Tampere University

Mikko Pajunen

COMMUNICATIVE PRESSURE AND MISSING SURFACE PHONOLOGICAL FEATURES

Faculty of Information Technology and Communication Sciences
Master's Thesis
March 2023

ABSTRACT

Mikko Pajunen: Communicative Pressure and Missing Surface Phonological Features
Master's Thesis
Tampere University
Master's Program in Language Studies
March 2023

The manner in which communicative pressure may cause variable forms in speech is occasionally discussed in L2 research. However, the discussion has been limited mostly to syntax and morphology, and what has remained largely unexplored is the way communicative pressure and cognitive processing limitations may affect phonological surface forms. Accordingly, L2 speaker phonology warrants investigation in this framework since phonology, much like morphology and syntax, similarly consists of organized and structured representations.

This thesis explores phonological optionality with consideration to cognitive processing limitations in L2 speech production. The presented investigation is a pilot study that tests a research design by which individuals are observed realizing their L2 phonology differently in two contexts: one which involves communicative pressure and one which does not. The hypothesis is that communicative pressure (i.e., situations involving "online" speech production) sets limits to working memory, temporarily making phonological features in specific categories inaccessible, resulting in non-target pronunciation. The pronunciation, it is hypothesized, becomes more target-like once the communicative pressure variable is removed.

Finnish university students undergoing language teacher training were recruited to the study and recorded during their practice lessons in which the language of communication was English. Based on several hours of classroom audio data, the participants' pronunciation was evaluated and tokens of non-target pronunciation in specific words were selected as stimuli for the study's second phase, which involved a recorded elicitation task. In addition to the elicitation task, the participants completed a receptive task based on the same stimuli. Both tasks were tailored to each participant as the stimuli were based on each participant's phonological performance as observed in the classroom.

The methods used in this study are qualitative, involving formant analyses with spectrographic and waveform illustrations conducted on the elicitation task audio data. The participants were additionally interviewed about their speech attitudes and strategies regarding L2 English. Specifically, the tokens analyzed in this study involved sibilant and plosive-sibilant clusters with phonation as a variable, lateral approximants with either velarized or non-velarized allophone as a variable, dental fricatives as part of assimilatory processes at word boundaries, L-vocalization in syllable rhymes, and word-final plosives with either presence or absence of release burst as a variable.

The results of the analysis may be summarized as follows: the study participants realized the allophone or phoneme that is closer to target language norms in the experimental situation in which communicative pressure was absent. They also judged the target-like version token as more correct than the non-target version in the receptive task. However, this finding did not replicate for all tokens. The validity of the investigated phonological features in addition to other methodological features of the study are assessed in the concluding chapters.

The findings are interpreted to mean that the classroom context, in which the participants were initially observed, involves communicative pressure, explaining why phonological performance errors were more frequent in the classroom than in the experimental context. Based on earlier work on morphological and syntactic categories, it is suggested that working memory may become burdened by, for instance, psychological stress, resulting in the target phonological features, which are underlyingly represented, being temporarily unrealized in the phonological surface representations of L2 speakers.

Keywords: Phonology, Phonetics, Second Language Acquisition, Second Language Speech, Working Memory

Contents

List of Figures

List of Tables

1 Introduction

The allophones selected to represent phonemes by second language (L2) speakers show high variation across individuals. However, for even one individual, the distribution of L2 phonological rules in various phonemic contexts is often inconsistent and a specific phoneme may be realized differently in the same word and environment. Setting aside the fact that the various possible features of phonemes are expressed differently depending on the phonetic environment, the question of what factors are regulating and constraining the variable access to the phonemic inventory remains. Second language research has long sought answers to questions of this sort, resulting in intriguing findings from morphological studies which in turn have led to hypotheses pertaining to how cognition affects the availability of morphemes in language production. Also important for understanding the phenomenon of L2 production is that certain forms appear in specific contexts, suggesting that the situation, or perhaps the psychological context of the situation, affects how language is expressed (Tarone 1979, 1983). Might similar phenomena underlie cases in which the phonemic or sub-phonemic level is accessed inconsistently?

