing the ratings according to the voicing feature, the L1 participants no longer performed significantly better than the L2 participants in the unvoiced alveolar fricative, but the difference between the voiced counterpart remained significant. This result was predicted because the unvoiced alveolar fricative is present in the Norwegian sound system, while the voiced alveolar fricative is not. The L2 participants would because they were much more familiar with the unvoiced alveolar fricative, be much more prone to produce errors in their productions that favored a correct production of the unvoiced alveolar fricative. The L2 participants were expected to perform well on the perception and production task because of the high-frequency of minimal pairs which contained the sound, and the presence of a separation between the voiced and unvoiced alveolar fricative in the orthography (<bus> vs. <buzz>). This hypothesis was supported by the findings, and the speakers performed well above chance in the perception experiment, and they showed some mastery of the contrast in their production. These results are similar to the ones found by Morland (2010) , since his results showed that the Norwegian L2 speakers of English did not use the ques of voicing to the same extent as L1 speakers when producing the voiced alveolar fricative /z/, even though they relied on the same ques as L1 speakers when separating the voiced and unvoiced alveolar fricative in perception.

Looking at the postalveolar fricatives /ʃ/-/ʒ/, the L2 participants showed a poor low accuracy in the perception task. The mean score on the postalveolar fricatives for the L2 participants was 60.2% (SD = 10%). When comparing the mean of the postalveolar fricatives was significantly lower than the means of the other sound pairs. The comparison of the L2 and the L1 participants on the postalveolar fricatives also showed that the L2 participants had a much lower mean accuracy than the L1 participants. The results suggest that the at least some of the L2 participants had not established a new phonetic category for the voiced postalveolar fricative, and that the L2 participants' ability to separate the two was not close to native-like. In the production data, the L2 participants received a lower score than the L1 participants for their production of the postalveolar fricatives when the voiced and unvoiced data were grouped, and the difference between the scores they received was significant.

When dividing the ratings according to the voicing feature of the target sound, the L2 participants were no longer significantly different from the L1 participants in the unvoiced postalveolar fricative, but the difference between the voiced counterpart remained significant. This result was predicted because the unvoiced postalveolar fricative is present in the Norwegian sound system, while the voiced postalveolar fricative is not. The L2 participants were expected to have a low accuracy on the perception and production task because of the low frequency of minimal pairs which contained the sound, and the absence of a separation between the voiced and unvoiced postalveolar fricative in the orthography (<mission> vs. <vision>). The results reflected the hypothesis, which stated the voiced postalveolar fricative might be a difficult to acquire as a Norwegian L2 speaker of English. Interestingly, the mean score of the production by the L2 participants of the postalveolar fricatives did not differ significantly from any of the means of production of the other sound pairs being investigated. This showed that the results of the production task did not follow the same pattern as the results of the perception task.

In the production data, the L2 group scored lower than the L1 group across all sound pairs when the voiced and unvoiced data were grouped. When dividing the data according to the voicing feature, the L1 group no longer performed significantly better than the L2 group in the voiced and unvoiced alveolar stops. This also reduced the significance in the difference of scores given to the voiced and unvoiced affricates when compared to the L1 group. The unvoiced alveolar fricatives and the unvoiced postalveolar fricatives were no longer significantly different from the L1 group, and the difference between the voiced counterparts of the fricatives remained significant.

The results showed a difference in the acquisition of the alveolar and postalveolar fricatives. This difference was hypothesized to be caused by the difference in the word-frequency and the number of minimal pairs containing the target contrast (Best and Tyler, 2007, p. 30), and the presence of a separation of the target contrast in the orthography (e.g. /z/ <buzz> vs. /s/ <bus>) (Best and Tyler, 2007, p. 32). SLM assumes that the volume of relevant input on native and nonnative sounds will influence a L2 speaker’s ability to acquire a nonnative phoneme (Flege, 1995). Other studies have also found similar result, such as van Dommelen and Hazan (2010), where the difference in percentage of correctly identified phonemes, voiced alveolar fricative /z/ (71.2%) and voiced postalveolar fricative /ʒ/ (50.7), was 20.5 % (2010, pp. 974-975).

