PERCEPTION OF ENGLISH VOWELS BY ADVANCED FINNISH LEARNERS

Master's Thesis

Riina Kivikangas

University of Jyväskylä
Department of Languages and Communication Studies
English

May 2019

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistis-yhteiskuntatieteellinen	Laitos – Department Kieli-ja viestintätieteet
Tekijä – Author Riina Kivikangas	
Työn nimi – Title Perception of English Vowels by Advanced Finnish Learners	
Oppiaine – Subject Englannin kieli	Työn laji – Level Maisterintutkielma
Aika – Month and year Toukokuu 2019	Sivumäärä – Number of pages 69 + liitteet 3 sivua

Tiivistelmä – Abstract

Vieraan kielen oppija kohtaa usein äänteitä, jotka eivät sisälly äidinkielen inventaarioon, ja joiden havainnointi saattaa tästä syystä olla vaikeaa. Tämän tutkimuksen tavoitteena oli selvittää, miten edistyneet suomalaiset oppijat havaitsevat englannin kielen vokaaleja. Lisäksi tutkimus pyrki selvittämään onko konsonanttikontekstilla vaikutusta vokaalien havainnointiin ja esiintyykö oppijoiden havainnoinnissa vokaalien periferaalisuudesta johtuvia direktionaalisia epäsymmetrioita.

Tutkimukseen osallistui kaksikymmentä englantia yliopistossa opiskelevaa oppijaa. Tutkimus koostui kolmesta kuuntelutestistä, jotka selvittivät kuuden englannin kielen vokaalin identifikaatiota, assimilaatiota ja diskriminaatiota. Konsonanttikontekstin vaikutuksen selvittämiseksi vokaalit esitettiin sekä bilabiaalisessa että alveolaarisessa kontekstissa, ja direktionaalisten epäsymmetrioiden selvittämiseksi vokaaliparit esitettiin kummassakin suunnassa.

Suomalaiset opiskelijat identifioivat ja diskriminoivat englannin vokaaleja yleisesti ottaen korkealla tarkkuudella ja assimiloivat ne samoihin suomen kielen vokaaleihin. /ʌ/ oli jokaisessa kuuntelukokeessa hankalin vokaali. Konsonanttikontekstin vaikutus ei saavuttanut tilastollista merkittävyyttä suurimmassa osassa, mikä tukee aiempia tuloksia edistyneiden oppijoiden havainnoinnista. Havainnoinnista löytyneet direktionaaliset epäsymmetriat eivät kaikissa tapauksissa välttämättä selity periferaalisuuden avulla.

Tutkimuksen tulokset osoittavat havainnoinnin eri osa-alueiden olevan läheisessä yhteydessä. Niitä voidaan hyödyntää esimerkiksi ääntämisen opettamisessa keskittymällä suomalaisille oppijoille vaikeisiin vokaaleihin ja opettamalla helposti toisiinsa sekoitettavia äänteitä yhdessä.

Asiasanat – Keywords non-native language, perception, phonetics, phonology, vowels

Säilytyspaikka – Depository JYX

Muita tietoja – Additional information

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

When learning a new language, learners often come into contact with sounds that are not included in their L1 inventories. These sounds can be challenging not only to produce but also to perceive (e.g. Iverson and Evans 2009). Native Finnish speakers learning English must not only learn to perceive new phonemes but also new acoustic features, such as vowel tenseness, which are not utilised in their L1 but are phonemic in English.

Faced with such a task, learners often approach the new sounds through pre-existing categories and strategies, native or non-native. These begin to shape perception in infancy, and already by the first year of life, perception of many non-native contrasts is no longer successful unless there is continued exposure (Mazuka et al. 2014). While the native language can facilitate non-native perception (e.g. Pajak and Levy 2014), overuse of native phonotactics can also impede it. However, Ylinen et al. (2010), among others, have shown that perceptual training can enable learners to perform in a more native-like way, even if they cannot reach native-like accuracy levels.

Models of non-native perception have approached the phenomenon from many different viewpoints. One of the most influential models, Perceptual Assimilation Model - L2 (Best and Tyler 2007), regards assimilation of non-native phonemes to native categories as the basis of perception. When two or more non-native phonemes are assimilated to the same native category, PAM-L2 proposes that they will be difficult to discriminate. Natural Referent Vowel Framework (Polka and Bohn 2011), on the other hand, argues that perception of non-native vowels is influenced by their acoustic salience. Peripheral vowels, or vowels closer to the edges of the vowel space, are considered to be more salient. Unlike Perceptual Assimilation Model - L2, which is L1-dependent, Natural Referent Vowel Framework suggests that salience is a universal feature, although it can be modulated by the native language. These models form the theoretical framework of the current study and are introduced in more detail in chapter two.

Relatively little research on the field has focused on advanced learners as well as on those learning a non-native language in a classroom context, as opposed to in a country where the language is widely spoken (Balas 2018). Because both the amount and nature of exposure alter perception, results from studies concentrating on naïve listeners, beginner learners, or learners residing in L2-speaking countries cannot be widely generalised to explain the perception of other learners, however. Indeed, the results of for example Levy and Strange (2008) show that advanced learners and beginner learners perceive non-native sounds differently.

The native language and learner level, however, are not the only factors influencing non-native perception. Indeed, various linguistic, mental and environmental factors have all been shown to play a role. Because sounds exist in relation to other sounds, one topic of research in vowel perception has been the effect of the surrounding consonants (e.g. Balas 2018). The consonant context can influence the acoustic realisation of the following vowel and consequently affect the way it is perceived. Allophonic variation of this type rarely impedes perception of native phonemes, however, and sufficiently advanced learners may also be able to ignore it (Levy and Strange 2008).

The aim of the present study is to investigate how advanced Finnish learners perceive English vowels. University students (N=20) of English completed three listening experiments, which focused on sound identification, sound assimilation and sound discrimination. In order to consider the influence of context on vowel perception, the stimuli were elicited in bilabial and alveolar contexts. The effect of vowel peripherality, as proposed by Natural Referent Vowel Framework, was also investigated to see if it applied to advanced learners.

The structure of the paper is as follows. Firstly, the paper reviews research on non-native speech perception, beginning with infant studies and proceeding to studies on adults. Various factors influencing the phenomenon are also considered. These are followed by discussion on the theoretical models of Perceptual Assimilation Model - L2 and Natural Referent Vowel Framework. The Finnish and American English vowel systems are then overviewed and briefly compared before the current study is introduced. The remainder of the paper is devoted to an analysis of the results, concluding with a discussion chapter.

2. Non-native speech perception

The present chapter looks firstly at infant studies to establish the developmental patterns that speech perception (native and non-native) can follow. Although the study reported in this paper (see chapter 4) focuses on adults, infants are included here as infant perception forms the basis of adult perception. The chapter then discusses studies investigating adult non-native speech perception and possible factors influencing the perception of non-native contrasts. The chapter concludes with an overview of two influential models of non-native speech perception, the Perceptual Assimilation Model-L2 (Best and Tyler 2007) and the Natural Referent Vowel Framework (Polka and Bohn 2011), which are also used in the present study.

2.1. Patterns of development

It is commonly thought that infants are born with the ability to discriminate between all sound contrasts. According to this view, exposure to the ambient language(s) results in perceptual reorganisation, leading to a gradual loss of the ability to perceive contrasts that are non-phonemic in the native language. In a seminal study, Werker and Tees (1984) investigated the perception of two non-native contrasts (Nlaka'pamuctsin velar/uvular ejective contrast /k'i/—/q'i/ and Hindi dental/retroflex plosive contrast /t̪a/—/ʈa/) by English-speaking infants. They found that infants at 6–8 months were better at discriminating the non-native contrasts than infants at 8–10 months, and that at 10–12 months, almost none of the infants displayed successful discrimination. Conversely, when these contrasts were native, discrimination was reliable at all ages (Werker and Tees 1984). In other words, differences in the perception of native and non-native sounds are exhibited already during the first year of life.

However, some studies have found that the perception of native contrasts is not always easy. Liu and Kager (2016) showed that Dutch infants did not discriminate the native near-close/close front unrounded vowel contrast /ɪ/—/i/ until at 11–12 months. Similar results have been achieved e.g. by Narayan, Werker and Beddor (2010), who claimed that Filipino infants at 6–8 months could not reliably discriminate the native alveolar/velar nasal contrast /na/—/ŋa/, but succeeded at 10–12 months. This suggests that not all native sounds are equally easy to discriminate, and factors such as frequency of the sound or a feature in the L1, as well as its acoustic features (for example, whether the sound is acoustically very close to another native sound), can result in infants needing more time until successful discrimination is possible.

Mazuka et al. (2014) propose six possible patterns that can be used to explain the sometimes conflicting results of infant speech perception studies. The classification is used here because it provides a way of comparing patterns both within the sphere of native perception and between native and non-native perception. Two patterns (maintenance and enhancement) categorise native perception, and four patterns (decline, maintenance without exposure, no discrimination, enhancement without exposure) categorise non-native perception. Below, each pattern is considered separately in more detail.

Pattern 1: Maintenance of a native contrast. Infants may maintain discrimination of a native contrast through the process of perceptual reorganisation. Werker and Tees (1984), discussed above, showed that the Nlaka'pamuctsin velar/uvular ejective contrast /k'i/—/q'i/ and Hindi dental/retroflex plosive contrast /t̪a/—/ʈa/ were successfully discriminated by infants at different ages when native.

Pattern 2: Enhancement of a native contrast. In some cases, continued input in the native language results in improved discrimination of a native contrast. Kuhl et al. (2006) found that American infants discriminated the native alveolar/postalveolar lateral approximant contrast /l/—/ɹ/ better at 10–12 months than at 6–8 months. Sato, Kato and Mazuka (2012) showed that Japanese infants at 4 months could not discriminate the native single/geminate stop contrast contrast /t/—/tt/ but succeeded at 9.5 months.

Pattern 3: Decline of a non-native contrast. Lack of exposure to specific contrasts in the native language may result in a decline in sensitivity and a subsequent inability to discriminate these contrasts. As discussed above, Werker and Tees (1984) attested to the decline in English-speaking infants' ability to perceive non-native Nlaka'pamuctsin and Hindi contrasts. Numerous other studies have also confirmed this pattern. Mazuka et al. (2014) found that at 4.5 months, Japanese infants could successfully discriminate the non-native German close back/front rounded vowel contrast /uː/—/yː/. At 10 months, however, the infants could no longer discriminate between the sounds. Similarly, Bosch and Sebastián-Gallés (2003) showed that at 4 months, Spanish infants discriminated the non-native Catalan close-mid/open-mid front unrounded /e/—/ɛ/ vowel contrast, but that at 8 months, they no longer could do so.

Pattern 4: Maintenance of a non-native contrast without exposure. It is also possible for sensitivity to a non-native contrast to continue beyond the age where it typically begins to decline even though perceivers are not exposed to the contrast in their ambient language. Best, McRoberts and Sithole (1988) investigated the perception of the non-native isiZulu voiceless unaspirated apical/lateral click contrast /ʇ/—/ʖ/ by English-speaking infants (N=40). They found that infants successfully discriminated the contrast at 6–8 months, 8–10 months, 10–12 months and 12–14 months. Similarly, Tyler et al. (2014a), investigating English-speaking infants' (N=48) perception of the non-native Nuu-chah-nult fricative uvular/pharyngeal /χ/—/ħ/ and uvular/velar /χ—/x/ contrasts, found that discrimination was successful at both 6 months and 11 months.

Pattern 5: No discrimination of a non-native contrast. A third possible pattern for non-native contrasts is a lack of discrimination. Mazuka et al. (2014) found that Japanese infants at neither 4.5 months nor 10 months could discriminate the non-native German close/close-mid back rounded vowel contrast /uː/—/oː/. This lack of discrimination is expected to continue into adulthood (unless the individual is later exposed to the contrast). As with contrasts following pattern 3, this can result in difficulty in learning later in life if the individual encounters a language where these contrasts are phonemic.

Pattern 6: Enhancement of a non-native contrast without exposure. It is also possible that sensitivity to a non-native contrast is enhanced even in the absence of exposure to a language in which the contrast is phonemic. Mazuka et al. (2014) showed that while Japanese infants at 4.5 months could not discriminate the non-native German close/close-mid front unrounded vowel contrast /iː/—/eː/, they succeeded in discriminating the sounds at 10 months. Intriguingly, at 10 months, the infants were not able to discriminate the other contrasts investigated in the study (/uː/—/yː/ and /uː/—/oː/; see discussion on patterns 3 and 5). Similarly, de Klerk et al. (2018) investigated Dutch infants' perception of the non-native English open-mid/near-open front unrounded vowel contrast /ɛ/—/æ/. They found that infants could discriminate the sound at 6 months but not at 8 months; however, they succeeded again at 10 months.

To summarise, during the process of perceptual reorganisation that occurs during the first year of life, the discrimination of native and non-native contrasts can be maintained or enhanced, or it may decline. The investigation of developmental patterns here is limited to infant perceivers for two reasons: firstly, the majority of the studies reviewed included infants at different ages, thus allowing for the examination of developmental trajectories; and secondly, because, as per the paradigm stated in the beginning of the subchapter, perceptual reorganisation occurs in infancy. The developmental trajectories of adult perceivers have been investigated for example by Kim, Clayards and Goad (2018); they are not discussed here in detail. Conversely, concentrating on adults enables the investigators to better compare variables such as length of exposure and proficiency for physiological, cognitive and methodological reasons. Adult perception of non-native contrasts will be the focus of the following section.

2.2. Adult non-native speech perception

The present section aims at overviewing adult speech perception, specifically, adult non-native speech perception. Because the participants in the empirical study reported in this paper (see Chapters 4 and 5) are adult L2 learners,[1] the focus here is limited to studies investigating perception of non-native contrasts.

Meador, Flege and MacKay (2000) investigated the perception of non-native English vowels by L1 Italian speakers (N=72) residing in Canada. They divided the participants into three groups: Early (median age of arrival=7, median length of residence=40), Middle (median AOA=14, median LOR=34) and Late (median AOA=19, median LOR=28). They found that the discrimination accuracy of the Early group was equal to that of native English speakers, whereas both Middle and Late groups discriminated the contrasts less successfully, with the Late group's discrimination being the poorest (Meador, Flege and MacKay 2000). Similarly, Evans and Alshangiti (2018) investigated the perception of non-native English consonants and vowels by native speakers of Saudi Arabian Arabic (N=35, median LOR=3 years) residing in the United Kingdom. The participants were divided into two groups, High proficiency and Low proficiency, on the basis of a grammaticality test. Lower age of arrival, but not longer length of residence, was associated with higher proficiency. Although neither group's discrimination accuracy equalled that of native speakers, the High proficiency group performed overall better than the Low proficiency group, and consonants were consistently better discriminated than vowels (Evans and Alshangiti 2018). Taken together, these studies appear to suggest that the age at which exposure to a non-native language begins influences discrimination accuracy more than the length of the exposure.

As discussed in the previous section, the perceptual reorganisation paradigm suggests that development akin to that seen in infants is unlikely to be seen in adults. Indeed, Meador, Flege and Mackay (2000) found that only those participants whose exposure began at an early age

[1] The term "L2 learners" is used in the present study over "FLA (foreign language) learners" because of the increasingly important role of English in the Finnish society (see Leppänen, Nikula and Kääntä (2008) for an overview of English in Finland).

could match native speakers in discrimination accuracy. Studies investigating the possible effects of training, however, have shown that adult non-native speech perception is not monolithic but can undergo development. Ylinen et al. (2010) discovered that before training, Finnish speakers prioritised duration as a cue when discriminating the non-native English close/near-close front unrounded vowel contrast /iː/—/ɪ/. Given that Finnish places great importance on duration (Sajavaara and Dufva 2001), it is not surprising that Finnish learners may extensively rely on it as a cue to sound identity. Following training, however, the Finnish speakers could better utilise spectral cues, which are also used by native English speakers more than durational cues. Thus, even though the Finnish speakers' discrimination accuracy did not reach native levels following training, their perception became more native-like (Ylinen et al. 2010).