Evidence is scant with respect to how emotions and the psychological state evoked by specific contexts and situations may influence L2 phonology, and how heightened stress, for instance, might distort the transfer from underlying representations to surface phonology. How second language production might be affected, and particularly the way L2 phonology might be temporarily degraded in certain contexts or psychological states, is a deficit in psycholinguistics literature that, to the knowledge of the author of this thesis, has been inadequately addressed. Of course, stress and its influence on language is difficult to investigate within ethical boundaries, particularly due to concerns of conferring minimal harm to those involved. However, one possible way to tackle this is by investigating the relatively harmless "stage fright" phenomenon which typically involves individuals speaking, on their own volition, in front of an audience. In such situations, psychological distress may be disproportionately high in some individuals, yet the "stakes" of the situation are objectively incongruous to the distress experienced in the situation. In short, the situation is not dangerous, yet the individual's nervous system may react to it as one would when perceiving a real threat.

The reasons for studying this phenomenon are important for multiple reasons. To shed light on the fact that L2 speakers may not always accurately represent their phonological competence, especially while under various forms of psychological pressure, brings with it implications that should be acknowledged in various professional contexts. To illustrate, job interviews may be conducted in languages which the applicants do not speak as their L1. Yet the linguistic performance of the L2 speaker in such situations may influence hiring decisions. Particularly, how a person sounds while speaking his or her L2 is potentially subject to prejudice, and L2 speakers may be in danger of becoming discriminated against based on incorrect representations of their competence. An important question, thus, is whether situations involving pressure arising from psychological factors – communicative, emotional, or otherwise – yield L2 speaker data that represents competence or, alternatively, whether competence is distorted by performance errors unique to the context and the psychological state experienced by the speaker.

During communication, the brain must meet the cognitive needs for forming coherent sentences and retrieving from memory the words relevant to what is being expressed. Central to this cognitive function is thought to be the working memory (WM). WM and related neurocognitive processes have emerged as promising concepts in explaining processing limitations involved in various language situations (e.g., Aboitiz 1995, 2018; Aboitiz *et al*. 2010, 2006; Ahmadian 2012a, 2012b; Acheson 2009, 2011; Juffs & Harrington 2011). However, most theories implicating WM in an attempt to account for optionality in L2 production do not involve underlying knowledge of L2 phonology, but instead focus on functional categories and morphological forms. The goal of this thesis is to test the hypothesis that phonological categories, i.e., phonemes to which different phones are assigned on a language-specific basis, are affected similarly to functional categories and inflectional morphology as described by the Missing Surface Inflection hypothesis (MSIH) (Prévost & White 2000; Haznedar & Schwartz 1997) in communicative situations involving highly proficient L1 Finnish-L2 English speakers. However, it should be specified that the analogy to missing surface inflection is not perfect, since phonological categories do not involve morphs that are left unexpressed, but instead, usually, the speaker's defaulting to L1 phones, to name one example. The focus is instead on whether psychological factors similarly affect realization of the appropriate language-specific sub-phonemic features, such as [+voice] and [-voice] features of certain sibilants in English words, and whether the contrasts have been acquired, but not realized in production.

L2 acquisition researchers have long struggled to account for why interlanguage (henceforth IL) grammars produce the target and non-target (henceforth TL) forms that they do. It is difficult to ascertain what has and has not been acquired in the L2 speaker IL. Optionality may offer some insight as to the nature of this phenomenon. MSIH suggests that, instead of an impairment in the representation, L2 speakers have problems with realizing the surface morphology when processing limitations impede proper retrieval (Meisel 1997; Lardiere 1998a, 1998b; Prévost & White 2000; Haznedar 2001, 2003). Crucially, the IL has underlying knowledge of the L2 morphology or phonology which is not necessarily realized in real-time production and may thus give an inaccurate representation regarding the competence of the speaker. Access to L2 forms is thought to be mediated by WM and the grammatical and, by extension, phonological representations are values transferred from the corresponding L1 systems (White 1991, 1992).