There are two available explanations for it being harder to acquire the voiced postalveolar fricative, than the voiced alveolar fricative. One is the near absence of minimal pairs containing the postalveolar fricatives and the conventional orthography of those which exist. In the case of the post-alveolar fricatives /ʃ/-/ʒ/, the frequency of minimal pairs containing them as the contrasting sound is low, for example only 1 out of 4 words in the minimal pairs that I found were among the 5000 most frequent words (COCA, 2017) and I was unable to find any minimal pairs that had a higher word-frequency. Compare this to the alveolar fricatives /s/- /z/, where the contrast appears in several high-frequent minimal pairs and three out of four words with this contrast in my list of minimal pairs for the perception experiment were among the 5000 most frequent words. The postalveolar fricatives could also be harder to acquire because the sound does not have its own conventional spelling (e.g. <mission> and <vision>). Compare this to the alveolar fricatives where in some cases the voiced alveolar fricative /z/ is spelled with a <z>, such as in the word <buzz>.

5.4 The correlation between perception and production

The earlier research on the topic of hypercorrection (Eckman, Iverson and Song, 2013, Eckman and Iverson, 2013) suggest that if errors of hypercorrection were to occur, then this should happen when the L2 speaker is in the process of acquiring a nonnative phoneme with acoustic or articulatory properties similar to a native phoneme. It is also possible for there to be both hypercorrection errors and native transfer errors present in the speakers' language at the same time (Eckman and Iverson, 2013).

Since the phonetic inventory of Norwegian contains the unvoiced alveolar and postalveolar fricatives the speakers would not have been expected to make a production error in their attempt of making unvoiced fricatives unless they had entered the process of adding the voiced fricative into their phonetic inventory. The counts of wrong productions could possibly include 'slips-of-the-tongue' since the independent raters were asked whether the speaker produced the target sound and not whether the participant produced the voiced or unvoiced sound in the target sound pairs. The results of the ratings given on the production of the target consonants were still considered to apply to the topic hypercorrection and the patterns of the results were similar to other studies on hypercorrection (see Eckman, Iverson and Song (2013)).

The L2 participants were expected to, if they made errors, that most of these errors would be made in the production of the voiced fricatives. These errors would most likely be errors of native transfer (Eckman, Iverson and Song, 2013), where the L2 participant produced the unvoiced fricative in cases where the target phoneme was a voiced fricative. If the L2 participant did made errors in the production of the unvoiced fricative, then the error would be a hypercorrection error (Eckman, Iverson and Song, 2013) and the participant would be expected to perform above chance in the perception task. The errors in the affricates and the alveolar stops were not expected to pattern in any particular way. The results (table 9) support this hypothesis, and the participants that had their productions of the unvoiced alveolar or postalveolar fricatives labeled wrong, performed above chance (the lowest score was 57.7, and the highest was 77.5). The results in table 9 suggest that some of the L2 participants had established separate phonetic categories for the voiced and unvoiced alveolar fricatives, while some of the L2 participants performed at chance in the perception task and had no signs of hypercorrection in the results of their production data.

One can see in table 9 that the number of L2 participants making errors in the unvoiced postalveolar fricatives were higher than the number of L2 participants making errors in the unvoiced alveolar fricatives. The performance was well above chance on the voicing contrast in the alveolar fricatives in the perception task. These results suggest that the L2 participants had acquired the voiced alveolar fricative in their phonetic inventory, and that they might be past the point in the process of acquisition where they produce hypercorrection errors.

5.5. Limitations of this study

The ratings of the target consonants by the independent rater had a high variability, and this was the case both for the productions by native and nonnative speakers. This could have influenced the results of the production task, which might be why some of the differences found between sound pairs in the perception task did not show up in the results of the production task. The scope of this study did not make room for measuring durations in the sound files, so this is a possible continuation of this study. The production of the alveolar and postalveolar fricatives can be investigated in more detail and be compared to the results of Morland (2010).

This study did not investigate whether the results in the perception test were related to the acoustic-phonetic properties of the target consonants. Wim and Hazan (2010) also looked at the perception of L2 English in native Norwegian speakers, and they found the acoustic-phonetic properties of the nonnative consonants to be the cause of their results. Any influential acoustic-phonetic properties should also have affected the perception of the L1 speakers, but since their results reached ceiling-effect in the perception experiment, this question remains open.