The type of training paradigm used may also play a role. Sennema, Hazan and Faulkner (2003) investigated native Japanese learners' (N=92) perception of the non-native English alveolar lateral/postalveolar approximant /l/—/ɹ/. They found that both audiovisual and audio-only training paradigms resulted in general improvement in discrimination, but that neither paradigm resulted in significantly better results than the other. In comparison, Hardison (2003), also examining native Japanese listeners' (N=16) perception of the non-native English /l/—/ɹ/ contrast, found that the audiovisual training paradigm led to more successful discrimination than the audio-only training paradigm. It is possible that the discrepant conclusions of Sennema, Hazan and Faulkner (2003) and Hardison (2003) are the result of for example variability in stimuli. Sennema and colleagues (2003) suggest that auditory cues may have been more salient than visual cues in their audiovisual stimuli. The possible lack of cue discrepancy in the stimuli of Hardison (2003) could explain why the audio-only paradigm did not result in equal improvement in her study.

It is also possible that individual differences between the participants in the two studies could explain the conflicting results. Perrachione et al. (2011), investigating how native American English speakers learned non-native pitch contrasts, found that individual differences were related to the training paradigm used: the performance of speakers whose pitch perception had been classified poor before training was impaired when the stimulus was highly variable, a condition that is usually considered to result in facilitated learning. In other words, inherent individual differences between participants can affect the efficiency of training. Consequently, comparing two training paradigms is likely to give more reliable results if the same group of participants is used, or if both studies equally consider individual differences.

The next section elaborates on individual differences, among other factors, as potential explanations for variability in non-native speech perception.

2.3. Factors influencing non-native speech perception

Numerous factors influence non-native speech perception. A fundamental one is clearly the class of the sound, that is, whether it is a vowel or a consonant. Syntactic processing relies more on vowels and lexical processing on consonants, and the acoustic properties of vowels are more salient (Bouchon et al. 2015); these differences also emerge relatively early in life, with consonants being preferred over vowels for lexical processing by 11 months (Poltrock and Nazzi 2015). Substitution studies also attest to the differences between vowels and consonants: speakers are more likely to modify the pseudoword *zobra* to *zebra* than to *cobra* (Cutler et al. 2000; although see Wiener and Turnbull 2016 for evidence that mutability may not apply to tonal languages). Carreiras et al. (2009) propose that different neural mechanisms are responsible for the processing of vowels and consonants. They showed that ERPs (event-related potentials, a brain response to stimulus or event) of native Spanish speakers were greater when the stimulus word was missing a consonant (e.g. *CHO O ATE*) than a vowel (e.g. *CHOC L TE*). Differences between vowels and consonants have also been found in studies of speech perception: Polka and Werker (1994) found that infants attune earlier to vowels (at 6—8 months) than they do to consonants (at 10—12 months). In other words, sensitivity to vowels begins to decline earlier than sensitivity to consonants. Indeed, as discussed in subchapter 2.1, Spanish infants could discriminate the non-native Catalan vowel contrast /e/—/ɛ/ at 4 months but not at 8 months (Bosch and Sebastián-Gallés 2003), suggesting that a process of perceptual narrowing had taken place.

Some of the studies overviewed in section 2.1 do not show early attunement to vowels, however (sixth pattern in Mazuka et al.'s [2014] classification). For example, Japanese infants were able to discriminate the non-native German vowel contrast /iː/—/eː/ at 10 months but not at 4.5 months (Mazuka et al. 2014), a feat that goes against Polka and Werker (1994). These results could be attributed to the influence of phonetically similar (but not identical) native phonemes

(Mazuka et al. 2014). Following this hypothesis, enhancement of a non-native contrast without exposure could potentially be related to the second pattern in Mazuka et al.'s (2014) classification, enhancement of a native contrast.

Larger-scale variation also plays a role. Evans and Alshangiti (2018), discussed in the previous section, found that Saudi Arabian adults perceived non-native English consonants more successfully than vowels. They proposed discrepancies between Arabic and English vowel and consonant inventory sizes as a possible explanation, noting that the size difference between the vowel inventories is bigger, which may result in Arabic speakers assimilating several English vowels to one native category (Evans and Alshangiti 2018). Cross-linguistically, the role of vowel inventory size was investigated by Iverson and Evans (2009). In their study, Spanish (N=17, vowel inventory=5) and German (N=16, vowel inventory=18) speakers were trained in the perception of English vowels. Both speakers were able to achieve the same post-training accuracy levels, although Spanish speakers required more training than the German speakers (Iverson and Evans 2009). In other words, although a bigger vowel inventory can offer an initial advantage, this advantage can be offset by training.

A better indicator of discrimination performance may be the acoustic closeness between the two languages (Elvin, Escudero and Vasiliev 2014; Alispahic, Mulak and Escudero 2017). The closer the phonemes of the languages are acoustically, the easier is it to discriminate non-native sounds. Section 2.1 discussed Best, McRoberts and Sithole (1988) in relation to the fourth pattern (maintenance of a non-native contrast without exposure) in Mazuka et al. (2014)'s classification. Best and colleagues (1988) proposed that the continued discrimination of the non-native isiZulu click contrast /ʇ/—/ʖ/ by English-speaking infants from 6 to 14 months could be because clicks are not included in the English language, meaning that they are unassimilable. (For further discussion on assimilation patterns, see section 2.2.1.) In other words, it is possible that speakers do not consider some non-native sounds as speech. This appears to be a somewhat extreme situation, however: the Spanish vowel inventory does not contain the Catalan open-mid front unrounded vowel /ɛ/, but Spanish speakers nevertheless recognise it as a speech sound (Bosch and Sebastián-Gallés 2003).

Language-specific factors are not limited only to vowel (or consonant) inventory sizes and acoustic closeness. Cue weighting, that is, whether perceivers use spectral or durational cues to identify sounds, was briefly discussed above in relation to Finnish. Several studies have also investigated the cue weighting strategies of speakers from other L1 backgrounds. McAllister, Flege and Piske (2002) studied the discrimination of Swedish vowels by native speakers of Spanish (N=20), English (N=20) and Estonian (N=20) who were all L2 Swedish learners. Duration differences are phonemic in Swedish as well as Estonian but not in English or Spanish. McAllister and colleagues (2002) found that Estonian speakers perceived quantity differences better than English speakers, who, in turn, performed better than Spanish speakers, suggesting that experience with duration in L1 can extend to L2. However, simply utilising similar cue weighting strategies does not necessarily lead to successful discrimination. Escudero, Benders and Lipski (2009) found that, although L1 naïve German speakers (N=31) preferred spectral cues as did L1 Dutch speakers (N=31), their discrimination of non-native Dutch contrasts was less successful than that of native Dutch speakers. In other words, L2 learners who utilise a different cue weighting strategy may perform better than naïve speakers who utilise the same cue weighting strategy as target language speakers. Nevertheless, cue familiarity can enhance between-class performance: experience with vowel duration as a phonemically distinguishing feature in L1 facilitates discrimination of differences in non-native consonant length (Pajak and Levy 2014). In other words, listeners can apply a familiar phonetic feature to contexts where it is not used in the L1.

Orthography is another factor that, like cue weighting and acoustic closeness, varies between languages. Escudero and Wanrooij (2010) investigated the perception of Dutch vowels by speakers of L1 Spanish (N=204). Spanish, like Finnish, has a relatively straightforward orthography, where one phoneme usually maps onto just one grapheme. In comparison, Dutch and English have what are termed as deep orthographies, whereby one phoneme can map onto several graphemes. Escudero and Wanrooij (2010) found that L1 (Spanish) orthography affected the orthographic labelling of L2 (Dutch) auditory stimuli. To illustrate, in an auditory-only task, Spanish speakers identified the Dutch vowel /y/ with the Spanish vowel /u/, labelled in Spanish as <u>. When asked to label the Dutch /y/, the majority of Spanish speakers chose <u> instead of the correct <uu> (Escudero and Wanrooij 2010). Of course, orthographic labelling tasks are not particularly natural situations; however, Escudero and Wanrooij's (2010) results carry implications for methodology in speech perception experiments to minimise orthographic interference.

Finally, because sounds do not exist in a vacuum, the vowel/consonant context in which they appear can also affect perception through differences in production as well as L1 coarticulatory interference. Bohn and Steinlen (2003) investigated the perception of English vowels by L1 Danish speakers (N=30) in three different consonant contexts (/hVt/, /dVt/ and /gVk/). They found that English vowels were assimilated to different Danish categories depending on the consonant context; for example, in both the /hVt/ and /dVt/ contexts the English /ɪ/ was assimilated to the Danish /e/, but in the /gVk/ context it was assimilated to the Danish /i/. Similarly, Levy and Strange (2008) found evidence for the influence of consonant context in their study of the perception of French vowels, presented in /rabVp/ and /radVt/ contexts, by L1 American English (N=20) speakers. In contrast to Bohn and Steinlen (2003), Levy and Strange (2008) also investigated the effect of listener experience by comparing experienced (had studied French and had spent time in France) and inexperienced (had never studied French and had not spent time in France) listeners. They discovered that inexperienced listeners had more difficulty perceiving vowels in the /radVt/ context than in the /rabVp/ context, whereas for experienced listeners, the context effect was not statistically significant. Context thus appears to be a possible factor in speech perception, at least together with other factors, such as language experience.

Naturally, differences exist not only between languages but also between language users. Consequently, intra-speaker factors can also influence non-native speech perception. Differences related to factors such as age of arrival and length of residence (see e.g. Baker and Trofimovich 2006) have been discussed briefly in section 2.2 and are not considered in more detail here.

The ability to successfully perceive non-native contrasts depends on the auditory and processing systems. It has been suggested that individual differences in L2 performance are based on speech-specific abilities: discrimination of non-native speech sounds, but not non-speech sounds (such as tonal stimuli), was linked to successful perception in Dutch speakers of L2 English (Díaz et al. 2016). Conversely, as discussed above, Best, McRoberts and Sithole (1988) proposed that the continued discrimination of the non-native isiZulu click contrast /ʇ/—/ʖ/ by English-speaking infants may be because the clicks are not considered speech sounds by English listeners.

Differences between sound classes can also be distinguished. Recall that Carreiras et al. (2009) found support for separate processing of vowels and consonants in the brain. Omote, Jasmin and Tierney (2017), investigating the link between auditory processing and non-native perception, found that greater frequency following responses (FFR, a brain response to periodic auditory stimuli) were related to successful discrimination of non-native consonants (but not vowels) in Japanese learners of L2 English. Thus, in addition to a general division, it appears that individual brain responses may also play a role.

Intra-speaker variation is not limited to intrinsic physiological factors. Kartushina and Frauenfelder (2013) investigated the relationship between native production and non-native perception. They found that L1 Spanish speakers (N=14) whose productions of the native mid front unrounded vowel /e/ were acoustically close to the non-native French close-mid front unrounded vowel /e/ were more successful at discriminating the non-native sound. Furthermore, the more compact the L1 Spanish speakers' phonological space for the native /e/, the better they were able to discriminate the non-native French open-mid front unrounded vowel /ɛ/ (Kartushina and Frauenfelder 2013). Recall that Mazuka et al. (2014) proposed native allophonic variation as a possible explanation for the enhanced discrimination of a non-native German contrast by Japanese infants with no exposure German. Acoustic closeness, discussed above in more general terms, thus appears to be a factor on the individual level as well.

Production is only one aspect of an individual's linguistic repertoire. Bundgaard-Nielsen, Best and Tyler (2011) investigated the role of L2 English vocabulary size on the assimilation of Australian English vowels by L1 Japanese learners (N=11). They found that learners in what was classified as the High-vocabulary group (Median vocabulary=7200) identified the target vowels with more consistency than the Low-vocabulary group (Median vocabulary=5017). While vocabulary size is certainly not the only factor that influences L2 discrimination, it appears at least to be beneficial to it, possibly as learners with bigger vocabularies have more exposure to non-native sounds in different contrasts, leading to more successful assimilation (Bundgaard-Nielsen, Best and Tyler 2011). The effect is similar to that of allophonic variation as discussed in relation to Mazuka et al. (2014). The apparent relation between vocabulary size and L2 perception also suggests that different sides of linguistic knowledge (here, lexis and phonology/phonetics) interact with each other (see also Escudero and Wanrooij [2010] for discussion on orthography).

Finally, there are, of course, factors outside language and language users, that is, the environment around them. In real life, the environment is rarely optimised for the non-native learner. Several studies have examined non-native speech perception under varying task conditions in an attempt to shed more light on the demands placed on the learner in real communicative situations. Asano (2018) investigated the impact of increased memory load (through an extended interstimulus interval) and attention control (through unrelated pitch changes) on the perception of non-native consonant length contrasts by German learners (N=48) of L2 Japanese. She found that increased memory load alone did not significantly impair performance, whereas increased demand on attention control was enough to result in less reliable discrimination on its own. Discrimination accuracy was poorest when both memory and attention control were under increased demand (Asano 2018). In other words, especially situations that force the learner to focus his or her attention amidst multiple competing stimuli can lead to weakened performance in the non-native language. Consequently, it is difficult to generalise the results of perception experiments performed in laboratory conditions to real communicative situations.

To conclude, the present section has aimed at offering an overview of some language-specific, individual-specific and environmental factors that may influence non-native speech perception. Many of these factors are interrelated, and one aspect of perception (such as differences in the discrimination of consonants and vowels) cannot be attributed to a single factor only.

The discussion now turns to theoretical models of non-native speech perception, more precisely, the Perceptual Assimilation Model-L2 (Best and Tyler 2007) and the Natural Referent Vowel Framework (Polka and Bohn 2011).

2.4. Models of non-native speech perception

Several theoretical models have been created to explain the phenomenon of non-native speech perception. Although these models have different approaches to the subject, they generally share the proposition that non-native perception is influenced by the native language. They include the Speech Assimilation Model (SLM; Flege 1995), the Native Language Magnet Model and its expanded version (NLM and NLM-e; Kuhl et al. 2008), the Automatic Selective Perception model (ASP; Strange 2011), the Second Language Linguistic Perception model (L2LP; van Leussen and Escudero 2015), the Perceptual Assimilation Model and its L2-specific version (PAM and PAM-L2; Best 1995, Best and Tyler 2007) and the Natural Referent Vowel Framework (NRVF; Polka and Bohn 2011). The present overview will focus on the latter two models because of their relevance for the study reported in this paper (see chapters 4 and 5): PAM-L2 because it specifically aims at describing speech perception by L2 listeners and NRVF because it specifically aims at describing vowel perception. The following two sections discuss each model in more detail.

2.4.1. Perceptual Assimilation Model-L2

Perceptual Assimilation Model-L2 (Best and Tyler 2007) is an extended version of Perceptual Assimilation Model (Best 1995), which focuses on non-native speech perception by naïve listeners. Central to perception as modelled by PAM-L2 are the notions of phonological distinctiveness and phonological constancy. Phonological distinctiveness refers to the ability of differences of one feature to cause a change in a word's meaning (Best et al. 2009: 539). In English, changing the height of a front unrounded vowel from the near-close /ɪ/ to the near-open /æ/ results in a phonologically distinctive difference, as illustrated by the minimal pair /bɪt/—/bæt/. Phonological constancy, on the other hand, refers to differences that do not cause a change in a word's meaning (Best et al. 2009: 539). Extending the duration of the near-open front unrounded vowel /æ/ results in a token that sounds unusual but whose meaning has not changed. Best et al. (2009) note that perceivers must learn phonological distinctiveness and phonological constancy early on, because they are necessary for understanding that a phoneme can be realised differently for example because of physiological or dialectal differences (allophonic variation; see discussion above in relation to Mazuka et al. [2014]).

The notions of phonological distinctiveness and phonological constancy shed light on the different levels (phonological, phonetic and gestural) that perceivers need to operate on once they encounter non-native sounds (Best and Tyler 2007). In differentiating such sounds, learners need to attend to both the phonetic and the phonological level (Best and Tyler 2007: 23). In other words, it is not enough that learners consider physical properties of sounds, they also need to consider the role of the sounds in the language: the English near-close back rounded vowel /ʊ/ and the close back rounded vowel /uː/ cannot be regarded simply as allophones of the same sound, as illustrated by the existence of the minimal pair /fʊl/—/fuːl/.