The hypothesis is as follows: IL phonologies of advanced L2 English speakers exhibit a mixture of representational and computational variability. The representational paradigm can be summarized as L1 transfer while the computational involves inconsistent suppliance of acquired target language phonemes. As stated in McCarthy (2008), with respect to MSIH, "[I]f variability is strictly a production-based phenomenon, then eliminating communication pressure should eliminate variability, or at least severely reduce its appearance." A parallel argument is adopted for this hypothesis, and it is presumed that IL phonology operates by a similar underlying mechanism: processing limitations may affect features of phonemes as well as surface morphology.

Phonological acquisition always involves assigning phones to phonemes, and in this sense there are three categories of phone: (1) matches the L1, (2) belongs to a different phoneme in the L1, and (3) does not exist in the L1 and has to be assigned by the learner. It is anticipated that removing communication pressures by changing from a classroom teaching context, which essentially involves some degree of performance, to a non-classroom context will effect changes on production. WM-related limitations may inhibit access to, or application of, a particular phonological rule determining, for instance, which phone is selected in a phonological context, while the phonemes may all be correctly organized (i.e., as sets of specific phones operating as allophones).

This thesis presents a pilot study that tests a method for capturing L2 production in situations requiring spontaneous and improvised oral production (i.e., communicative pressure) by highly

proficient L2 English speakers, in the form of classroom English teaching. L2 production errors are nuanced and highly variable as they may be systematic or exhibit optionality. However, it is likely that conclusions drawn regarding the degree to which a discrete category for a given phoneme has been established are misleading if target language phonemes are realized differently depending on the situation, a notion that shapes the following research questions:

Research Question 1: Does the nature of the communicative situation affect phonological production, giving rise to inaccurate representation of L2 phonological proficiency?

Research Question 2: When production is elicited in an experimental situation where the production is not part of a communicative situation, does this affect phonological output in comparison with a communicative situation?

The situational contingency under investigation here is a matter of degree of psychological pressure, one which both questions seek to answer. It is presumed that phonemes that match the language being spoken are retrieved with constant success in situations where the speaker is allowed to plan and self-correct his or her production, providing that the appropriate categories have been established by the IL phonology. However, planning and the production of utterances is impeded in real-time performance situations. It could thus be anticipated that if L2 speakers show optionality in a communicative situation in one or more speech sounds, optionality should decrease in a setting where communicative pressures have been obviated. The findings are of potential significance for various disciplines, including teachers and language researchers.

Answers to the research questions outlined above are sought through a research design involving multiple phases. The first phase consists of observing, recording, and transcribing the L2 English production during English classes taught by university students undergoing English teacher training. The data yielded by the first phase is thought to represent the type of output which is constrained by psychological factors described above. From this data, inaccurate single-segment L2 English tokens are extracted and reproduced for both oral and aural language tasks in the second phase of the study. The second phase is designed to place the participants observed in the classroom into a situation where they demonstrate L2 output free of communicative pressure. This output is guided by reading out loud from a list of words and phrases with similar phonetic environments in which the non-target segments were heard in the first phase of the study. The same phrases and words are used in a listening task where the participants judge the accuracy of the tokens in two versions of the word or phrase. The

elicitation data from the listening and production tasks are analyzed with respect to how they compare to the classroom data. The production task results are also analyzed using spectrographic techniques. This procedure is accompanied by a semi-structured interview by which additional details regarding pronunciation, and more broadly the use of L2 English, are obtained.

This thesis is divided into five sections. The second section discusses some of the concepts and theoretical frameworks central to the thesis, namely, phonological and second language acquisition literature. The methods and the procedure of the study are presented in the third section. The fourth section consists of analysis of the data obtained with the methods outlined in the third section. The results are discussed in the fifth section and the sixth section delivers concluding remarks on the presented work.