This study only touched the surface of hypercorrection, and the manner of which the production data was collected and analyzed was not optimal for this research question. Further research on hypercorrection and native transfer in the voiced/unvoiced alveolar and postalveolar fricatives could potentially reveal results not found in this study.

6.0 Conclusion

The hypotheses of the experiments were based primarily the *speech-learning model* by Flege (SLM) (Flege, 1995) and the *perceptual assimilation model* suggested by Best, which was adapted to L2 learning by Best and Tyler (PAM-L2) (Best and Tyler, 2007, Best, 1995, 1994). The experiments failed to find a significant difference between L2 participants who had been on exchange to an English-speaking country, and the L2 participants who had never been on exchange to an English-speaking country. It was suggested that this might be because of the already high exposure of English to Norwegian speakers. This could mean that a stay in an English-speaking country of (up to) 10 months does not significantly improve perception or production of nonnative English consonants in Norwegians speakers. This suggests that highly proficient speakers of a L2 would require much more native input if it was to influence their phonetic inventory.

The analysis of the results failed to find significant linear models to describe the data from the two experiments. The most influential variable in the results of the perception experiment was *final grade in English in high school*. The most influential variable in the results of the production experiment was *self-evaluation of oral competence.* All measures of L2 competence correlated with the performance by the participants, and no correlations were found to any of the explanatory variables indicating exposure to native English input or amount of usage of English in every-day life.

All the results with respect to the different sound pairs conformed with the hypotheses of this project. The results on the alveolar stops showed that the L2 participants did not significantly differ from the L1 participants in neither of the two tests, which was expected since the alveolar stops are present in the phonetic inventory of Norwegian. The data from both the perception and production tests suggested that two new categories had been established in the phonetic inventory of the L2 participants for the unvoiced and voiced affricates /tʃ/ and /dʒ/. The performance on the unvoiced and voiced alveolar fricatives /s/ and /z/ resembled the results of previous research conducted on Norwegian speakers, and the results suggested that a new phonetic category had been added to the phonetic inventory of the L2 participants for the voiced alveolar phoneme/z/. However, the perceptions and productions of the L2 participants were not as good as the ones for the L1 participants. The perception of the unvoiced and voiced postalveolar fricatives /ʃ/ and /ʒ/ by the L2 participants were significantly less accurate than the perception of any of the other sound pairs. The mean accuracy was close to chance and the standard deviation was high. This indicates that some of the speakers might have acquired the contrast, while some did not perceive the difference in the minimal pairs at all.

Previously suggested variables that might influence the ease of acquiring a nonnative phoneme was supported by the results of these experiments. The L2 participants had a significantly harder time when separating the postalveolar fricatives, and this might be due to a low density in minimal pairs, and a low frequency in the ones that exist. Another possible explanation to why the alveolar fricatives appear to be easier to acquire is the presence of the contrast in the orthography.

The correlation between the results in the perception and production data was significant. The data also suggests that the participants who had errors that looked like hypercorrection errors had a low number of native transfer errors and they performed better than chance in the perception task.

This study investigated several factors which could have influence the likelihood of a L2 speaker acquiring a nonnative phoneme. For most part it seems as the results are determined by how competent the L2 speaker is, but the study fails to discover any other explanatory factors within the speakers. Different pairs of sounds were used to investigate how the volume of native input influence the acquisition of nonnative phonemes, and this was also seen in relation to whether the nonnative phoneme had a similar native phoneme. This study adds to a small body of research that has been conducted on first-language users of Norwegian and English as a second language.

Sources

Best, C. T. (1994) The emergence of native-language phonological influences in infants: A perceptual assimilation model, *The development of speech perception: The transition from speech sounds to spoken words*, 167(224), pp. 233-277.

Best, C. T. (1995) Chapter 6: A Direct Realist View of Cross-Language Speech Perception, *Speech perception and linguistic experience: Issues in cross-language research*, pp. 171-204.

Best, C. T. and Tyler, M. D. (2007) *Nonnative and Second-Language Speech Perception: Commonalities and Complementarities*. Amsterdam, Netherlands: Amsterdam, Netherlands: Benjamins.

Boersma, P. and Weenink, D. (2018) Praat: doing phonetics by computer [Computer program] (6.0.39 edn.). Available at: http://www.praat.org/ (Accessed: 3 April 2018).