PAM-L2 proposes that the basis of non-native perception is the ability to distinguish articulatory gestures as the result of perceptual learning (Best and Tyler 2007: 20). Articulatory gestures are defined as the "active movements of one or more articulators (lips, tongue tip, tongue body, tongue root, velum, glottis) to achieve constrictions of varying degrees (closed, critical, narrow, mid, wide) at specific locations within the vocal tract (e.g. at the lips, upper front teeth, alveolar ridge" (Best et al. 2009: 540). It is through this process that perceivers form new phonetic categories or modify existing ones to account for within-category variation (Best and Tyler 2007: 21-25).

The studies reviewed in subchapters 2.1. and 2.3. have clearly demonstrated that the formation of new phonetic categories is not an easy task, and that categories do not always remain stable beyond the first few months of life. Nevertheless, people are capable of understanding and using non-native languages. PAM-L2 argues that, in the absence of native-like phonetic categories, learners assimilate non-native sounds to L1 categories (Best and Tyler 2007: 27). A total of six assimilation patterns can be distinguished (Best, McRoberts and Goodell 2001):

1. **Two Category (TC).** Two non-native sounds are assimilated to two separate native sounds.
2. **Single Category (SC).** Two non-native sounds are assimilated to one native sound.
3. **Category Goodness (CG).** Two non-native sounds are assimilated to one native sound, but one is considered a better exemplar than the other.
4. **Uncategorised-Categorised (UC).** One non-native sound is assimilated to one native sound, while the other non-native sound is not assimilated to any native sound (but is recognised as speech).
5. **Uncategorised-Uncategorised (UU).** Neither non-native sound is assimilated to any native sound (but both are recognised as speech).
6. **Non-Assimilable (NA).** Neither non-native sound is assimilated to any native sound (neither is recognised as speech).[2]

On the basis of assimilation patterns it is possible to make predictions regarding discrimination accuracy (Best, McRoberts and Goodell 2001). In TC and UC cases, the non-native sounds should be discriminated in an excellent manner because they assimilate to two separate native phonetic categories. Similarly, discrimination in NA cases should also be excellent to good, depending on the perceived similarity between the two sounds. In UU cases, discrimination should vary between good and intermediate, again depending between the perceived similarity

[2] See Best, McRoberts and Sithole (1988). To the author's knowledge, no study has investigated how English learners of isiZulu assimilate the click contrasts.

between the two sounds. In CG cases, discrimination is predicted to be intermediate, depending on the strength of the perceived quality difference between the sounds. Finally, SC cases are predicted to have the poorest discrimination, as both sounds are assimilated to one native phonetic category.

The six patterns and their predicted difficulty levels are illustrated in Table 1.

Assimilation pattern	Explanation	Difficulty
Two Category	A→ B, C→ D	Easy
Single Category	A→ B, C→ B	Difficult
Category Goodness	A→ B, C→ B	Intermediate
Uncategorised-Categorised	A→ B D, C→ E	Easy
Uncategorised-Uncategorised	A→ B D, C→ E F	Easy to intermediate
Non-Assimilable	A→ *, C→ *	Easy

Table 1. Assimilation patterns according to PAM-L2 (Best et al. 2001).

To summarise, PAM-L2 proposes that perceptual relationships between two languages can be classified into six patterns based on how non-native sounds are assimilated to native sounds. Furthermore, PAM-L2 predicts that assimilation patterns can be used to predict discrimination accuracy, with TC contrasts being the easiest and SC contrasts the most difficult to discriminate. As a result, PAM-L2 is suitable for the present study (see chapter 4), which investigates both assimilation and discrimination. Because it is not class-specific, PAM-L2 can be used to analyse both consonant and vowel discrimination. The discussion now turns to the Natural Referent Vowel Framework (Polka and Bohn 2011), a model that concentrates on vowels.

2.4.2. Natural Referent Vowel Framework

PAM-L2, as discussed above, focuses on relations between L1 and L2 sounds as encapsulated by the six possible assimilation patterns. The Natural Referent Vowel Framework (Polka and Bohn 2011), on the other hand, focuses on what are considered to be language-independent vowel features.[3] More specifically, NRVF attempts to explain the presence of directional

[3] See Nam and Polka (2016) for an application of NRV principles to consonants.

asymmetries in vowel perception, a phenomenon attested to in numerous studies on infant speech perception (Polka and Bohn 2011: 467; see Polka and Bohn [2003] for a review of relevant studies).

Consider Figure 1, which shows the complete International Phonetic Alphabet (IPA) vowel chart.

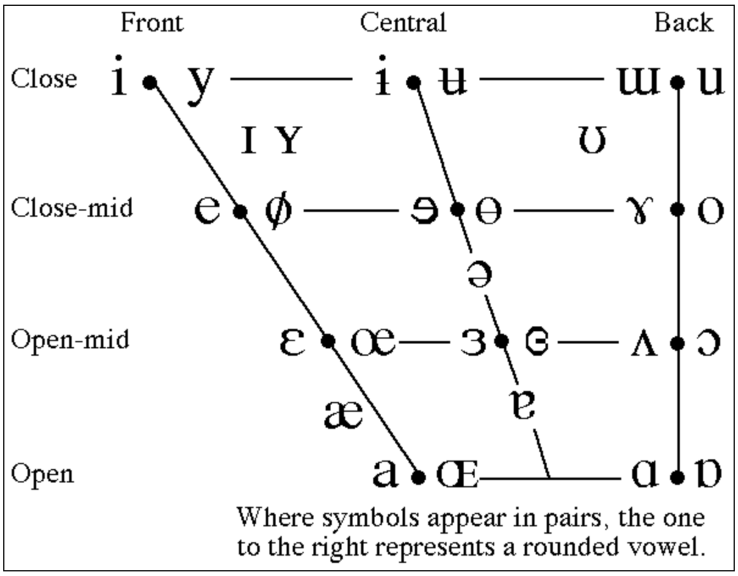

Figure 1. The IPA vowel chart. Adapted from UCLA Phonetics Lab.

The vowel space is defined by peripheral vowels, such as the close front unrounded vowel /i/ and the open front unrounded vowel /a/. Infant studies appear to show that a change from a more peripheral vowel to a more central one is less sensitive for discrimination than the other way around (Polka and Bohn 2011: 467). For example, Polka and Werker (1994) found that English-speaking infants (N=34) discriminated the non-native German close back/near-front rounded vowel contrast /uː/—/yː/ and near-close near-back/near-front rounded vowel contrast /ʊ/—/ʏ/ in an asymmetric manner, that is, discrimination was easier from /yː/ to /uː/ and /ʏ/ to /ʊ/ than the other way around, respectively.

NRVF suggests that the basis of this bias is formant frequency convergence at the edges of the vowel space, which causes a raise in each amplitude, resulting in focalisation. In other words, the peripheral vowels are more acoustically salient than the central vowels (Polka and Bohn 2011: 474). This salience makes these vowels easier for the infant listener to perceive. Given the ubiquity of /i/, /a/ and /u/ in the vowel inventories of languages around the world, attraction to peripheral vowels is presumably universal (Polka and Bohn 2011: 474-475). Following exposure to the native language, listeners either maintain or eliminate this initial bias in a manner that best facilitates attunement to the L1 phonological system (Polka and Bohn 2011: 474).

Directional asymmetries for non-native contrasts have also been documented in adult listeners. Kriengwatana and Escudero (2017) investigated Spanish learners (N=79) of L2 Dutch, showing that discrimination of the Dutch open back/front unrounded vowel contrast /ɑ/—/aː/ was more accurate from /aː/ to /ɑ/ than vice versa. Conversely, Balas (2018), investigating the perception of English vowels by adult Polish learners (N=35) of L2 English, found no directional asymmetries. She considered two explanations for this: firstly, she utilised an AXB discrimination paradigm whereas Polka and Bohn (2011) utilised a go/no-go paradigm; secondly, the participants in her study were advanced learners. In comparison, the participants in Kriengwatana and Escudero (2017) were beginner-to-intermediate learners. Thus, methodology, as well as the learner status of the participants, may influence the presence of directional asymmetries.

This is further supported by Tyler et al. (2014b), who investigated the perception of French, Norwegian and Thai vowels by naïve American English listeners (N=13). They found directional asymmetry in contrasts assimilated in the single category pattern (as per PAM-L2), but not in other patterns. In contrast to Balas (2018) and Kriengwatana and Escudero (2017), the participants in Tyler et al. (2014b) had no previous experience with the languages studied. This could suggest that adult speakers exhibit directional asymmetry at least for single category non-native contrasts. As Kriengwatana and Escudero (2017) did not include assimilation patterns in their analysis, their data neither support nor deny the hypothesis. Balas (2018), as noted above, found no directional asymmetries for any pattern, a result she attributed partially to the advanced status of the participants. Recall that Polka and Bohn (2011) found that adult German speakers tested showed no directional asymmetries in their discrimination of the native /uː/—/yː/ and /ʊ/—/ʏ/ contrasts; consequently, advanced learners might perceive vowels in a sufficiently native-like manner that directional asymmetries are no longer present, even in single-category assimilations (Balas 2018).

To summarise, NRVF suggests that directional asymmetries, based on vowel peripherality, are a universal feature of both native and non-native vowel perception. Both infant and adult listeners have been found to exhibit directional asymmetries, although adult studies appear to suggest that methodology and learner status may affect their visibility in the data and, in the case of advanced listeners, potentially lead to their absence.

The discussion now turns to a description of the vowel inventories of Finnish and English, respectively.

3. Finnish and English vowels

The present chapter overviews the Finnish and American English[4] vowel systems, concluding with a brief acoustic analysis of Finnish and American English vowels.

3.1. Finnish vowel system

The Finnish phonological system includes the following eight monophthong vowels, classified in Table 2 in terms of height, backness and roundedness:

Vowel	Height	Backness	Roundedness
i	Close	Front	Unrounded
y	Close	Front	Rounded
u	Close	Back	Rounded
e	Mid	Front	Unrounded
ø	Mid	Front	Rounded
o	Mid	Back	Rounded
æ	Near-open	Front	Unrounded
ɑ	Near-open	Back	Unrounded

Table 2. Finnish monophthongs. Adapted from Suomi et al. (2008: 20).

[4] American English is used because it is spoken by the native speakers in the present study.

Vowel height is based on whether the tongue is positioned towards the upper or lower part of the mouth, backness on whether the tongue is positioned towards the front or back of the mouth, and roundedness on whether the lips are rounded or not (Edwards 2003).[5] Finnish does not utilise all the available dimensions: for example, the Finnish vowels /æ ɑ/, which are sometimes categorised as open, are more precisely near-open, and the Finnish vowel system does not contain any central vowels (vowels where the tongue is positioned centrally with regards to the front and back of the mouth) (Suomi et al. 2008).

However, the Finnish system has several features that make it distinct from many other vowel systems, including the American English one. Firstly, vowel length is contrastive in Finnish: each vowel phoneme can be long or short, resulting in meaning-distinguishing pairs such as /tuːli/ "wind" and /tuli/ "fire" (Suomi et al. 2008: 23). In terms of PAM-L2, length is a phonologically distinctive feature in Finnish. As discussed above, this can result in Finnish speakers relying heavily on duration when identifying vowel sounds (Ylinen et al. 2010).

Secondly, Finnish follows vowel harmony. The eight Finnish vowel phonemes can be divided into three classes (front=/y ø æ/, back=/u o ɑ/, neutral=/i e/) on the basis of whether they can appear in the same (native) word or not (Suomi et al. 2008: 51). It is possible to combine vowels from the front (e.g. /tytːø/ "girl") , front and neutral (e.g. /æiti/ "mother"), neutral (e.g. /tie/ "road"), back (e.g. /lupɑ/ "permission") and back and neutral (e.g. /teko/ "act") classes; other combinations, such as front, neutral and back /olympiɑlɑiset/ "the Olympic Games" break the vowel harmony, and Finnish speakers consequently often have difficulties in realising them, producing instead the phonotactically acceptable */olumpiɑlɑiset/ (Suomi et al. 2008).

In summary, the Finnish vowel system consists of eight monophthong vowels, which can be classified according to height, backness and roundedness. Vowel length and vowel harmony are characteristic features of the Finnish vowel system that separate it from the American English vowel system, which is the focus of the next section.

---

[5] Note that in the present study, the terms *close*, *mid* and *open* are used to describe the vowels; *high*, *mid* and *low*, respectively, can also be used.

3.2. American English vowel system

General American English contains ten monophthong vowels, classified in Table 3 according to height, backness, roundedness and tenseness:

Vowel	Height	Backness	Roundedness	Tenseness
ɪ	Near-close	Front	Unrounded	Lax
i	Close	Front	Unrounded	Tense
ʊ	Near-close	Back	Rounded	Lax
u	Close	Back	Rounded	Tense
ɛ	Open-mid	Front	Unrounded	Lax
ə	Mid	Central	Unrounded	-
ʌ	Open-mid	Back	Unrounded	Lax
ɔ	Open-mid	Back	Rounded	Tense
æ	Near-open	Front	Unrounded	Lax
ɑ	Open	Back	Unrounded	Tense

Table 3. General American English monophthongs. Adapted from Edwards (2003).

In contrast to the Finnish vowel system, American English vowels are more varied with regards to height and backness. They can also be classified on the basis of tenseness, that is, how much muscular effort is required to produce the sound (Edwards 2003). Thus a native speaker of Finnish encountering the American English vowel system will have to perceive, identify and produce not only new vowel phonemes but also new acoustic features (tenseness). Furthermore, unlike in Finnish, vowel length is not a phonemic feature in American English, and the vowel system also does not adhere to vowel harmony (Edwards 2003).

The discussion now turns to an acoustic comparison of Finnish and American English vowels.

3.3. Comparison of Finnish and American English vowels

The present section aims at offering a brief acoustic comparison of Finnish and American English vowels.[6] As discussed in 2.3, acoustic closeness is a good indicator of discrimination

[6] The author is aware of one comparative study of Finnish and English (Wiik 1965); as it focuses on Received Pronunciation speakers, its data are not applicable here.

performance, and thus comparisons of the Finnish and American English vowel systems can illustrate potential problems in cross-language sound perception. The Finnish data in the present section come from Iivonen (2012), while the American English data come from Hillenbrand et al. (1995). Data from the current study are discussed in more detail in section 4.2.

Acoustic analysis of vowel phonemes is typically based on formants, that is, peaks of acoustic resonance (Edwards 2003). Vowel features (notably, height and backness) are linked to formant values (Ogden 2009: 62-63). Firstly, vowel height and first formant (F1) value correspond so that vowels in each of the height categories (close, mid, open) are near each other and distant from vowels in the other categories. Mid and open vowels have higher F1 values than close vowels. Secondly, vowel backness and second formant (F2) value correspond so that front vowels are closer to each other than to back vowels and vice versa. Front vowels have higher F2 values than back vowels. In other words, F1 and F2 values can separately be used to investigate a vowel's height and backness; together, they can be used to identify vowels.

Hillenbrand et al. (1995) recorded American English vowels in a /hVd/ context by 45 men and 48 women from Midwestern United States. Table 4 shows the average F1 and F2 values for American English monophthongs /ɪ i ɛ ʌ ɔ æ/ as averaged across all the speakers:

Vowel	F1	F2
ɪ	427	2034
	483	2365
i	342	2322
	437	2761
ɛ	580	1799
	731	2058
ʌ	623	1200
	753	1426
ɔ	652	997
	781	1136
æ	588	1952
	669	2349

Table 4. Average F1 and F2 values (Hz) for American English vowels /ɪ i ɛ ʌ ɔ æ/.
The data are averaged across the productions of 45 male speakers (upper row) and 48 female speakers (bottom row) from the Midwestern United States. Adapted from Hillenbrand et al. (1995: 3103).