2 Background

2.1 Speech and the Phoneme

A basic concept in phonology, the phoneme is a mental representation of a speech sound, and within a particular language system, phonemes are distinctive. As such, the definition of a phoneme is that it can minimally distinguish words in a language, i.e., word pairs that differ in place of just one phone, such as *tin* and *pin*. One phoneme may also have features that divide it into several allophones, which are always defined in the context of one language, i.e., a phonological system. To illustrate, aspirated and non-aspirated plosives, /p p^h t t^h k k^h/[1], are phonemes in Hindi because they distinguish word meanings (Spencer 1996: 8). However, this is not the case in English in which aspirated and non-aspirated plosives can never form a minimal pair and thus they are allophones in English phonology. Therefore, in English, the corresponding phonemes are /p t k/. Allophony is contingent on complementary distribution, meaning that two sounds never occur in the same phonetic environment. For instance, [p^h] and [p] are allophones in English because [p^h] occurs only in syllable-initial positions and [p] in other positions.

Conceptually, phonemes are helpful for understanding linguistic sound systems and they are often typically, and perhaps erroneously, thought to imply one-to-one correspondence with speech sounds. They are also generally thought to be ordered in the actual phonetic speech chain, i.e., one speech sound, represented by one phoneme, occurs after the other, discretely, in a linear fashion. However, phonology observed at the level of connected speech does not support this view as speech sounds can be shown to overlap in articulation or "collapse" and become represented by a single phoneme. In a rapid stream of speech, word boundaries are deleted or suppressed as speech sounds within a syllable of a word carry over to the second

[1] The notational conventions in phonetics and phonology dictate that abstract phonemes be transcribed with slashes, / . . . /, while their concrete counterparts, phones, are placed inside square brackets, [. . .]. Phones are understood to represent the physical manifestation of a speech sound, which is why the transcription of phones is often accompanied by additional symbols to denote their various realizations more specifically. This is known as narrow transcription and its purpose is to illustrate what features have been added to the realization of the phoneme (e.g., aspiration represented by a superscript h-letter).

word which then becomes represented by a phoneme that is "supposed" to be part of the preceding word. This is illustrated by Lass (1984: 295-296), who provides an example of his own speech with the phrase *the difficulty is that I'm not sure about it*, in which the first word, *the*, consists of two phonemes, the consonant /ð/ and the vowel /ə/ but becomes realized as [dː], which proceeds to represent the first segment of the following word, *difficulty*. In other words, the surface representation of the first segment of *difficulty* subsumes the entire preceding word, *the*. Further, *that*, a word which is normally transcribed as [ðæʔt], is realized as [ə] in the same phrase.

The problem on display is that there is no one-to-one mapping between phonemes and sounds produced in speech, which may cause doubts about the phoneme's very existence. To rescue the concept, before they are realized in speech, words are thought to exist in a speaker's mind as phonetic citation forms, which are shared among the speakers of a given language. Individual speakers articulate the citation forms in their own ways, violating in the process restrictions on phoneme sequences which are typically rigid in citation forms[2], but the listener nevertheless understands that the single lax vowel, [ə], can represent the entire word, [ðæʔt]. This in turn effectively implies that listeners constantly hear things that are not there, and instead force a meaningful interpretation into the process of comprehension, a finding commonly replicated in perception studies.

Conceptually, citation forms and casual speech are closely related to underlying representations (URs) and their surface representations (SRs) from the SPE model (Chomsky & Halle 1968). In their model, the UR encodes all the information about a word's pronunciation that cannot be predicted by a rule. Phonological rules then effect changes on the word's pronunciation, such as segment substitution, reordering, deletion, insertion and so on, resulting in the SR (Spencer 1991: 63). URs may be thought to consist of feature matrices, constructed by the minds of language learners, but the exact nature of these matrices depends on which representational approach is accepted (see, for instance, Lahiri & Reetz 2002, 2010 for an alternative theory of URs to that of Chomsky & Halle).

Multiple definitions exist for the term phoneme (for a discussion, see Sommerstein 1977: 24) and various views exist regarding how phonemes and phonological systems are acquired by

[2] For instance, English syllable onsets prohibit consonant clusters such as [ts] and consequently loan words such as *tsunami* are regularly pronounced as [sunɑmi]. However, in word boundaries, phonotactic restructuring tends to occur and phrases such as *it seems* may be realized as [tsimz] (Lass 1984: 297). This shows that phonemes and phonotactic rules generally are valid but may be violated when expressed in a stream of speech.