Catherine, P. B. and Louis, M. G. (1986) Towards an Articulatory Phonology, *Phonology Yearbook*, 3, pp. 219-252. Available at: http://www.jstor.org/stable/4615400.

COCA (2017) *Corpus of Contemporary American English*. Available at: https://www.wordfrequency.info/ (Accessed: May 14th 2018).

Davidsen-Nielsen, N. (1977) *English phonetics*. Oslo: Gyldendal.

Eckman, F. and Iverson, G. K. (2013) The Role of Native Language Phonology in the Production of L2 Contrasts, *Studies in Second Language Acquisition*, 35(1), pp. 67-92. doi: 10.1017/S027226311200068X.

Eckman, F. R., Iverson, G. K. and Song, J. Y. (2013) The Role of Hypercorrection in the Acquisition of L2 Phonemic Contrasts, *Second Language Research*, 29(3), pp. 257-283. doi: 10.1177/0267658312467029.

Fintoft, K. (1961) The Duration of Some Norwegian Speech Sounds, *Phonetica*, 7, pp. 19-39.

Flege, J. E. and Hillenbrand, J. (1986) Differential Use of Temporal Cues to the /s/-/z/ Contrast by Native and Non-Native Speakers of English, *Journal of the Acoustical Society of America*, 79(2), pp. 508-517. doi: 10.1121/1.393538.

Flege, J. E. (1993) Production and Perception of a Novel, Second-Language Phonetic Contrast, *Journal of the Acoustical Society of America*, 93(3), pp. 1589-1608. doi: 10.1121/1.406818.

Flege, J. E. (1995) Second Language Speech Learning Theory, Findings, and Problems, in Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross Language Research.* Timonium, MD: New York Press, pp. 233-277.

Flege, J. E., Munro, M. J. and Mackay, I. R. (1995) Factors affecting strength of perceived foreign accent in a second language, *The Journal of the Acoustical Society of America*, 97(5 Pt 1), pp. 3125. doi: 10.1121/1.413041.

Fromkin, V. A. (1971) The non-anomalous nature of anomalous utterances, *Language*, pp. 27-52.

Hammond, M. (1999) *The phonology of English : a prosodic optimality-theoretic approach*. Oxford: Oxford University Press.

Iverson, P. *et al.* (2003) A perceptual interference account of acquisition difficulties for non-native phonemes, *Cognition*, 87(1), pp. B47-B57. doi: 10.1016/S0010-0277(02)00198-1.

Johnson, K. (2011) *Acoustic and Auditory Phonetics*. 3rd ed. edn. Chicester: Wiley.

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

Kristoffersen, G. (2015) Innføring i norsk fonologi, 4. utgave 2015.

Ku, H. and Zussman, A. (2010) *Lingua franca: The role of English in international trade*. Exhibited at 250-260.

Kunnskapsdepartementet (1997) *Læreplanverket for den 10-årige grunnskolen*. Available at: https://urn.nb.no/URN:NBN:no-nb_digibok_2008080100096.

Kunnskapsdepartementet (2006a) *English subject curriculum*. Available at: https://www.udir.no/kl06/ENG1-01?lplang=http://data.udir.no/kl06/eng.

Kunnskapsdepartementet (2006b) *Læreplan i fremmedspråk - programfag i utdanningsprogram for studiespesialisering*. Available at: https://www.udir.no/kl06/PSP1-01.

Kunnskapsdepartementet (2010) English subject curriculum. Available at: https://www.udir.no/kl06/ENG1-02?lplang=http://data.udir.no/kl06/eng.

Kunnskapsdepartementet (2013) *English subject curriculum*. Available at: https://www.udir.no/kl06/ENG1-03?lplang=http://data.udir.no/kl06/eng.

Lundskær-Nielsen, T., Barnes, M. P. and Lindskog, A. (2005) *Introduction to Scandinavian phonetics : Danish, Norwegian and Swedish*. København: Alfabeta.

Marinova-Todd, S. H., Marshall, D. B. and Snow, C. E. (2000) Three Misconceptions about Age and L2 Learning, *TESOL Quarterly*, 34(1), pp. 9-34. doi: 10.2307/3588095.