As illustrated by Table 4, open and mid vowels /ɛ ʌ ɔ æ/ have higher F1 values than close vowels /ɪ i/. As the formant-feature relation is converse, /ɪ i/ are produced higher in the mouth. Similarly, front vowels /ɪ i ɛ æ/ have higher F2 values than back vowels /ʌ ɔ/; this relation is not converse. Differences can also be distinguished between male and female speakers: on average, male speakers have lower formant values, reflecting physical properties of the vocal tract.

Formant data for Finnish are offered by Iivonen (2012). He does not elaborate on the consonant contexts in which the vowels were elicited. He also presents data only from a male speaker; as shown above, productions of male and female speakers are generally different, and consequently only the male productions from Hillenbrand et al. (1995) are used here to compare Finnish and American English.

Table 5 shows the average F1 and F2 values of the Finnish short monophthongs /y u e ø o æ ɑ/ and the long monophthong /iː/:

Vowel	F1	F2
i:	294	2039
y	330	1601
u	332	690
e	443	1722
ø	436	1451
o	433	958
æ	600	1535
ɑ	609	1256

Table 5. Average F1 and F2 values (Hz) for Finnish vowels.
Data averaged across 20 productions of a male native speaker. Adapted from Iivonen (2012: 31).

Again, mid and open vowels, here /e ø o æ ɑ/, have higher F1 values than close vowels, here /i y u/. Front vowels /i y e ø æ/ have higher F2 values than back vowels /u o ɑ/.

Brief comparisons can be made between Finnish and American English. Because Iivonen's (2012) data come from only one speaker, generalising differences on the basis of the two data discussed here is impossible. However, as the purpose of the present study is not to offer a complete acoustic comparison of Finnish and American English vowel inventories, the data were deemed sufficient to account for its needs.

In terms of height, the English near-close front vowel /ɪ/ appears closest to the Finnish mid back vowel /o/, while in terms of backness, it is almost identical to the Finnish close front /i:/. The English close front vowel /i/, meanwhile, is very close to its Finnish counterpart, especially with regards to vowel height. The English open-mid front vowel /ɛ/ falls between the Finnish mid front /e/ and near-open front /æ/ in terms of height, and is very close to the Finnish /e/ in terms of backness. The English open-mid back /ʌ/ is closest to the Finnish near-open back /ɑ/ in both respects, while the English open-mid back /ɔ/ is also closest to the Finnish /ɑ/ in terms of height but to the Finnish mid back /o/ in terms of backness. Finally, the English near-open front /æ/ is close to the Finnish near-open front /æ/ in terms of height and the Finnish mid front /e/ in terms of backness. Overall, differences between Finnish and American English are visible especially with regards to backness, with American English vowels constantly having higher F2 values than Finnish vowels, meaning that they are more fronted.

In conclusion, a comparison of the Finnish and American English vowel systems shows both similarities and differences. Certain challenges for the Finnish learner of American English can be distinguished. Firstly, some phonemes are closely related, and thus may be confused with each other. Similarly, the AE inventory includes some phonemes that the Finnish inventory does not contain, which require the Finnish learner to form new phonetic categories or, failing that, to assimilate them to pre-existing ones. However, as shown above, because the acoustic relationships between Finnish and American English vowels can be multi-faceted, creating a new phonetic category can be difficult if two or more AE phonemes are assimilated to one Finnish phoneme.

The discussion now turns to the empirical study of how advanced Finnish learners perceive English vowels.

4. Empirical study

The present chapter introduces the empirical study reported in this paper. Section 4.1. describes the research questions and hypotheses. Sections 4.2. and 4.3. describe the stimuli and participants. Section 4.4. describes the procedures for experiments 1, 2 and 3, respectively. Section 4.5. discusses the design of the study. The chapter concludes with section 4.5., which concerns the statistical analysis methods used.

4.1. Research questions

The present study investigated the perception of non-native English vowels by advanced Finnish learners of English. The study aimed at answering the following research questions:

1. How do advanced Finnish learners of English assimilate the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?
2. How accurately do advanced Finnish learners of English identify the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?

3. How accurately do advanced Finnish learners of English discriminate the English vowel contrasts /iː/—/ɪ/, /ɛ/—/æ/, /ɪ/—/ɛ/, and /ɔː/—/ʌ/?

3a. Can discrimination be predicted by assimilation patterns and overlap scores?

3b. Are there directional asymmetries in discrimination, as proposed by NRVF?

4. Does the consonant context (bilabial versus alveolar) affect perception?

Taken together, the research questions aimed at offering an overview of different aspects of non-native vowel perception, as well as enabling the investigation of how these aspects are related.

The purpose of the first research question was to find out how accurately advanced Finnish learners identify English vowels. High accuracy levels in identification were presumed to show that the participants' mental representations of the phonemes are well-defined. Cross-language perception was investigated through the second research question, which aimed at finding out how advanced Finnish learners assimilate English vowels to Finnish vowel categories. Here, it was presumed that overlap in assimilation, that is, assimilating two or more phonemes to one Finnish category, suggests that discriminating those phonemes will be more difficult. The purpose of the third research question, then, was to investigate how advanced Finnish learners discriminate English vowel contrasts. The third research question also aimed at investigating whether there are directional asymmetries in advanced Finnish learners' discrimination of English vowels. Balas (2018) found no evidence of directional asymmetries in advanced Polish L2 learners of English, whereas Tyler et al. (2014b), investigating naïve listeners, found directional asymmetries in those vowel contrasts that were assimilated in the single-category pattern. The third research question, then, aimed at establishing whether the advanced status of learners can result in the elimination of directional asymmetries.

The fourth research question concerned the influence of context. It was not considered separately, but instead alongside the overall discussion on the identification, assimilation and discrimination experiments, respectively.

The six English monophthongs investigated (/iː ɪ ɛ æ ʌ ɔː/) were chosen so that it was possible to form vowel pairs where at least one phoneme is not included in the Finnish inventory (/ɪ ɛ ɔː ʌ/). Finnish speakers thus have to either create a new phonological category for these vowels or assimilate them to pre-existing Finnish categories. For the discrimination tasks, the vowels were divided into pairs on the basis of acoustic similarity. In order to investigate the effect of consonantal context on non-native vowel perception, the vowels were elicited in two different contexts, the bilabial /bVb/ and alveolar /dVd/. These contexts are commonly used in phonological research (e.g. Levy 2009, Balas 2018). Studying the perception of French vowels by English learners, Levy (2009) found that discrimination was more successful in bilabial context for both more and less experienced learners; Balas (2018) also found some evidence for Polish advanced learners better discriminating English vowels in bilabial context, although consonantal influence was overall relatively minor. For the present study, then, it could be hypothesised that discrimination will be mildly-to-moderately better for vowels in the bilabial /bVb/ context.

4.2. Stimuli

The stimuli were recorded in February 2019 at the University of Jyväskylä. Two native speakers of American English, a 57-year-old female from Minnesota and a 35-year-old male from Pennsylvania, were instructed to read a list of English words containing the vowels /i/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/ in symmetrical /bVb/ and /dVd/ syllables (see Appendix 1). The recordings were made using a Roland R-05 digital recorder with a built-in microphone. The stimuli had a sample rate of 44.1 kHz, 16-bit resolution, and were on a stereo channel. The files were later transferred to a MacBook Pro computer. The Audacity audio editing software was used to edit the files into three different stimulus lists. A list randomiser was used to order the tokens (see section 4.4. for description of the experiments). The Praat software was then used to measure the first and second formants of each vowel as produced by both speakers in both contexts. The remainder of this section offers a brief acoustic description of the stimuli. In particular, as the present study also investigates the influence of consonant context, the focus is on possible differences between the speakers' productions in bilabial and alveolar contexts.

Table 6 shows the average F1 and F2 values for American English monophthongs /ɪ iː ɛ ʌ ɔː æ/ in bilabial (bVb) and alveolar (dVd) contexts, as produced by a male native speaker:

Vowel	F1	F2
ɪ	428	1805
	394	1819
i:	293	2251
	620	1964
ɛ	945	2493
	522	1823
ʌ	616	1695
	507	1489
ɔ:	643	877
	600	1052
æ	728	1480
	569	1738

Table 6. Average F1 and F2 values (Hz) for American English vowels /ɪ i: ɛ ʌ ɔ: æ/ in bilabial (upper row) and alveolar (bottom row) contexts, as produced by a male native speaker of AE.

As illustrated by Table 6, formant values change depending on the consonant context in which the vowel is elicited. Here, the vowel phonemes are generally closer, that is, produced with the tongue higher in the mouth, in alveolar contexts. An exception is the phoneme /i:/, which is considerably closer in the bilabial than the alveolar context. The productions are more divided across the two contexts when vowel backness is considered. The phonemes /ɪ ɔ: æ/ all have higher F2 values and are thus more fronted in alveolar contexts, while the phonemes /i: ɛ ʌ/ are more fronted in bilabial contexts. In other words, the consonant context appears to have influenced the speaker's productions.

Table 7 presents the average F1 and F2 values for American English monophthongs /ɪ i: ɛ ʌ ɔ: æ/ in bilabial (bVb) and alveolar (dVd) contexts, as produced by a female native speaker:

	F1	F2
Vowel		
ɪ	651	2041
	469	2197
i:	613	1954
	320	2839
ɛ	671	2394
	375	2135
ʌ	400	1179
	745	1338
ɔ:	775	1250
	437	1696
æ	769	1790
	823	1769

Table 7. Average F1 and F2 values (Hz) for American English vowels /ɪ i: ɛ ʌ ɔ: æ/ in bilabial (upper row) and alveolar (bottom row) contexts, as produced by a female native speaker of AE.

Again, the vowels are generally more closed in alveolar contexts. Exceptions are the phonemes /ʌ æ/, which have lower F1 values in the bilabial context. With regards to vowel backness, only /ɛ/ has a higher F2 value, and is thus more fronted, in the bilabial context, with all the other phonemes being more fronted in the alveolar context. Context effects are thus distinguishable also in the female speaker's productions.

A comparison of the male and female speaker's productions reveals that her /ɪ æ/ are higher than the male speaker's in both contexts, her /i:/ is higher in the bilabial context, and her /ʌ ɔ:/ are higher in the alveolar context. Her /ɛ/, meanwhile, is lower in both bilabial and alveolar contexts. Furthermore, her vowels are more often fronted in the alveolar context than the male speaker's, and are generally more fronted than his; only her bilabial /i: ɛ/ and bilabial and alveolar /ʌ/ are less fronted than his.

Overall, the productions of the current study's male speaker are close to the productions measured in Hillenbrand et al. (1995), discussed in section 3.3. Looking at vowel height, his /ɪ ɔ:/ are near in both bilabial and alveolar contexts, with the bilabial /ɪ/ being almost identical (428 Hz vs. 427 Hz). His /i: ʌ/ are near in the bilabial and /æ/ in the alveolar context. With regards to vowel backness, his /ɪ/ is less fronted in both contexts. His /iː/ is nearer in the the bilabial and /ɛ ʌ æ/ in the alveolar context. Finally, his /ɔː/ is rather close in both contexts.

The female speaker's productions are also close to Hillenbrand et al. (1995). Considering vowel height, her productions appear to be equally divided between the contexts, with /ɪ iː ʌ/ near in the alveolar and /ɛ ɔ æ/ in the bilabial one. Regarding vowel backness, her /ɪ æ/, as well as her bilabial /iː/, are less fronted than corresponding productions in Hillenbrand et al. (1995); her /ɛ/ is near in both contexts, and her /ʌ/ is closer in the alveolar and /ɔː/ in the bilabial context.

In conclusion, context appears to have affected the production of both male and female speakers, although the effect is overall relatively minor. Differences can also be distinguished between the male and female speaker. However, as advanced learners, the participants have presumably been exposed to enough allophonic variation that the minor differences are not expected to cause differences in perception, and are not consequently discussed in detail in the present study.

4.3. Participants

The participants of the study (N=20, 4 male, median age=21.75, age range=19-28) were English majors (N=11) or minors (N=9) at a Finnish university. At the time of data collection, the participants were taking an introductory phonetics and phonology course. Participation in the study was voluntary and non-compensated.

Prior to the actual experiments, the participants filled in a background questionnaire. No participant reported any diagnosed hearing problems. They had studied English for 13.47 years on average (range=10-28 years). With the exception of two participants, who had spent an extended period of time (defined as at least three months) in the United States and Australia, respectively, no other participant had spent extended time in an English-speaking country. One participant had an English-speaking parent (from Australia). Five participants reported frequent spoken interaction with native speakers of English apart from their instructors, averaging 2.6 hours per week (range=1-5 hours).

4.4. Procedure

The study consisted of three listening experiments. The first group of participants (N=9) participated in the study in March 2019, and the second group (N=11) in April 2019. The experiments were piloted in March 2019 in a separate session, with the first group participating. Following the pilot session, a goodness-of-rating section was removed from the second experiment, as it was not considered to answer the research questions. No other major changes were made.

4.4.1. Experiment 1: Sound discrimination

Experiment 1, a discrimination task, took place in a language laboratory at a Finnish university (see Appendix 2). Through a headset, the participants heard two English words and were asked to determine whether the vowel in the second word was the same as in the first word or not by circling their answer on an answer sheet. The interstimulus interval (=ISI) was 2 seconds, and the intertrial interval was 5 seconds. For each contrast, there were four trials (AB, BA, BB, and AA), with 50% of the trials being change trials and 50% being no-change (catch) trials (for example, /biːb/—/bɪb/, /bɪb/—/biːb/, /biːb/—/biːb/, /bɪb/—/bɪb/). Each vowel was presented in both the bilabial and the alveolar contexts, with no mixing (so */biːb/—/dɪd/). The inclusion of both AB and BA trials allowed for the investigation of possible directional asymmetries. The order of presentation was randomised. For each trial, both tokens were produced by a different speaker, with the order of speakers randomised. There were 4 contrasts × 4 orders × 2 contexts = 32 responses per participant, for a total of 640 responses.

4.4.2. Experiment 2: Sound identification

Experiment 2, a forced-choice identification task, took place immediately after experiment 1 (see Appendix 3). Prior to the experiment, the researcher presented the IPA symbols used, and the participants were allowed to ask for clarification if necessary. The IPA symbols were also written on the answer sheet.

Through a headset, the participants heard an English word and were asked to identify its vowel by circling their answer on an answer sheet. The intertrial interval was 5 seconds. Each vowel was presented in both the bilabial and the alveolar contexts, with the order randomised. Half of the stimuli (4 words) were spoken by the male speaker and half by the female speaker, with the order randomised. There were 6 vowels × 2 contexts = 12 responses per participant, for a total of 240 responses.

4.4.3. Experiment 3: Sound assimilation

Experiment 3, a perceptual assimilation task, took place immediately after experiment 2 (see Appendix 4). Prior to the experiment, the researcher again presented the IPA symbols used, and the participants were allowed to ask for clarification if necessary.

In the first part, through a headset, the participants heard an English word and were asked to choose which Finnish vowel its vowel resembled the most by circling their answer on an answer sheet. The intertrial interval was 5 seconds. Each vowel was presented in both the bilabial and the alveolar contexts, with the order randomised. Half of the stimuli (4 words) were spoken by the female speaker and half by the male speaker, with the order randomised. There were 6 vowels × 2 contexts = 12 responses per participant, for a total of 240 responses.

The second part followed immediately after the completion of the first part. The participants heard the same words again, in the same order and spoken by the same speaker. They were asked to rate the resemblance between the English and Finnish vowels using a five-point scale (1=”very remotely”, 5=”identical”). There were 6 vowels × 2 contexts = 12 responses per participant, for a total of 240 responses.

4.5. Methodology

The present study was designed partially after Balas (2018), who investigated the assimilation and discrimination of English vowels by advanced Polish learners. However, some methodological differences can be seen between the two studies.

Firstly, two different speakers were utilised in the present study, while Balas (2018) only had one speaker. The purpose of this was to prevent the participants from making judgements solely on the basis of the acoustic features of the phonemes (Strange and Shafer 2008: 162). In other words, using only one speaker makes it possible for the participants to judge (dis)similarity on the basis of whether the tokens are physically identical or not, unless each token is recorded separately. Thus, in the present study, two speakers were used to negate this possibility.