language learners and what the underpinnings of the acquisition mechanisms are. For instance, Universal Grammar (Chomsky 1965, 1975, 1980, 1986), referred to as UG henceforth, is one theory that attempts to explain the acquisition of first language phonologies during the first years of life, as well as other areas of grammar, such as syntax. With respect to phonology, the UG framework postulates that human infants at birth have access to the full range of phonological features that can conceivably be produced by the human speech organs, and the infant's mind automatically acquires the phoneme features which are present in the speech of the ambient language environment. During the early years of life, the UG device switches off and becomes "deaf" to the rest of the possibilities that are not present, or contrastive, in the L1, even though they may be in other languages (White 2003: 154). Thus, in this view, Japanese L1 speakers are not able to hear the difference between /l/ and /r/ because Japanese lacks the phonemic feature [coronal], since it has been "switched off" in early childhood due to its absence in Japanese, leading L1 Japanese speakers to categorize both /l/ and /r/ as interchangeable speech sounds (ibid.: 155-156).

Although UG has been influential since its conception, many remain unconvinced and offer alternative theories. For instance, Boersma (2012) points out that although phonological segments can be classified in terms of feature bundles, their phonetic realizations may significantly vary cross-linguistically. According to Boersma, the phoneme category /u/ cannot be innately given because it has a different SR depending on the language, despite embodying the feature values [+back, +high, +rounded]. Although UG does not rule out the possibility of acquiring the language-specific phoneme /u/ from "formant clouds" which demarcate vowels and their closest neighbors into specific regions that can be illustrated by axes formed by the first and second formant[3], Boersma further points out that even though UG could make languages take on different shapes in terms of specific vowel formant boundaries, the perceptual boundaries are also language-specific. Thus, /u/ in one language may be perceived as its neighboring vowel /o/ in another language.

At odds here is the nativist view (the UG proponents) and the empiricist or emergentist view which implies that phonological categories are learnable from input, and language acquisition is simply utilized by a more general mental faculty instead of a specialized "language learning

[3] In acoustic phonetics, the first and second formants are resonant frequencies that correspond to specific vowels. The formant frequencies change in accordance with changing tongue and lip positions as different vowels are articulated. For instance, the "neighboring vowels" of /u/ are /o ɯ ʊ/, though whether these are classified as phonemes is specific to each language (to name one example, /ɯ/ and /ʊ/ are not phonemic in Finnish, but /u/ and /o/ are).

device" postulated by Chomsky (1965). For instance, Boersma (2012) suggests that infants individually construct their URs and acquire contrasts, long before the capacity to articulate them, directly from auditory speech signals, without requiring an innate endowment of phonemes and their features. The implications of UG and the early life constraining mechanisms for phonological L2 learning will be further discussed in section 2.2.3. The section to follow will present an overarching summary of the relevant theories that attempt to account for the underpinnings of phonological second language acquisition and the individual differences in its outcomes.

2.2 Acquisition Models

2.2.1 Negative Transfer

Phonological interference is the primary principle that causes foreign accents. It is thought to operate by way of transferring segments from the L1 phonology to that of the L2 and can impede speech comprehensibility in less proficient individuals. As such, it could be presumed that the degree of success in the acquisition of L2 phonologies is contingent on the degree of phonological similarity between the L1 and L2 systems. The view is accepted here that this is true, at least to some degree. Nevertheless, L1 transfer is a complex phenomenon and is not simply a case of importing wholesale, for instance, the parameter settings from the L1 grammar to be applied in the L2 grammar, but the patterns are rather unpredictable (see Westergaard 2021 for an overview). This has been demonstrated in numerous studies. For instance, White (1990) showed that French L2 English learners perform syntactic movement operations required by English declaratives and questions, which are different from French syntax, indicating that negative transfer does not always occur and what is being transferred are fine distinctions instead of simplistic analyses based on the L1 grammar. In the case of phonology, the items that are transferred are phonemes along with their features and phonological rules (for example, word-initial h-dropping, in keeping with the example of French learners). Conversely, while phonemes and their features may be acquired, not necessarily all the phonological rules governing their distribution are.