Mathôt, S., Schreij, D. and Theeuwes, J. (2012) OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314-324 doi: doi:10.3758/s13428-011-0168-7.

Morland, A. (2010) Norwegians' production and perception of the voiceless - voiced contrast in English (Mastergradsavhandling, NTNU). Trondheim: Morland, Aleksander.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available at: https://www.R-project.org/.

RStudio Team (2015) RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Available at: http://www.rstudio.com/.

Schütze, C. T. and Sprouse, J. (2014) Judgment data, *Research methods in linguistics*, pp. 27-50.

Skommer, G. (2014) Norsk fonologi - en utfordring for polske innlærere?, *NOA : norsk som andrespråk*, 30(2), pp. 67-85.

St. meld. nr. 9 (2011–2012) *Landbruks- og matpolitikken* Available at: www.regjeringen.no.

Utdanningsdirektoratet (2018) *Karakterer i videregående skole*. Available at: https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/karakterer-vgs/ (Accessed: 26.03.2018).

van Dommelen, W. A. and Hazan, V. (2010) Perception of English consonants in noise by native and Norwegian listeners, *Speech Communication*, 52(11), pp. 968-979. doi: 10.1016/j.specom.2010.05.001.

Vanvik, A. (1979) *Norsk fonetikk : lydlæren i standard østnorsk supplert med materiale fra dialektene*. Oslo: Universitetet i Oslo, Fonetisk institutt.

Vikør, L. S. (2017, 17th october) Språk I Norge, *Store norske leksikon*. Available at: https://snl.no/spr%C3%A5k_i_Norge.

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wittgenstein, L. and Ogden, C. K. (2014) *Tractatus Logico-Philosophicus: German and English*. Florence: Taylor & Francis Group.

Appendix 1 – Relevance for the teaching profession

On an academic level, this thesis addresses the acquisition of nonnative phonemes in English with respect to Norwegian as the first-language. This plays directly into the aims of the English curricula set for the English teaching in lower and upper secondary school. The curricula mention, among other aims, the aim «use the central patterns for pronunciation, intonation, word inflection and different types of sentences in communication» (Kunnskapsdepartementet, 2013). I interpret this aim to mean that the students should learn about the phonology of English, and that they should be made aware of the differences between the Norwegian and English phonology. My thesis goes in-depth on the two phonologies and does experimental work to see how the sounds are realized in both Norwegian speakers of English and native speakers of English. The purpose of the curriculum also states that «When we are aware of the strategies that are used to learn a language, and strategies that help us to understand and to be understood, the acquisition of knowledge and skills becomes easier and more meaningful» (Kunnskapsdepartementet, 2013). The knowledge I have acquired in my work with this thesis will help me at making the students aware of the differences that are between the two languages, and thus help ease their process of acquisition.

The process of writing this thesis has prepared me for both how to teach academic writing to the students, and how to teach the process of writing. Through my work, I have had to make hypotheses and find relevant theories to support these. The process of writing my thesis has made me more aware of statements of truth, which in fact are not truths at all. I started writing my thesis with the assumption in mind, that surely going on exchange and being exposed to a high volume of native input must play an effect, and help the speaker acquire nonnative phonemes. It appears that more research must be conducted on this topic if one wants to find out which explanatory factors helps ease the process of acquiring nonnative phonemes.

My results showed me that the only explanatory variables that could describe the results I got were which final grade the participant had received in the English subject in upper secondary school and at which level they evaluated their oral competence. Their self-evaluation of oral competence had a significant positive correlation with their attitude towards speaking English. However, none of my findings can tie these findings to exposure to native input. The questions of how to improve the students' self-evaluation of oral competence, attitude towards speaking English, and how exactly final grade is influencing the perception and production of these phonemes remain open.