Secondly, the present study also included a forced-choice identification task. The purpose of this task, as discussed above, was to determine if the participants could distinguish the phonemes when presented individually. The inclusion of an identification task also made it possible to see if identification results could be used to explain assimilation and discrimination results, although this was not a separate research question.

Thirdly, the present study utilised a slightly different discrimination task design. Balas (2018) used the AXB paradigm, where the participants were presented with three tokens per each contrast, with one token differing from the other two. For the purposes of the present study, an AX paradigm was used, with only two tokens presented per each contrast. This was deemed easier for the participants, while still allowing the investigation of both discrimination and potential directional asymmetries.

Finally, following Balas (2018), the present study used /cVc/ syllables in bilabial and alveolar contexts. Separate contexts enabled the investigation of context effects. Balas (2018) also elicited the vowels in the velar /gVg/ context; it was not included in the present study because of limited resources. Closed syllables allowed the inclusion of lax vowels (Balas 2018) and were also deemed more natural than isolated and/or synthetic stimuli.

In all experiments, IPA symbols were chosen instead of orthographic labels to ensure that the participants' mental representations of the vowels under investigation were constant. Prior to the experiments, the participants were familiarised with the symbols to ensure that they could correctly link them to their respective phonemes.

4.6. Statistical analysis

The data were analysed statistically using the SPSS program. Mean correct identification scores were calculated by adding all correct responses and dividing them by the total number of responses. Assimilation scores were calculated similarly by adding all responses where a specific Finnish vowel category had been chosen and dividing them by the total number of responses. Mean goodness-of-fit ratings in the assimilation task were calculated by adding all ratings for one assimilation together and then dividing them by their number. Finally, discrimination scores were calculated in the same manner as mean correct identification scores. For all three experiments, scores were calculated first for both contexts together and then separately for the bilabial and alveolar contexts.

For the assimilation task, an English vowel contrast was classified as Two Category if both sounds were assimilated to two different Finnish categories; Single Category if both sounds were assimilated to one Finnish category with the same goodness-of-rating; and Category Goodness if both sounds were assimilated to one Finnish category with different goodness-of-fit ratings. An overlap score (Levy 2009) was also calculated for each vowel contrast in the assimilation task, defined as the “smaller percentage of responses when two members of a pair of non-native (or L2) speech sounds are assimilated to the same native category”. To demonstrate, if the non-native vowel A is assimilated to the native vowel category C 70% of the time and the non-native vowel B 40% of the time, then A and B perceptually overlap for 40% of the time. A full overlap score can be calculated by combining individual overlap scores for all assimilation categories: if A if also assimilated to the native D 30% of the time and E 60% of the time, the full overlap score is 40% + 30% = 70%.

5. Perception of English vowels by advanced Finnish learners

The current section presents the results of the study reported in this paper. In order to consider the influence of assimilation patterns on discrimination, the results of the experiments are presented in a mixed order, beginning with sound identification (experiment 2) and ending with sound discrimination (experiment 1).

5.1. Sound identification

Sound identification was investigated in the second experiment, which was motivated by the following research question:

2. How accurately do advanced Finnish learners of English identify the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?

Table 8 presents the mean overall identification of English vowels.

	English vowel					
English vowel	iː	ɪ	ɛ	æ	ʌ	ɔː
iː	**97.5%**	2.5%	-	-	-	-
ɪ	5%	**85%**	10%	-	-	-
ɛ	-	-	**100%**	-	-	-
æ	-	-	-	**97.5%**	2.5%	-
ʌ	-	-	-	-	**72.5%**	27.5%
ɔː	-	-	-	-	5%	**95%**

Table 8. Mean overall identification of English vowels (%). Modal responses **bolded**.

As can be seen from Table 8, perfect (100%) identification was reached for /ɛ/, excellent (>90%) for /iː æ ɔː/ and good (>80%) for /ɪ/; only for /ʌ/ was discrimination overall poor (>70%). In other words, Finnish learners generally identified English vowels very well, with /ʌ/ as the clearly most difficult phoneme to identify. Each vowel will now be discussed in more detail.

/iː/ was identified accurately at a rate of 97.5%, suggesting that Finnish learners had very little trouble with it. In 2.5% of the cases, it was misidentified as /ɪ/, possibly because the two sounds are acoustically very close.

/ɪ/ was also identified correctly at a high rate (85%). It was the only vowel identified as three separate English vowels (/iː/, /ɪ/ and /ɛ/), which indicates that it had fuzzier category boundaries than other vowels. In 10% of the cases, it was misidentified as /ɛ/. Acoustically, /ɪ/ and /ɛ/ are further apart than /ɪ/ and /iː/; however, one explanation for the frequency of the /ɛ/ misidentification could be that, unlike /ɪ/ and /iː/, /ɪ/ and /ɛ/ are both lax vowels, and they are also closer in duration. These additional similarities between /ɪ/ and /ɛ/ could have resulted in the participants misidentifying /ɪ/ as /ɛ/ more frequently than as /iː/, which was chosen in 5% of the cases.

As noted above, /ɛ/ was identified with 100% accuracy, suggesting that while /ɪ/ was misidentified as /ɛ/, this did not happen the other way around. A possible explanation could simply be that the participants' mental representations of /ɛ/ were clearer than their representations of /ɪ/, maybe because /ɪ/ and /iː/ both have to fit to the native /i/. In other words, the participants may have been more likely to give /ɛ/-like attributes to /ɪ/.

/æ/ was identified accurately in 97.5% of the cases, suggesting that, as with /iː/, Finnish learners had minimal trouble with correctly recognising it. Acoustically, the English /æ/ and the Finnish /æ/ are very close; it is thus possible that the participants had already formed a strong mental representation of /æ/ through exposure to the vowel in their native language, and could easily identify the English phoneme as an instance of it. Interestingly, in 2.5% of the cases, /æ/ was misidentified as /ʌ/. Although the two vowels are relatively close in height, /æ/ is a front vowel and /ʌ/ a back vowel. The very small number of these identifications, however, suggests that this misidentification is not a pattern typical of Finnish learners, and could consequently be the result of an individual perception error.

/ʌ/, as noted above, was the only vowel with an overall poor discrimination accuracy; it was identified correctly in just 72.5% of the cases. Participants appeared to misidentify it as /ɔː/ relatively frequently, in 27.5% of the cases. However, this does not hold true the other way around: /ɔː/ was identified correctly with 95% accuracy, and was mistaken for /ʌ/ only 5% of the time. As neither sound is included in the Finnish inventory, Finnish listeners do not have corresponding native representations, as is the case with /æ/, suggesting that they either need to form new phonetic categories for the non-native vowels or try to assimilate them to pre-existing Finnish categories (see next section for discussion on assimilation results). Although /ʌ/ and /ɔː/ differ in roundedness and tenseness, they are nonetheless acoustically close, which could make it more difficult to distinguish and identify them. The overall high identification accuracy for /ɔː/ suggests that the participants may have clearer mental representations for this phoneme than for /ʌ/, which could explain the misidentification discrepancy.

In order to investigate if the consonant context was statistically significant, two-sample t-tests were performed. (No two-sample t-test could be performed for /ɛ/, which was identified with 100% accuracy in both contexts.) The results showed that consonant context did not reach statistical significance for any of the vowels: /iː/ $t(19)=-1.000$, $p=.330$; /ɪ/ $t(19)=.000$, $p=1.000$, /æ/ $t(19)=1.000$, $p=.330$, /ʌ/ $t(19)=-1.143$, $p=.267$, and /ɔː/ $t(19)=1.453$, $p=.163$. Consequently, is possible that the participants' mental representations of the English vowels are relatively context-independent, allowing them to identify the phonemes with generally high success levels in different consonant contexts. Indeed, Levy and Strange (2008) found that context influenced the perception of inexperienced but not experienced listeners, suggesting that even if advanced listeners do not reach native-like performance, some aspects of their perception could have reached native-like levels.

This does not mean, however, that no differences could be distinguished between bilabial and alveolar contexts. Table 9 presents the mean identification scores of English vowels in the two different environments.

		English vowel					
English vowel	Context	i:	ɪ	ɛ	æ	ʌ	ɔ:
i:	Bilabial	**95%**	5%	-	-	-	-
	Alveolar	**100%**	-	-	-	-	-
ɪ	Bilabial	5%	**85%**	10%	-	-	-
	Alveolar	5%	**85%**	10%	-	-	-
ɛ	Bilabial	-	-	**100%**	-	-	-
	Alveolar	-	-	**100%**	-	-	-
æ	Bilabial	-	-	-	**100%**	-	-
	Alveolar	-	-	-	**95%**	5%	-
ʌ	Bilabial	-	-	-	-	**65%**	35%
	Alveolar	-	-	-	-	**80%**	20%
ɔ:	Bilabial	-	-	-	-	-	**100%**
	Alveolar	-	-	-	-	10%	**90%**

Table 9. Mean identification of English vowels as separated by context. Modal responses **bolded**.

As illustrated by Table 9, context effect does not appear to be very strong, with none of the vowels displaying an effect of >15%. Given that statistical significance was not reached for any of the vowels, this is as expected.

Above, it was found that the participants identified the English /i:/ as both /i:/ and /ɪ/. Here, it can be seen that the misidentification occurred only in the bilabial context. At a rate of just 5%, it could be attributed to an individual perception error.

/ɪ/ showed no context effect at all, with 85% correct identification, 10% misidentification as /ɛ/ and 5% misidentification as /i:/ occurring in both the bilabial and the alveolar context. Thus, even though Finnish learners were the most divided in their identification of this vowel, this division did not appear to be context-dependent.

/æ/ showed an effect comparable to that shown by /i:/, albeit in reverse: it was identified with 100% accuracy in the bilabial context, whereas the misidentification as /ʌ/ occurred only in the alveolar context. As with /i:/, the 5% rate means that this mistake was not frequent among Finnish learners, and consequently may have more likely been an individual error than the result of for example wide-scale acoustic similarity.

The largest context effect was visible for /ʌ/, which was identified with 65% accuracy in the bilabial context and with 80% accuracy in the alveolar context. This discrepancy could potentially be explained by coarticulation. As noted by Levy and Strange (2008), in order to achieve high-level performance, learners must learn not only the static acoustic properties of individual vowels but also the non-native language's coarticulatory patterns. The bilabial /b/ is produced with both lips, and could affect the unrounded /ʌ/, making it more rounded and thus causing it to sound more similar the rounded /ɔ:/. In contrast, the alveolar /d/ is unlikely to have a similar effect.

Finally, /ɔ:/ displayed a smaller context effect than /ʌ/. The vowel was identified with 100% accuracy in the bilabial context and with 90% accuracy in the alveolar context, where it was misidentified as /ʌ/ in 10% of the cases. Although the effect is still small enough that the alveolar misidentification could potentially be explained by individual errors, another possibility is that, as /ɔ:/ is a rounded vowel, its identification was facilitated by the bilabial context more than the alveolar one.

In conclusion, vowel identification ranged from perfect to poor, with only /ʌ/ identified with an overall rate of <80%. /ɪ/ was the only vowel that was identified as three different vowels, suggesting that it was the most divisive phoneme to identify, possibly because it bears higher resemblance to several other phonemes. Consonant influence was distinguishable but failed to reach statistical significance, potentially because of the small sample size and potentially because of the advanced status of the participants. /ʌ/ showed also the largest context effect, suggesting that the participants' mental representation of it is possibly less abstract than their representations of the other vowels.

5.2. Sound assimilation

Sound assimilation was investigated in the third experiment, which was motivated by the following research question:

> **1.** How do advanced Finnish learners of English assimilate the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?

Table 10 presents the overall results of the assimilation task.

	Finnish vowel							
English vowel	ɑ	e	i	o	u	y	ø	æ
iː	-	5% (2.50)	**95%** (4.18)	-	-	-	-	-
ɪ	-	10% (2.75)	**85%** (3.41)	-	-	5% (2)	-	-
ɛ	-	**100%** (3.95)	-	-	-	-	-	-
æ	2.5% (1)	-	-	-	-	-	2.5% (1)	**95%** (4.58)
ʌ	**57.5%** (3.22)	-	-	27.5% (3)	12.5% (2.60)	-	-	-
ɔː	42.5% (3.47)	-	-	**55%** (3.95)	2.5% (2)	-	-	-

Table 10. Assimilation of English vowel stimuli to Finnish vowel categories and mean goodness-of-fit ratings (in parentheses). Modal responses **bolded**.

As can be seen from Table 10, most of the English vowels were assimilated to one Finnish category at a high rate (>85%). Similar results were reached by Balas (2018), who found that Polish learners generally assimilated English vowels to one Polish category at a rate of >80%. This suggests that Finnish learners found it easy to classify these vowels in terms of Finnish vowel categories. The exceptions were /ʌ/ and /ɔː/, both of which were assimilated to one Finnish category at a rate barely over 50%. Above, it was proposed that participants' mental representations of these vowels may be weaker, making it more difficult not only to identify them but also to assimilate them.

Both the English /iː/ and /ɪ/ were assimilated to the same Finnish category, /i/, at high frequencies (95% and 85%, respectively). However, a clear difference could be distinguished between the mean goodness-of-fit ratings: whereas /iː/ was considered very similar to the Finnish /i/, receiving a rating of 4.18 out of 5, /ɪ/ was considered only moderately similar to /i/ with a rating of 3.41. This is not surprising, given that /iː/ is acoustically closer to the Finnish /i/ than /ɪ/, and that the learners rarely misidentified /iː/ and /ɪ/ as each other (see section 5.1.).

Although /i/ was by far the most frequent Finnish category chosen for the English /iː/ and /ɪ/, both were also assimilated to the Finnish /e/, the former at a rate of 5% and the latter at a rate of 10%. Goodness-of-fit ratings reveal that those participants who assimilated /iː/ and /ɪ/ to /e/ found them only remotely similar to the Finnish vowel, rating /iː/ as 2.50 and /ɪ/ as 2.75.
The /ɪ/ to /e/ assimilation is potentially explained by the fact that /ɪ/ was misidentified as /ɛ/ in 10% of cases, and that /ɛ/ was assimilated to the Finnish /e/ in 100% of cases. Finally, /ɪ/ was also assimilated to the Finnish /y/ 5% of the time, with a mean goodness-of-fit rating of 2, again suggesting that it was not considered similar to the Finnish vowel. This was the only assimilation to the Finnish /y/, which indicates that overall, participants did not find /y/ similar to any of the English vowels investigated.

The English /ɛ/ was the only vowel that was assimilated to the same Finnish category (/e/) 100% of the time. It received a mean goodness-of-fit rating of 3.95, suggesting that the listeners considered it moderately-to-very similar to the Finnish vowel. As participants also identified /ɛ/ with 100% accuracy (see section 5.1.), it appears that in general, this phoneme caused them little trouble. A potential explanation was briefly discussed above in relation to /æ/: like /æ/, /ɛ/ could be identified through, and assimilated to, the corresponding Finnish category, in this case /e/. Another possible factor could be orthography, whose role in the perception of non-native sounds was discussed in section 2.3. In the present study, when the vowels were introduced to the participants prior to the actual experiments, /ɛ/ was exemplified by the word BET. Indeed, the corresponding grapheme for /ɛ/ is often <e>. Thus, the relatively strong (for English) grapheme-phoneme-mapping between /ɛ/ and <e> could also have influenced the participants' perception.

The English /æ/ was assimilated to the Finnish /æ/ 95% of the time, with a very high mean goodness-of-rating (4.58), suggesting that it was considered very similar to its Finnish counterpart. Indeed, as discussed in section 3.3., the two vowels are acoustically very close, especially in terms of height, although the English /æ/ is more fronted than the Finnish /æ/. /æ/ was also assimilated to the Finnish /ɑ/ and /ø/, in both cases at a rate of 2.5% and with a mean goodness-of-fit ratings of 4 and 1, respectively. In other words, these assimilations were exceedingly rare. Similarity to /ɑ/ was considered very high, whereas similarity to /ø/ was considered very low. Because of the rarity of these assimilations, they are likely to be the result of individual differences in perception. Similar rare assimilations were also found by Balas (2018), although she did not discuss potential explanations; nevertheless, their rarity could be taken to mean that overall, advanced learners assimilate L2 sounds in a largely uniform manner.