Since no language has the same segmental inventory, it may be expected that L2 learners have varying acquisition outcomes depending on their L1s, regardless of individual differences. Setting aside individual predispositions and aptitude for language learning, the acquired L2s

that have a greater number of matching phones in the L1 most likely result in more comprehensible, or "native sounding" L2 phonologies. Hypotheses may be created based on this principle. For instance, Swedish not only has more vowels, but also more consonant sounds than Finnish, which may place Swedes at a presumed advantage as English L2 learners in relation to Finns. Richer segmental inventories make it easier to assign a more diverse array of L1 speech sounds onto L2 speech sounds that are considered by the learner to match the L1 sounds. While L1-L2 phoneme and allophone mapping may not be perfect in that it may cause incorrect assignments (and thus a foreign accent), this is something that L2 learners appear to regularly opt for; learners are known to create analyses that result from converting L2 phonemes to the corresponding L1 phonemes, presumably because they are familiar and easy to produce (Obler 1982). This mapping is often presumed to be one-directional, but this should not be taken for granted as L2 speakers may in some cases develop what appears to be a foreign accent in the L1 after and during predominant L2 use (ibid.). Similar effects may be seen after immediate L2 to L1 switches in specific allophones (in the example provided by Obler, allophonic realizations of /r/).

The bilingual mind thus constructs complex parallel sound systems. Crucially however, the definition of a constructed parallel sound system, i.e., a segmental L2 inventory alongside the L1 inventory, is not simply one of production, but extends also to perception, and the production-perception systems do not appear to represent the native-foreign language system in equal measure. A study by Obler (1978) investigated balanced Hebrew-English bilinguals who were found to have perception and production systems governed by different principles. The participants' production showed "maximal differentiation" in the voice-onset time of voiced and voiceless plosive pairs /p b/, /t d/, and /k g/ in that the words with voiced plosives were closer to Hebrew monolingual controls and similarly the voiceless plosives were closer to English monolinguals. In other words, the bilinguals were tending toward the extremes, performing differently when compared to monolinguals. However, the same participants, when given perception tasks involving the same pairs, appeared to use parsing that was much more unified and in line with the performance of the monolingual control groups. For the bilingual, Obler concludes, "optimal processing is achieved by a primarily dual system for production, which exaggerates the discrepancies between the two languages, and by a broad, unitary system for perception, incorporating the extremes of the monolingual systems".

Negative transfer transcends the segmental level and extends to the prosodic dimension and the application of certain phonotactic rules. As will be shown, L1 speakers of certain languages

retain a high degree of loyalty to the syllable structure of their earliest acquired language, which is often observed in speakers whose L1s have strict rules regarding what syllables are allowed. This results in an apparent difficulty in realizing syllables that would not be permissible in the L1, an effect that subconsciously influences L2 pronunciation. For instance, Spanish L1 speakers insert epenthetic vowels in onset clusters of English words that begin with [s] because Spanish syllable onsets prohibit the segment in the initial position. Words such as *splash*, then, tend to be realized as [ɪsplæʃ] (Carlisle 1997).

Similarly, Arabic speakers tend to struggle to pronounce English words with complex onsets and complex codas in syllables (shapes such as CCV and VCC) as these are not permitted in many Arabic dialects, resulting in vowel insertions in, for instance, clusters with three consecutive consonants (Broselow 1983, 1988, 2018). These restrictions are arbitrary in the sense that they are language-specific. As mentioned previously, English does not permit sequences such as [ts] as onset clusters and therefore English speakers pronounce words such as *tsunami* as [sunɑmi], while Finns, for example, find realizing the [ts] in the onset unproblematic, a consequence of relatively lenient syllable structure rules in Finnish. The transfer effect at the syllabic level has been shown to also affect coda positions. For instance, Broselow and colleagues (1998) found L1 speakers of Mandarin, whose codas only allow a single nasal or a single approximant, to be unable to pronounce imaginary words such as *vig*. The Mandarin speakers instead often omitted the consonant or added a vowel following it. Taken together, these anecdotes indicate that not only segments, but also syllable onsets, codas, and possibly rhymes, may show interference from the L1, and must therefore be separately acquired by the learner, if possible.