Appendix 2 – Perception word list

batch	/bætʃ/
badge	/bædʒ/
lunch	/lʌn(t)ʃ/
lunge	/lʌndʒ/
Confucian	/kənˈfyu ʃən/
confusion	/kənˈfjuːʒən/
Aleutian	/əˈluːʃən/
allusion	/əˈluːʒən/
ice	/aɪs/
eyes	/aɪz/
bus	/bʌs/
buzz	/bʌz/
bat	/bæt/
bad	/bæd/
hat	/hæt/
had	/hæd/
spent	/spent/
spend	/spend/
location	/lə(ʊ)ˈkeɪʃn/
equation	/ɪˈkweɪʒən/, /ɪˈkweɪʃən/
mission	/ˈmɪʃən/
vision	/ˈvɪʒən/

Appendix 3 – Production word list

batch	/bætʃ/
badge	/bædʒ/
location	/lə(ʊ)ˈkeɪʃn/
equation	/ɪˈkweɪʒən/, /ɪˈkweɪʃən/
Confucian	/kənˈfyu ʃən/
confusion	/kənˈfjuːʒən/
Aleutian	/əˈluːʃən/
allusion	/əˈluːʒən/
mission	/ˈmɪʃən/
vision	/ˈvɪʒən/
ice	/aɪs/
eyes	/aɪz/
bus	/bʌs/
buzz	/bʌz/
bat	/bæt/
bad	/bæd/
hat	/hæt/
had	/hæd/

Appendix 4 – Questionnaire Norwegian speakers

Informant nr.:

Hvor gammel er du?

Hvilket klassetrinn tilhører du?

Hvilke fag/fagretninger har du studert på universitetsnivå?

Snakket du bare norsk hjemme under oppveksten?

Hvis nei; hvilke(t) andre språk snakket du under oppveksten?

List opp andrespråk du har lært på skolen/språkkurs/ect (untatt Engelsk) og ditt nivå (sett X).

Språk	grunnleggende	middels	høyt	Flytende

Har du vært på utveksling i et engelsk-språklig land (mer enn 6 uker)?

Hvis ja; hvor gammel var du?

Hvis ja; Hvor var du (land/stat)?

Hvis ja; hvor mange måneder var du der?

Har du bodd utenlands (mer enn 6 uker) (unntatt utveksling hvis du har nevnt det ovenfor)?

Hvis ja; hvor gammel var du?

Hvis ja; Hvor har du bodd

Hvis ja; hvor lenge har du bodd utenlands?

Hvor mange timer (ca.) i uka er du eksponert for muntlig engelsk, f.eks. spill, tv, musikk og skole?

Hvor mye engelsk snakker du I løpet av en uke? (gi et ca. timetall på hvor mye engelsk samtale du aktivt deltar i)

Hvordan vil du beskrive din holdning til det å skrive engelsk:

- ☐ 1 – Misliker sterkt å skrive engelsk
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 – Liker veldig godt å skrive engelsk

Hvordan vil du beskrive din holdning til det å snakke engelsk:

- ☐ 1 – Misliker sterkt å snakke engelsk
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 – Liker veldig godt å snakke engelsk

Hvordan vil du beskrive din holdning til engelskfaget når du gikk på skolen:

- ☐ 1 – Veldig negativ
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 – Veldig positiv

Hvordan vil du vurdere din muntlige kompetanse i engelsk:

- ☐ 1 – Veldig lav kompetanse
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 – Veldig høy kompetanse

Hva var din siste standpunktkarakter i engelsk?

Har du noen hørsel eller taleproblemer?

- ☐ Hørsel
- ☐ Tale
- ☐ Ingen kjente hørsel eller taleproblemer

Er det noe annet som kan ha påvirket din engelsk?

Appendix 5 – Questionnaire English speakers

Subject no:

How old are you?

Did you grow up speaking only English?

If not, which other language(s) did you grow up speaking?

Which dialect of English did you grow up speaking?

Please list any second languages that you speak and indicate your level of competence:

Language	Basic	Intermediate	Advanced	Fluent

Do you have any hearing or speech impairments?

- ☐ Hearing
- ☐ Speech
- ☐ No known hearing or speech impairments

Appendix 6 – Table of production and perception results

Participants marked in yellow belonged to the native English-speaking control group