As noted above, assimilation for /ʌ/ and /ɔː/ was much more divided than for other vowels, suggesting that the learners found it difficult to classify them according to Finnish categories. /ʌ/ was assimilated to the Finnish /ɑ/ at a rate of 57.5%, with a mean goodness-of-fit rating of 3.22. The second highest assimilation was to the Finnish /o/ at a rate of 27.5%, with a mean rating of 3. The Finnish /u/ was chosen the least often, at a rate of 12.5%; with a mean rating of 2.6, it was also considered to bear the least similarity to /ʌ/. The small differences between the goodness-of-fit ratings also shows the difficulty the participants had with assimilating /ʌ/, as it was not considered very similar to any of the Finnish vowels chosen. The English /ɔː/ was assimilated to the Finnish /o/ at a rate of 55% and with a mean goodness-of-fit rating of 3.95. It was also assimilated to /ɑ/ at a rate of 42.5% with a mean rating of 3.47, and /u/ at a rate of 2.5% with a mean rating of 2.

The converging of Finnish categories chosen suggests that participants found /ʌ/ and /ɔː/ similar to each other. This is also supported by the fact that /ʌ/ was misidentified as /ɔː/ in 27.5% of the cases. /ɔː/, however, was only misidentified as /ʌ/ in 5% of the cases, suggesting that the participants have a better grasp of it. Above, coarticulation was discussed as a possible explanation for this; more specifically, it was proposed that in the bilabial context, the acoustic properties of /ʌ/ may have become closer to those of /ɔː/, resulting in more frequent misidentification. Because further rounding of /ɔː/ does not lead to increased acoustic resemblance to the unrounded /ʌ/, this effect would be more noticeable precisely for /ʌ/. The same explanation could explain the assimilation results.

Tables 11 and 12 illustrate the assimilation of English vowels to Finnish vowel categories in bilabial and alveolar contexts, respectively.

	Finnish vowel							
English vowel	ɑ	e	i	o	u	y	ø	æ
iː	-	-	**100%** (4.15)	-	-	-	-	-
ɪ	-	15% (2.33)	**80%** (3.31)	-	-	5% (2)	-	-
ɛ	-	**100%** (3.80)	-	-	-	-	-	-
æ	-	-	-	-	-	-	5% (1)	**95%** (4.55)
ʌ	35% (3.14)	-	-	**50%** (3.20)	15% (2.67)	-	-	-
ɔː	-	-	-	**100%** (4.05)	-	-	-	-

Table 11. Assimilation of English vowel stimuli in bilabial context to Finnish vowel categories and mean goodness-of-fit ratings. Modal responses **bolded**.

	Finnish vowel							
English vowel	ɑ	e	i	o	u	y	ø	æ
iː	-	10% (2.50)	**90%** (4.22)	-	-	-	-	-
ɪ	-	5% (4)	**90%** (3.50)	-	-	5% (2)	-	-
ɛ	-	**100%** (4.10)	-	-	-	-	-	-
æ	5% (4)	-	-	-	-	-	-	**95%** (4.60)
ʌ	**80%** (3.25)	-	-	5% (1)	10% (2.50)	-	-	5% (3)
ɔː	**85%** (3.47)	-	-	10% (3)	5% (2)	-	-	-

Table 12. Assimilation of English vowel stimuli in alveolar context to Finnish vowel categories and mean goodness-of-fit ratings. Modal responses **bolded**.

As can be seen from Tables 11 and 12, the only English vowel that shows no influence of context is /ɛ/, which was assimilated to the Finnish /e/ at a 100% rate in both the bilabial and the alveolar context. Other vowels show context effect to varying degrees, with /ʌ/ again showing the strongest effect; assimilation of /ɔː/ also appears to be more context-dependent than identification. The investigation of Polish learners by Balas (2018) found similar results, suggesting that consonant context influences vowel assimilation more than vowel identification.

The English /iː/, assimilated 100% to the Finnish /i/ in the bilabial context, was also assimilated to the Finnish /e/ in the alveolar context, albeit at a rate of just 10%. Difference between the mean goodness-of-fit ratings for the /iː/ → /i/ assimilation was minimal (4.15 in the bilabial, 4.22 in the alveolar context), further supporting the suggestion that context did not have a major effect.

More influence can be seen in the case of the English /ɪ/. The vowel was slightly more likely to be assimilated to the Finnish /i/ in the alveolar (90%) than the bilabial (80%) context. The mean goodness-of-fit ratings again differed only very little (3.31 in the bilabial, 3.50 in the alveolar context), suggesting that context had only a minor effect on how similar to Finnish categories the vowels were considered. Conversely, /ɪ/ was slightly more frequently assimilated to the Finnish /e/ in the bilabial (15%) than the alveolar (5%) context. Interestingly, mean goodness-of-fit ratings showed that it was considered more similar to the Finnish /e/ in the alveolar (4) than the bilabial (2.33) context. Previously, it was shown that /ɪ/ was misidentified as /ɛ/ in 10% of the cases in both contexts; given the strong connection between /ɛ/ and the Finnish /e/ the participants appeared to have, it is possible that this perceptual link resulted in their assimilating /ɪ/ to /e/.

The English /æ/ shows a likewise small context effect. In the alveolar context, it was assimilated to the Finnish /ɑ/ in 5% of the cases, with a mean goodness-of-fit rating of 4; in the bilabial context, assimilation to the Finnish /æ/ was 100%. Again, a possible explanation for the unusual assimilation could be identification: the English /æ/ was misidentified as /ʌ/ in the alveolar context in 5% of the cases, and /ʌ/ was strongly assimilated to the Finnish /ɑ/ in the alveolar context.

A considerably stronger context effect is visible for the English /ʌ/. In the bilabial context, it was assimilated most commonly to the Finnish /o/, although only at a 50% rate. In the alveolar context, on the other hand, it was assimilated to the Finnish /ɑ/ at a 80% rate. In the bilabial context, /ɑ/ was the second most common Finnish vowel chosen at 35%; in the alveolar context, however, /o/ was chosen just 5% of the time. In other words, in the bilabial context, categorisation between /ɑ/ and /o/ was less sharp than in the alveolar context.

Finally, the English /ɔː/ also shows a strong context effect. In the bilabial context, it was assimilated to the Finnish /o/ in 100% of the cases, and received a mean goodness-of-fit rating of 4.05, suggesting that it was considered very similar to /o/. Conversely, in the alveolar context, it was assimilated to the Finnish /ɑ/ at a rate of 85%, while the Finnish /o/ was only chosen in 10% of the cases.

In other words, /ʌ/ and /ɔː/ followed a similar pattern, where in the bilabial context they were assimilated more frequently to /o/, and in the alveolar context to /ɑ/. Coarticulation, as discussed above, appears to be a potential solution for this phenomenon. The alveolar /d/ is unlikely to have a similar rounding effect as the bilabial /b/, which could explain the preference for /ɑ/ in the alveolar context. Indeed, acoustically the alveolar /ʌ/ and /ɔː/ are closer to the Finnish /ɑ/ than /o/.

The discussion now turns to assimilation patterns. As stated in section 2.4.1., PAM-L2 proposes that the assimilation of non-native vowel contrasts to native categories can be divided into six patterns (Best and Tyler 2007). In order to distinguish possible variation between participants, assimilation patterns are presented here separately for each individual; detailed individual analysis, however, is beyond the scope of the current study.

Tables 13 (bilabial context) and 15 (alveolar context) present individual assimilation patterns for the English vowel contrasts /iː/—/ɪ/, /ɪ/—/ɛ/, /ɛ/—/æ/ and /ɔː/—/ʌ/.

	English vowel contrast			
Participant	/iː/—/ɪ/	/ɪ/—/ɛ/	/ɛ/—/æ/	/ɔː/—/ʌ/
1	SC	TC	TC	CG
2	SC	TC	TC	CG
3	CG	TC	TC	TC
4	SC	TC	TC	TC
5	CG	TC	TC	CG
6	SC	TC	TC	SC
7	CG	TC	TC	SC
8	SC	TC	TC	TC
9	CG	TC	TC	SC
10	CG	TC	TC	TC
11	TC	TC	TC	TC
12	TC	CG	TC	TC
13	CG	TC	TC	CG
14	CG	TC	TC	CG
15	SC	TC	TC	TC
16	CG	TC	TC	TC
17	CG	TC	TC	CG
18	CG	TC	TC	TC
19	TC	CG	TC	TC
20	TC	CG	TC	CG

Table 13. Individual assimilation patterns for English vowel contrasts in bilabial context.

Overall frequencies of assimilation patterns per contrast are presented in Tables 14 and 16 for bilabial and alveolar contexts, respectively.

	English vowel contrast			
Pattern	/iː/—/ɪ/	/ɪ/—/ɛ/	/ɛ/—/æ/	/ɔː/—/ʌ/
SC	30%	-	-	15%
TC	20%	**85%**	**100%**	**50%**
CG	**50%**	15%	-	35%

Table 14. Frequency of assimilation pattern types in bilabial context. Most common pattern bolded.

As illustrated by Tables 13 and 14, three patterns (Single Category, Two Category, and Category Goodness) could be observed in the bilabial data. None of the four contrasts were predominantly assimilated in Single Category pattern. /ɪ/—/ɛ/ (85%), /ɛ/—/æ/ (100%) and /ɔ:/—/ʌ/ (50%) were all assimilated primarily in Two Category pattern, although the rate for /ɔ:/—/ʌ/ was relatively low. Finally, /i:/—/ɪ/ (50%) was assimilated primarily in Category Goodness pattern, although again at a relatively low rate. /i:/—/ɪ/ and /ɔ:/—/ʌ/ thus appear to be the most divisive contrasts to assimilate, with all three patterns present. This is supported by the general assimilation results, as discussed above.

	English vowel contrast			
Participant	/i:/—/ɪ/	/ɪ/—/ɛ/	/ɛ/—/æ/	/ɔ:/—/ʌ/
1	CG	TC	TC	CG
2	CG	TC	TC	CG
3	TC	TC	TC	CG
4	SC	TC	TC	TC
5	CG	TC	TC	CG
6	SC	TC	TC	TC
7	SC	TC	TC	SC
8	SC	TC	TC	TC
9	CG	TC	TC	CG
10	SC	TC	TC	SC
11	TC	TC	TC	CG
12	TC	SC	TC	SC
13	CG	TC	TC	CG
14	CG	TC	SC	CG
15	TC	TC	TC	CG
16	CG	TC	TC	CG
17	SC	TC	TC	SC
18	CG	TC	TC	TC
19	CG	TC	TC	SC
20	CG	TC	TC	TC

Table 15. Individual assimilation patterns for English vowel contrasts in alveolar context.

	English vowel contrast			
Pattern	/iː/—/ɪ/	/ɪ/—/ɛ/	/ɛ/—/æ/	/ɔː/—/ʌ/
SC	30%	5%	-	25%
TC	20%	**95%**	**100%**	25%
CG	**50%**	-	-	**50%**

Table 16. Frequency of assimilation pattern types in alveolar context. Most common pattern **bolded**.

The same patterns (Single Category, Two Category, and Category Goodness) were also observed in the alveolar context, as illustrated by Tables 15 and 16. Again, no contrast was most frequently assimilated in Single Category pattern. /ɪ/—/ɛ/ and /ɛ/—/æ/ were assimilated in Two Category pattern most of the time (95% and 100%, respectively). /iː/—/ɪ/ (50%) and /ɔː/—/ʌ/ (50%) were assimilated primarily in Category Goodness pattern, although at relatively low rates. /iː/—/ɪ/ and /ɔː/—/ʌ/ again showed the most spread-out patterns, which could indicate that once it reaches a certain level, difficulty assimilating a phoneme extends to different contexts.

Overlap scores can further illustrate perceptual relationships. Recall that the "smaller percentage of responses when two members of a pair of non-native (or L2) speech sounds are assimilated to the same native category" forms the overlap score of a non-native contrast (Levy 2009). Overlap scores are thus a way of measuring how similar to one native category two non-native phonemes are judged to be. The higher the overlap score, the more difficult discrimination is predicted to be (Levy 2009).

Table 17 presents the full overlap scores for each English vowel contrast in both bilabial and alveolar contexts.

		Overlap score (%)
English vowel contrast	Context	
/iː/—/ɪ/	Bilabial	80
	Alveolar	95
/ɪ/—/ɛ/	Bilabial	15
	Alveolar	5
/ɛ/—/æ/	Bilabial	0
	Alveolar	0
/ɔː/—/ʌ/	Bilabial	50
	Alveolar	90

Table 17. Overlap scores for the English vowels in bilabial and alveolar contexts.

As Table 17 illustrates, overlap scores are very high (>80%) for the /iː/—/ɪ/ contrast in both contexts, and for the /ɔː/—/ʌ/ contrast in alveolar context. /ɔː/—/ʌ/ also overlaps at 50% in the bilabial context. Overlap for /ɪ/—/ɛ/ is small (<15%) in both contexts, and /ɛ/—/æ/ do not overlap in either context, suggesting that they are not perceived similarly by advanced Finnish learners; indeed, these vowels were also not identified as each other, which indicates that their mental representations are separate.

On the basis of assimilation patterns and overlap scores, it is possible to predict how successfully non-native sounds are discriminated. Recall that PAM-L2 proposes that SC assimilations are more difficult to discriminate than CG assimilations, while TC assimilations are easy to discriminate (Best and Tyler 2007). Similarly, contrasts with high overlap scores are more difficult to discriminate than those with low overlap scores (Levy 2009). Table 18 combines the assimilation patterns and overlap scores of English vowel contrasts in both contexts to predict their discrimination difficulty.

English vowel contrast	Context	Assimilation pattern	Overlap score (%)	Difficulty
/iː/—/ɪ/	Bilabial	CG	80	Moderate/difficult
	Alveolar	CG	95	Moderate/difficult
/ɪ/—/ɛ/	Bilabial	TC	15	Easy
	Alveolar	TC	5	Easy
/ɛ/—/æ/	Bilabial	TC	0	Easy
	Alveolar	TC	0	Easy
/ɔː/—/ʌ/	Bilabial	CG	50	Moderate
	Alveolar	CG	90	Moderate/difficult

Table 18. Predicted discrimination difficulty of English vowel contrasts in both contexts according to assimilation pattern and overlap score.

On the basis of Table 18, it is thus possible to predict that /ɪ/—/ɛ/ and /ɛ/—/æ/ should be easy to discriminate in both contexts. /iː/—/ɪ/ should be moderate-to-difficult to discriminate in both contexts, and /ɔː/—/ʌ/ in the alveolar context. In the bilabial context, /ɔː/—/ʌ/ should be discriminated with moderate success.

In conclusion, with the exception of /ʌ/ and /ɔː/, English vowels were assimilated to one Finnish vowel category at a high rate (>85%). /ʌ/ and /ɔː/ also displayed the strongest context effects. Three assimilation patterns (Single Category, Two Category and Category Goodness) proposed by PAM-L2 could be distinguished, again with only minor differences in their distribution between the bilabial and alveolar contexts. Taken together, assimilation patterns and overlap scores predict that discrimination of English vowel contrasts will vary between easy and moderate-to-difficult, and will be context-dependent.

5.3. Sound discrimination

Sound discrimination was investigated in the first experiment, which was motivated by the following research question:

> **3.** How accurately do advanced Finnish learners of English discriminate the English vowel contrasts /iː/—/ɪ/, /ɛ/—/æ/, /ɪ/—/ɛ/, and /ɔː/—/ʌ/?

3a. Can discrimination be predicted by assimilation patterns and overlap scores?

3b. Are there directional asymmetries in discrimination, as proposed by NRVF?

Table 19 presents the mean overall discrimination results of English vowel contrasts.

	Mean correct discrimination
English vowel contrast	
/iː/—/ɪ/	82.5%
/ɪ/—/ɛ/	70%
/ɛ/—/æ/	98.75%
/ɔː/—/ʌ/	55%

Table 19. Mean overall discrimination of English vowel contrasts.

Above, it was predicted that /iː/—/ɪ/ and /ɔː/—/ʌ/ would be moderate-to-difficult to discriminate, while /ɪ/—/ɛ/ and /ɛ/—/æ/ would both be easy to discriminate. As illustrated by Table 19, the predictions held true for /iː/—/ɪ/, /ɛ/—/æ/ and /ɔː/—/ʌ/. The first was discriminated moderately at 82.5%, while the third was the most difficult contrast, discriminated accurately only 55% of the time. The second was the easiest contrast, discriminated accurately 98.75% of the time. The combination of assimilation pattern types and overlap scores thus offered a relatively reliable way to predict discrimination accuracy.

The results for /ɪ/—/ɛ/, however, deviated more from the predictions. Although discrimination for this contrast was predicted to be easy, it was only discriminated accurately in 70% of the cases. Recall that /ɪ/ was misidentified as /ɛ/ at a rate of 10%, whereas /ɛ/ was not misidentified as /ɪ/. Assimilation overlap was also small: the vowels were assimilated to the Finnish /e/ at rates of 10% (for /ɪ/) and 100% (/ɛ/), respectively, giving an overlap score of just 10%. It could be that, although the participants were capable of differentiating the vowels when they were presented separately, they struggled more in the discrimination task, where the vowels were presented together and where they were specifically asked to consider their similarity.

/iː/—/ɪ/ was discriminated accurately at a rate of 82.5%. Above, it was found that /iː/ was misidentified as /ɪ/ in 2.5% of the cases, while /ɪ/ was misidentified as /iː/ in 5% of the cases.

Both vowels were also assimilated primarily to the Finnish /i/ at rates of 95% for /iː/ and 85% for /ɪ/. The moderate confusion of these two vowels, as illustrated by the identification and assimilation tasks, could explain the difficulties the participants had in discriminating them.

/ɛ/—/æ/ was the easiest contrast to discriminate, reaching an almost perfect accuracy rate at 98.75%. Recall that neither vowel was identified as the other, and that /ɛ/ was not assimilated to the Finnish /æ/ and vice versa. In other words, the participants did not appear to find /ɛ/ and /æ/ similar; consequently, discrimination was very reliable.

/ɔː/—/ʌ/ had by far the poorest discrimination rate, with just 55% of the cases successfully discriminated. The problems the participants had with both identifying and assimilating these sounds thus extended to discrimination.

Two-sample t-tests were performed to see if there was a statistically significant context effect. Statistical significance was reached only for the bilabial /ɔː/—/ʌ/, $t(19)=-5.339$, $p=.000$. Above, statistical significance was also not found for any of the English vowels when investigating context effect on vowel identification. Furthermore, Levy and Strange (2008) likewise found that context was not a statistically significant factor in the discrimination of non-native contrasts by advanced learners, although Balas (2018) found that context was statistically significant in the case of velar /gVg/ discriminations. As the discrimination results of beginner learners in Levy and Strange (2008) did show statistically significant consonant effect, it could be that the advanced status of the learners resulted in the lack of significance in the present study.

However, as noted above, even though statistical significance was not reached for most contrasts, context effects can still be discerned in the data. They are illustrated in Table 20, which shows the mean correct discrimination of English vowel contrasts in bilabial and alveolar contexts, respectively.

		Mean correct discrimination
English vowel contrast	Context	
/iː/—/ɪ/	Bilabial	85%
	Alveolar	80%
/ɪ/—/ɛ/	Bilabial	72.5%
	Alveolar	67.5%
/ɛ/—/æ/	Bilabial	97.5%
	Alveolar	100%
/ɔː/—/ʌ/	Bilabial	40%
	Alveolar	70%

Table 20. Mean correct discrimination of English vowel contrasts in bilabial and alveolar contexts.

As can be seen from Table 20, overall, discrimination rates do not appear to vary considerably between the two contexts. Indeed, it was noted above that upon reaching a certain level, difficulty identifying or assimilating phonemes appears to generally extend to both contexts.

Again, the largest difference between bilabial and alveolar contexts can be found for /ɔː/—/ʌ/, which was discriminated with just 40% accuracy in the bilabial context and at 70% in the alveolar context. Recall that /ʌ/ was identified 65% accurately in the bilabial context, and was mistaken for /ɔː/ in the remaining 35% of cases. In the alveolar context, /ʌ/ was identified 80% accurately and was mistaken for /ɔː/ in the remaining 20% of the cases. It is thus not surprising that the participants struggled with discriminating the sounds, especially in the bilabial context.

On the basis of discrimination results, it is also possible to evaluate the NRVF claim that a change from a more peripheral vowel to a more central one is more difficult to discriminate than the other way around (Polka and Bohn 2011). Figure 2 shows the peripherality relationships of the English vowels /iː ɪ ɛ æ ɔː ʌ/.

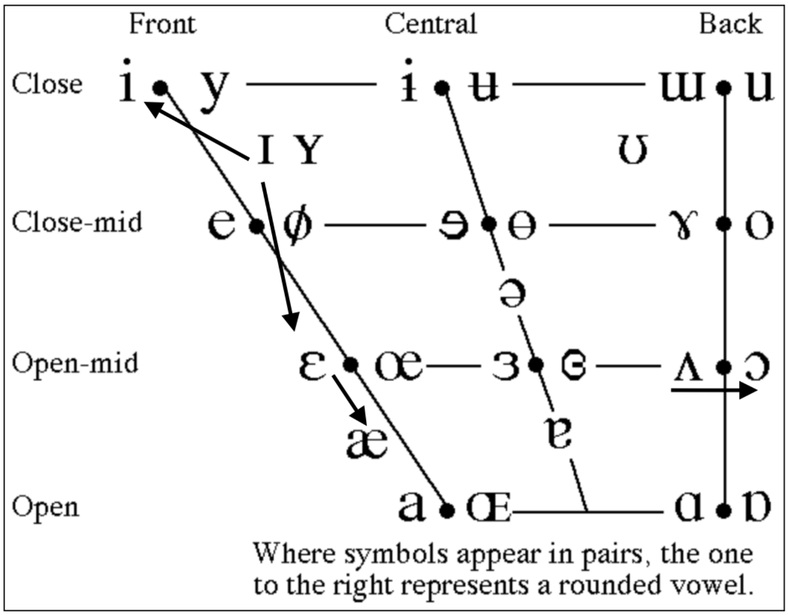

Figure 2. Peripherality relationships of the English vowels /iː ɪ ɛ æ ɔː ʌ/. Arrows point towards more peripheral vowels. Adapted from UCLA Phonetics Lab.

Tyler et al. (2014) found directional asymmetries only in contrasts assimilated in Single Category pattern, whereas Balas (2018), who also investigated the perception of advanced learners, did not find them. In the present study, no contrast was primarily assimilated in Single Category pattern; indeed, in the bilabial context, only /iː/—/ɪ/ and /ɔː/—/ʌ/ were assimilated in this pattern at rates of 30% and 15%, respectively, while in the alveolar context, the rates were 30% /iː/—/ɪ/, 5% for /ɪ/—/ɛ/ and 25% for /ɔː/—/ʌ/. However, as the participants in Tyler et al. (2014) were naïve listeners, it does not necessarily hold that directional asymmetries are found in the present study (only) for contrasts assimilated in SC pattern.

Two-sample t-tests were performed to see if the direction in which the vowels were presented was statistically significant. The results of the tests confirmed statistical significance for the alveolar /ɔː/—/ʌ/ contrast, $t(19)=-3.684$, $p=.002$. Thus, in contrast to Balas (2018), some evidence of directional asymmetries in the perception of advanced Finnish learners could be found in the present study. Possible explanations for this are discussed below.

Table 21 presents the mean correct discrimination of English vowel contrasts in the bilabial context and in both directions.

		Mean correct discrimination
English vowel contrast	Direction	
/iː/—/ɪ/	Peripheral→ central	85%
/ɪ/—/iː/	Central→ peripheral	85%
/ɛ/—/ɪ/	Peripheral→ central	60%
/ɪ/—/ɛ/	Central→ peripheral	85%
/æ/—/ɛ/	Peripheral→ central	100%
/ɛ/—/æ/	Central→ peripheral	95%
/ɔː/—/ʌ/	Peripheral→ central	45%
/ʌ/—/ɔː/	Central→ peripheral	35%

Table 21. Mean correct discrimination of English vowel contrasts in the bilabial context and in both directions.

As can be seen from Table 21, moderate directional asymmetries can be distinguished in the bilabial context for all the English vowel contrasts with the exception of /iː/—/ɪ/, which was discriminated with 85% accuracy in both directions. Only /ɛ/—/ɪ/ ~ /ɪ/—/ɛ/ was discriminated more accurately (60% versus 85%) when the change was from the central to the peripheral vowel, as suggested by NRVF. The other two contrasts, /æ/—/ɛ/ ~ /ɛ/—/æ/ and /ɔː/—/ʌ/ ~ /ʌ/—/ɔː/, were both discriminated more accurately when the first vowel presented was the peripheral vowel. This suggests that, for these contrasts, vowel peripherality was not the reason for the asymmetry.

Again, identification and assimilation results offer a potential explanation. Recall that /ʌ/ was mistaken for /ɔː/ more often than the other way around in the bilabial context. In other words, /ʌ/ was perceived as more similar to /ɔː/. When the order of presentation was peripheral to central, that is, /ɔː/—/ʌ/, the participants, when presented with /ʌ/, had already heard /ɔː/, which could have helped them differentiate between the sounds, resulting in a slightly more accurate discrimination rate than when /ʌ/ was presented first.

This explanation does not hold for the /æ/—/ɛ/ ~ /ɛ/—/æ/ contrast, as these two vowels were not misidentified as each other and also had no assimilation overlap. However, it is worth noting that the asymmetry was only 5%, suggesting that in the vast majority of the cases, direction does not influence the perception of this context. Such a small effect could consequently possibly be attributed to an individual perception error.

As noted above, Tyler et al. (2014) found that naïve listeners displayed directional asymmetries in the perception of contrasts assimilated in the Single Category pattern. Figure 3 shows the mean correct discrimination of English vowel contrasts, categorised by the most frequent assimilation pattern, in the bilabial context and in both directions.

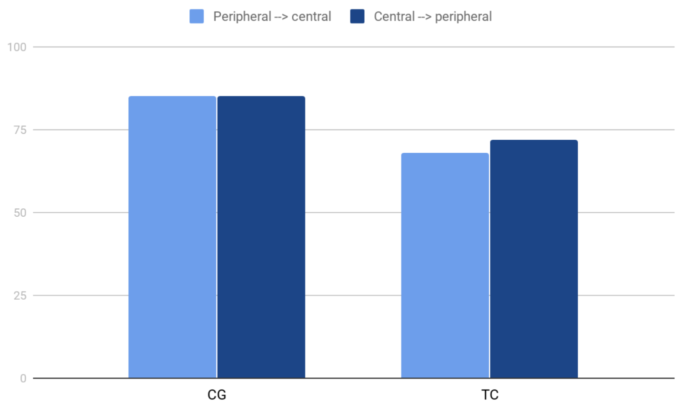

Figure 3. Mean correct discrimination (%) of English vowel contrasts in both directions per assimilation pattern. Bilabial context.

As can be seen from Figure 3, overall, discrimination was more accurate for the Category Goodness pattern, which consisted of the /i:/—/ɪ/ contrast. This goes against the proposition of PAM-L2 that contrasts assimilated in Two Category pattern should be easier to discriminate than those assimilated in Category Goodness pattern. This discrepancy is potentially explained by that, as stated, only one contrast was assimilated in the CG pattern, while the remaining three were assimilated in the TC pattern. In other words, there may not have been enough instances of different patterns to produce statistically valid results.

Furthermore, as noted above, the NRVF proposition that perceiving change from a more central to a more peripheral vowel is easier only held true for contrasts assimilated in TC pattern (more specifically, only in the case of the /ɛ/—/ɪ/ ~ /ɪ/—/ɛ/ contrast). Some reasons for the divergent results achieved for other contrasts were briefly discussed above.

Table 22 presents the mean correct discrimination of English vowel contrasts in the alveolar context and in both directions.

		Mean correct discrimination
English vowel contrast	Direction	
/iː/—/ɪ/	Peripheral→ central	75%
/ɪ/—/iː/	Central→ peripheral	85%
/ɛ/—/ɪ/	Peripheral→ central	55%
/ɪ/—/ɛ/	Central→ peripheral	80%
/æ/—/ɛ/	Peripheral→ central	100%
/ɛ/—/æ/	Central→ peripheral	100%
/ɔː/—/ʌ/	Peripheral→ central	45%
/ʌ/—/ɔː/	Central→ peripheral	95%

Table 22. Mean correct discrimination of English vowel contrasts in the alveolar context and in both directions.

As illustrated by Table 22, directional asymmetries are also present in the alveolar context. An exception is the /æ/—/ɛ/ ~ /ɛ/—/æ/ contrast, which showed a very small asymmetry of 5% in the bilabial context but was discriminated with 100% accuracy in both directions in the alveolar context. The /iː/—/ɪ/ ~ /ɪ/—/iː/ contrast, which showed no asymmetry in the bilabial context, shows moderate asymmetry in the alveolar context, as it was perceived more accurately (85% versus 75%) when the change was from the central to the peripheral vowel. A stronger asymmetry is visible for /ɛ/—/ɪ/ ~ /ɪ/—/ɛ/, which was discriminated with 80% accuracy when the central vowel came first but with just 55% accuracy when the peripheral vowel came first. Finally, as in the bilabial context, the greatest asymmetry is present for /ɔː/—/ʌ/ ~ /ʌ/—/ɔː/ contrast, perceived with 45% accuracy when the vowel change was from peripheral to central and with 95% accuracy when the vowel change was from central to peripheral. In the alveolar context, the NRVF claim that a change from a more central vowel to a more peripheral one is easier to discriminate than the other way around held true for all contrasts that exhibited directional asymmetry.

The alveolar /ɔː/—/ʌ/ ~ /ʌ/—/ɔː/ contrast was the only one where directional asymmetry reached statistical significance, $t(19)=-3.684$, $p=.002$. Indeed, at 50%, the asymmetry is larger than for any other contrast in either context. Recall that /ɔː/ was misidentified as /ʌ/ at a rate of 10% in the alveolar context. The two vowels also had an overlap score of 90 in the alveolar context (in comparison to 50 in the bilabial one). Above, in relation to the bilabial context, it was suggested that the discrimination rate for /ɔː/—/ʌ/ may have been slightly higher because the participants were overall better at distinguishing /ɔ:/ than /ʌ/. Conversely, it is possible that in the alveolar context, the two sounds were considered less acoustically similar, as the unrounded /ʌ/ would not have been rounded to resemble /ɔ:/. Subsequently, the difference may have been starker when /ʌ/ was presented first, leading to excellent discrimination.

Figure 4 shows the mean correct discrimination of English vowel contrasts, categorised by the most frequent assimilation pattern, in the alveolar context and in both directions.

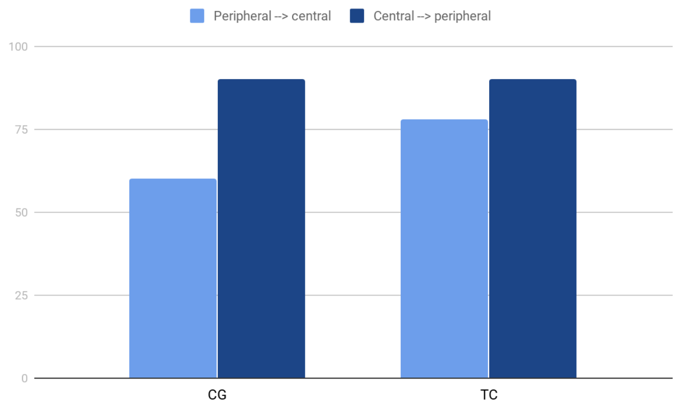

Figure 4. Mean correct discrimination (%) of English vowel contrasts in both directions per assimilation pattern. Alveolar context.