Participant no.	sound	wrong v- %	wrong v+ %	perception %	Participant no.	sound	wrong v- %	wrong v+ %	perception %
13	postalveolar fricative	33.33	0.00	72.50	15	affricate	0.00	0.00	65.00
7	postalveolar fricative	25.00	0.00	57.50	2	affricate	0.00	0.00	70.00
16	postalveolar fricative	25.00	25.00	60.00	4	affricate	0.00	0.00	70.00
11	postalveolar fricative	16.67	0.00	65.00	14	affricate	0.00	0.00	70.00
9	postalveolar fricative	16.67	8.33	67.50	8	affricate	0.00	0.00	82.50
10	postalveolar fricative	8.33	0.00	77.50	3	affricate	0.00	0.00	85.00
6	postalveolar fricative	8.33	16.67	62.50	1	affricate	0.00	0.00	87.50
19	postalveolar fricative	8.33	0.00	92.50	19	affricate	0.00	0.00	87.50
14	postalveolar fricative	0.00	0.00	42.50	7	affricate	0.00	0.00	90.00
4	postalveolar fricative	0.00	0.00	57.50	9	affricate	0.00	0.00	90.00
17	postalveolar fricative	0.00	0.00	95.00	10	affricate	0.00	0.00	90.00
20	postalveolar fricative	0.00	0.00	97.50	12	affricate	0.00	0.00	90.00
21	postalveolar fricative	0.00	0.00	97.50	13	affricate	0.00	0.00	92.50
18	postalveolar fricative	0.00	0.00	100.00	11	affricate	0.00	0.00	97.50
22	postalveolar fricative	0.00	0.00	100.00	17	affricate	0.00	0.00	100.00
1	postalveolar fricative	0.00	16.67	50.00	18	affricate	0.00	0.00	100.00
8	postalveolar fricative	0.00	16.67	50.00	20	affricate	0.00	0.00	100.00
3	postalveolar fricative	0.00	16.67	57.50	21	affricate	0.00	0.00	100.00
2	postalveolar fricative	0.00	16.67	72.50	22	affricate	0.00	0.00	100.00
12	postalveolar fricative	0.00	25.00	52.50	6	affricate	0.00	33.33	80.00
5	postalveolar fricative	0.00	25.00	70.00	5	affricate	0.00	33.33	95.00
15	postalveolar fricative	0.00	91.67	47.50	16	affricate	0.00	100.00	57.50
16	alveolar fricative	50.00	50.00	60.00	14	alveolar stop	0.00	0.00	76.67
11	alveolar fricative	16.67	16.67	95.00	7	alveolar stop	0.00	0.00	85.00
12	alveolar fricative	0.00	0.00	80.00	1	alveolar stop	0.00	0.00	86.67
14	alveolar fricative	0.00	0.00	80.00	15	alveolar stop	0.00	0.00	86.67
6	alveolar fricative	0.00	0.00	85.00	11	alveolar stop	0.00	0.00	88.33
9	alveolar fricative	0.00	0.00	87.50	13	alveolar stop	0.00	0.00	88.33
10	alveolar fricative	0.00	0.00	92.50	8	alveolar stop	0.00	0.00	91.67
1	alveolar fricative	0.00	0.00	97.50	10	alveolar stop	0.00	0.00	91.67
17	alveolar fricative	0.00	0.00	97.50	12	alveolar stop	0.00	0.00	91.67
22	alveolar fricative	0.00	0.00	97.50	3	alveolar stop	0.00	0.00	93.33
18	alveolar fricative	0.00	0.00	100.00	4	alveolar stop	0.00	0.00	93.33
19	alveolar fricative	0.00	0.00	100.00	9	alveolar stop	0.00	0.00	93.33
20	alveolar fricative	0.00	0.00	100.00	18	alveolar stop	0.00	0.00	93.33
21	alveolar fricative	0.00	0.00	100.00	19	alveolar stop	0.00	0.00	93.33
13	alveolar fricative	0.00	16.67	82.50	6	alveolar stop	0.00	0.00	96.67
3	alveolar fricative	0.00	16.67	90.00	17	alveolar stop	0.00	0.00	98.33
7	alveolar fricative	0.00	16.67	92.50	21	alveolar stop	0.00	0.00	98.33
2	alveolar fricative	0.00	33.33	77.50	20	alveolar stop	0.00	0.00	100.00
4	alveolar fricative	0.00	33.33	77.50	22	alveolar stop	0.00	0.00	100.00
5	alveolar fricative	0.00	50.00	95.00	16	alveolar stop	0.00	16.67	68.33
15	alveolar fricative	0.00	66.67	82.50	5	alveolar stop	0.00	16.67	88.33
8	alveolar fricative	0.00	66.67	92.50	2	alveolar stop	0.00	33.33	90.00