As illustrated by Figure 4, discrimination for Two Category pattern, consisting of the /ɪ/—/ɛ/ and /ɛ/—/æ/ contrasts, was more accurate than discrimination for Single Category pattern, consisting of the /i:/—/ɪ/ and /ɔ:/—/ʌ/ contrasts. In opposition to the bilabial context, the alveolar context thus supports the PAM-L2 prediction that TC contrasts are easier to discriminate. Similarly, in the alveolar context, discrimination was overall more accurate when the change was from a more central to a more peripheral vowel, which supports the NRVF.

In conclusion, discrimination of English vowel contrasts varied between almost perfect (98.75%) for /ɛ/—/æ/ to very poor (55%) for /ɔ:/—/ʌ/. With the exception of /ɪ/—/ɛ/, assimilation patterns and overlap scores reliably predicted discrimination accuracy. Context differences could be distinguished, and were especially strong for /ɔ:/—/ʌ/. In the alveolar context, /ɔ:/—/ʌ/ was the only contrast reaching statistically significant direction asymmetries, as proposed by NRVF. The results were more mixed with regards to both PAM-L2 and NRVF predictions in the bilabial context than in the alveolar one. As no contrast was assimilated primarily in Single Category pattern, the findings of Tyler et al. (2014) could not be evaluated for advanced learners; the findings of Balas (2018), on the other hand, were not replicated, as directional asymmetries could be distinguished for the majority of the contrasts.

6. Discussion

The present study aimed at answering the following research questions:

1. How do advanced Finnish learners of English assimilate the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?
2. How accurately do advanced Finnish learners of English identify the English vowels /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ɔː/?
3. How accurately do advanced Finnish learners of English discriminate the English vowel contrasts /iː/—/ɪ/, /ɛ/—/æ/, /ɪ/—/ɛ/, and /ɔː/—/ʌ/?
  3a. Can discrimination be predicted by assimilation patterns and overlap scores?
  3b. Are there directional asymmetries in discrimination, as proposed by NRVF?
4. Does the consonant context (bilabial versus alveolar) affect perception?

As can be seen from the research questions, the main aim of the paper was to investigate three different aspects of non-native speech perception: assimilation, identification and discrimination. The paper also investigated whether assimilation could predict discrimination and whether directional asymmetries were present. Finally, the effect of consonant context was also considered.

Regarding the first research question, advanced Finnish learners generally assimilated English vowels to one Finnish vowel category at rates of >85%. The exceptions were /ʌ/ and /ɔː/, which were assimilated at rates <60%. Somewhat similar results were reached by Balas (2018), in whose study the majority of vowels were assimilated to one category at rates above >80%. Balas (2018) did not consider the influence of context on assimilation, but in the present study, its effect proved to be relatively minor, with /ʌ/ and /ɔː/ also displaying the strongest context effects.

Regarding the second research question, advanced Finnish learners identified English vowels generally accurately, with only /ʌ/ identified with an overall rate of <80%. /ɪ/ was identified as three separate English vowels, suggesting that the participants found it similar to other English vowels more frequently than in other cases. Although consonant effect could be distinguished, especially for /ʌ/, it failed to reach statistical significance, suggesting that the participants' mental representations of the vowels were relatively context-independent. This aligns with Levy and Strange (2008), who found consonant context did not reach statistical significance for advanced learners; Balas (2018), meanwhile, found that the velar context, but not the bilabial or the alveolar context, was enough for statistical significance to be reached. Another study involving other consonant contexts in addition to the bilabial and alveolar ones, as well as comparing beginner and advanced learners, could help determine the reasons for these findings.

Another proposition for future research is a slightly different methodological approach. The identification task in the present study was forced-choice, and only the vowels investigated were offered as response options. An open-choice task, where participants can freely respond with any English vowel they consider the most suitable, could possibly result in additional identification patterns; however, because it would require a relatively high level of familiarity with the English language, this approach might not be suitable when investigating the perception of beginner learners.

Regarding the third research question, discrimination was overall relatively good, although for /ɔː/—/ʌ/, it reached only a rate of 55%. The initial hypothesis that discrimination will be mildly-to-moderately better in the bilabial context did not hold true for /ɛ/—/æ/ and /ɔː/—/ʌ/, although in the former case, the difference was minimal (2.5%). Assimilation patterns and overlap scores accurately predicted discrimination with the exception of /ɪ/—/ɛ/. As with the identification task, context effects did not reach statistical significance with the exception of the alveolar /ɔː/—/ʌ/, but could be distinguished for all of the contrasts, albeit generally at low rates. Directional asymmetries could furthermore be distinguished, in contrast to Balas (2018), although on two occasions in the bilabial context they actually went against NRVF propositions. Individual perception errors and, in the case of /ɔː/—/ʌ/, the overall difficulty Finnish learners had with perceiving these sounds were offered as explanations. Again, a study comparing advanced and beginner learners could either confirm or deny this suggestion.

Some methodological issues and general future research suggestions can be briefly considered. The sample size in the present study was relatively small at twenty. This raises the possibility that not enough statistical power was present to reach valid conclusions, although the lack of statistical significance with regards to both identification and discrimination is supported by other studies. The stimuli used were also not judged by a native speaker. Although they were compared to pre-existing acoustic data and did not appear to deviate significantly from it, suggesting that they were reliable representations, native judgements could have pointed out any possible anomalies, and could be utilised in future studies as an additional measurement. While the use of closed syllables allowed the inclusion of lax vowels and increased control over the consonant context, using isolated and synthetic stimuli, as well as stimuli extracted from natural speech, could potentially produce different results, although in the case of the first two it might not be possible to generalise them to perception of everyday speech. Methods in additional to behavioural ones, such as neuroimaging techniques, could also be utilised to further investigate the basis of perception. Finally, future research could the production dimension to shed light on the perception-production link, which was not investigated here.

Overall, the results of the present study confirm the predictions of PAM-L2 and NRVF, and suggest that the three aspects of pronunciation investigated all interact with each other, so that difficulty in identifying and assimilating vowels generally carries over to difficulty in discriminating them. Several implications for teaching and learning English vowels can also be mentioned. On a general level, as illustrated by for example Hardison (2003), explicit training can help learners perceive difficult sounds more successfully. As accurate perception can potentially result in more accurate production, instruction in how to differentiate vowels, especially phonemic vowel pairs, can improve both listening and speaking skills. Of course, this can also benefit teachers, many of whom are L2 speakers themselves. Knowledge of the interrelatedness between different aspects of perception could also be employed to devise teaching methods that utilise these connections. Firstly, sounds that were misidentified at higher rates could be taught not only on their own but also in relation to the sounds as which they were misidentified. This could help learners refine their mental representations and consequently better distinguish these sounds. Secondly, the discrimination of two sounds assimilated to one Finnish category could be facilitated by making the learners more aware of this overlap. Thirdly, if learners are able to generalise familiar to the unfamiliar, as suggested for example by Pajak and Levy (2014), then teaching the /iː/—/ɪ/ distinction by focusing on the spectral cues could enable learners to more successfully weigh spectral cues with regards to other relevant contrasts as well. Finally, more attention overall could be paid to /ʌ/ and /ɔː/, which appear to be the most difficult vowels for Finnish learners to perceive, even at the advanced level.

BIBLIOGRAPHY

Alispahic, S., Mulak, K. E., and Escudero, P. (2017). Acoustic Properties Predict Perception of Unfamiliar Dutch Vowels By Adult Australian English and Peruvian Spanish Listeners. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2017.00052.

Asano, Y. (2018). Discriminating Non-Native Segmental Length Contrasts Under Increased Task Demands. *Language and Speech* 61 (3), 409-429.

Baker, W. and Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching* 44 (3), 231-250.

Balas, A. (2018). English vowel perception by Polish advanced learners of English. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 63 (3), 309-338.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press, 171-204.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., and Quann, C. A. (2009). Development of Phonological Constancy. Toddlers' Perception of Native and Jamaican-Accented Words. *Psychological Science* 20 (5), 539-542.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109 (2), 775-794.

Best, C. T., McRoberts, G. W. and Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance* 14 (3), 345-360.

Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro and O.-S. Bohn (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 13-34.

Bohn, O.-S. and Steinlen, A. (2003). Consonantal context affects cross-language perception of vowels. *Proceedings of the 15th International Congress of Phonetic Sciences,* 2289-2292.

Bosch, L. and Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech* 46 (2-3), 217-243.

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M. and Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science* 18 (4), 587-598.

Bundgaard-Nielsen, R., Best, C. T., and Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32 (1), 51-67.

Carreiras, M., Gillon-Dowens, M., Vergara, M., and Perea, M. (2009). Are Vowels and Consonants Processed Differently? Event-related Potential Evidence with a Delayed Letter Paradigm. *Journal of Cognitive Neuroscience* 21 (2), 275-288.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory and Cognition* 28 (5), 746-755.

De Klerk, M., de Bree, E., Kerkhoff, A. and Wijnen, F. (2018). Lost and Found: Decline and Reemergence of Non-Native Vowel Discrimination in the First Year of Life. *Language Learning and Development,* https://doi.org/10.1080/15475441.2018.1497490.

Díaz, B., Mitterer, H., Broersma, M., Escera, C., and Sebastián-Gallés, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition* 19 (5), 955-970.

Edwards, H. T. (2003). *Applied Phonetics: The Sounds of American English*. Clifton Park: Delmar Learning.

Elvin, J., Escudero, P. and Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2017.00052.

Escudero, P., Benders, T. and Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics* 37, 452-465.

Escudero, P. and Wanrooij, K. (2010). The Effect of L1 Orthography on Non-native Vowel Perception. *Language and Speech* 53 (3), 343-365.

Evans, B. and Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics* 68, 15-31.

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

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24 (4), 495-522.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97 (5), 3099-3111.

Iverson, P. and Evans, B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America* 126 (2), 866-877.

Kartushina, N. and Frauenfelder, U. H. (2013). On the role of L1 speech production on L2 perception: Evidence from Spanish learners of French. *Interspeech 2013*. https://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_2118.pdf. (9 December 2018.)

Kim, D., Clayards, M. and Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics* 67, 1-20.

Kriengwatana, B. P. and Escudero, P. (2017). Directional Asymmetries in Vowel Perception of Adult Nonnative Listeners Do Not Change Over Time With Language Experience. *Journal of Speech, Language, and Hearing Research* 60, 1088-1093.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9 (2), 13-21.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1493), 979-1000.

Leppänen, S., Nikula, T. and Kääntä, L. (2008). *Kolmas kotimainen: lähikuvia englannin käytöstä Suomessa*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *Journal of the Acoustical Society of America* 126 (5), 2670-2682.

Levy, E. and Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36, 141-157.

Liu, L. and Kager, R. (2016). Perception of a native vowel contrast by Dutch monolingual and bilingual infants: A bilingual perceptual lead. *International Journal of Bilingualism* 20 (3), 335-345.

Mazuka, R., Hasegawa, M. and Tsuji, S. (2014). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology* 56 (2), 192-209.

McAllister, R., Flege, J. E., and Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229-258.

Meador, D., Flege, J. E., and MacKay, I. R. A. (2000). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106 (5), 2973-2987.

Nam, Y. and Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition* 155, 57-66.

Narayan, C. R., Werker, J. F. and Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science* 13 (3), 407-420.

Ogden, R. (2009). *An introduction to English phonetics*. Edinburgh: Edinburgh University Press.

Omote, A., Jasmin, K. and Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex* 93, 146-154.

Pajak, B. and Levy, R. (2014). The role of abstraction in non-native speech perception. *Journal of Phonetics* 46, 147-160.

Perrachione, T. K., Lee, J., Ha L. Y. Y., and Wong P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America* 130 (1), 461-472.

Polka, L. and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication* 41 (1), 221-231.

Polka, L. and Bohn, O.-S. (2011). Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics* 39 (4), 467-478.

Polka, L. and Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20 (2), 421-435.

Poltrock, S. and Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology* 131, 135-148.

Sajavaara, K. and Dufva, H. (2001). Finnish-English Phonetics and Phonology. *International Journal of English Studies* 1 (1), 241-256.

Sato, Y., Kato, M. and Mazuka, R. (2012). Development of single/geminate obstruent discrimination by Japanese infants: early integration of durational and nondurational cues. *Developmental Psychology* 48 (1), 18-34.

Sennema, A., Hazan, V. and Faulkner, A. (2003). The role of visual cues in L2 consonant perception. *Proceedings of the 15th International Congress of Phonetic Sciences*.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39 (4), 456-466.

Strange, W. and Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. H. Edwards and M. L. Zampini (eds.), *Phonology and second language acquisition*. Amsterdam: John Benjamins, 153-192.

Suomi, K., Toivanen, J. and Ylitalo, R. (2008). *Finnish sound structure: Phonetics, phonology, phonotactics and prosody*. Oulu: Oulu University Press. Studia humaniora ouluensia 9.

Tyler, M. D., Best, C. T., Goldstein, L. M. and Antoniou, M. (2014a). Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental Psychobiology* 56 (2), 210-227.

Tyler, M.D., Best, C. T., Faber, A. and Levitt, A. G. (2014b). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71 (1),

UCLA Phonetics Lab. http://www.phonetics.ucla.edu/course/chapter1/vowels.html. (30 April, 2019.)

van Leussen, J.-W. and Escudero, P. (2015). Learning to perceive and recognize a second language: the L2P2 model revised. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2015.01000.

Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7 (1), 49-63.

Wiener, S. and Turnbull, R. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech* 59 (1), 59-82.

Wiik, K. (1965). *Finnish and English vowels: A comparison with special reference to the learning problems met by native speakers of Finnish learning English.* Turku: University of Turku. Annals of the University of Turku 94.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R. and Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience* 22 (6), 1319-1332.

APPENDICES

Appendix 1: Stimulus words

	Bilabial context	Alveolar context
	/biːb/	/diːd/
	/bɪb/	/dɪd/
	/bɛb/	/dɛd/
	/bæb/	/dæd/
	/bʌb/	/dʌd/
	/bɔːb/	/dɔːd/

Appendix 2: Experiment 1 answer sheet

EXPERIMENT 1: SOUND DISCRIMINATION

You will hear two English words. Pay attention to the **vowel** sounds in the words. Is the vowel in the second word the same (S) or different (D) as in the first word? Circle your answer.

1. S D	**9.** S D	**17.** S D	**25.** S D
2. S D	**10.** S D	**18.** S D	**26.** S D
3. S D	**11.** S D	**19.** S D	**27.** S D
4. S D	**12.** S D	**20.** S D	**28.** S D
5. S D	**13.** S D	**21.** S D	**29.** S D
6. S D	**14.** S D	**22.** S D	**30.** S D
7. S D	**15.** S D	**23.** S D	**31.** S D
8. S D	**16.** S D	**24.** S D	**32.** S D

Appendix 3: Experiment 2 answer sheet

EXPERIMENT 2: SOUND IDENTIFICATION

You will hear an English word. Pay attention to the **vowel** sound in the word. Which of the following English vowels matches the one in the word you heard? Circle your answer.

1. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
2. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
3. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
4. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
5. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
6. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
7. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
8. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
9. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
10. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
11. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/
12. /iː/ /ɪ/ /ɛ/ /æ/ /ʌ/ /ɔː/

Appendix 4: Experiment 3 answer sheet

EXPERIMENT 3: SOUND ASSIMILATION

You will hear an English word. Pay attention to the **vowel** sound in the word. Which Finnish vowel does it resemble the most? Circle your answer.

1. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
2. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
3. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
4. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
5. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
6. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
7. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
8. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
9. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
10. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
11. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/
12. /ɑ/ /e/ /i/ /o/ /u/ /y/ /ø/ /æ/

You will hear the word again. How well does the **vowel** sound in the word resemble the Finnish vowel you have chosen? Circle your answer.

1	**2**	**3**	**4**	**5**
Very remotely	Not very well	Fairly well	Very well	Identical

1.	1	2	3	4	5
2.	1	2	3	4	5
3.	1	2	3	4	5
4.	1	2	3	4	5
5.	1	2	3	4	5
6.	1	2	3	4	5
7.	1	2	3	4	5
8.	1	2	3	4	5
9.	1	2	3	4	5
10.	1	2	3	4	5
11.	1	2	3	4	5
12.	1	2	3	4	5