In summary, the mental linguistic apparatus used to formulate utterances appears to rely on rules from the earliest language to which the individual was exposed, usually long before any conscious memories have formed. Accordingly, these rules become subconscious. The imprinting that language leaves on the mind is a process that begins in infancy and possibly before birth, which attests to how deeply ingrained the first language is. This will be discussed further in section 2.2.3 on the effects of age on language acquisition.

2.2.2 Interlanguage

Since L2 speaker production is a complex phenomenon that empirically cannot be said to rest on mere transfer of elements from the L1, additional theories are required to fill in the gaps. The theory of interlanguage (henceforth IL) accounts for the notion that L2 learner errors

cannot be explained by language transfer alone, supported by studies showing that child L2 learners make similar errors as monolingual L1 learners (Dulay & Burt 1972), and that L2 learners, both children and adults, from different L1 backgrounds make similar errors, largely underpredicted by L1 grammar (George 1972; Lance 1969; Dulay & Burt 1974). The term interlanguage was coined by Selinker (1969, 1972), though at the time there were others who described the same phenomenon. For instance, Nemser (1971) and Corder (1967) called IL an "approximative system" and "transitional competence," respectively. Conceptually, ILs resemble the phenomenon of children "creating their own grammars" in the process of L1 acquisition in the sense that L2 forms do not entirely adhere to target norms yet appear to be systematic in the IL (Selinker *et al*. 1975) and, as time goes by, the forms are re-adjusted to match target forms.

A crucial concept associated with ILs is that of fossilization, entailing that learners may not be able to acquire specific features of the L2 despite ample exposure to positive evidence, and in many cases even negative evidence does not seem to affect development. In other words, the non-target form persists despite overt explanation and instruction not to use it. It is hypothesized that children can circumvent fossilization more successfully as they are using the same cognitive power with which they acquired their L1s to acquire an L2, whereas by adulthood this ability will have largely atrophied (Tarone 2018).

IL may be thought of as generally being in a state of flux in contrast to the L1, which is static in the sense that it has a developmental endpoint after which no more learning occurs, at least in a phonological sense. In addition to language transfer, Selinker (1972) describes four "central processes" of the L2 learner, expanding the acquisition process to include such strategies as overgeneralization, transfer-of-training, strategies of second-language learning and second-language communication. Some of these processes, such as overgeneralization, are well-known and rather ubiquitous among L2 learners. Errors related to overgeneralization may be thought to result from reasoning based on applying already learned L2 rules inappropriately, rather than the transfer of rules from the L1 system. Likewise, in lieu of the L1 system being the sole source of transfer, some elements may transfer from a particular method of training or specificities of the contexts in which the L2 speaker communicates with native speakers.

Tarone (1979) was the first to suggest that IL is not a singular system, but may contain multiple sets (e.g., IL1, IL2, IL3, and so on) which depend on context of use, citing systematically variable performance, both phonological and syntactic, depending on the language task in

question. In this view, the IL consists of a capability continuum with various styles with varying grammatical accuracy. Depending on context, certain linguistic features are supplied either all the time or, perhaps, half of the time (Tarone 1983). In Tarone's model, accuracy varies in accordance with how much speakers pay attention to language form, predicting, paradoxically, that spontaneous and unmonitored speech styles are less influenced by the L1 than careful and monitored styles.

Due to the "chameleon-like" character of IL, Tarone (2018) suggests that researchers should account for contextual factors involved in each elicitation of research subjects due to the systematic variability contingent on, for instance, social context, task, topic, focus on form, and interlocutor. Such a perspective is fairly radical, as most linguistic research does not take performance variability into account. The prevailing view implies that the mental grammar that is assumed to reside inside a study subject's mind is always there, yet the grammar is reflected inconsistently in production. Indeed, it is not far-fetched to suggest that, for instance, subjects' performance in elicitation tasks will be affected by how nervous they are.

2.2.3 Age Effects and Maturational Constraints

Individuals with earlier L2 acquisition onsets in childhood tend to perform closer to native pronunciation norms. However, late L2 learners seem to rarely approximate L1 speaker proficiency, a finding which is considered particularly true of phonological acquisition. For instance, Scovel (1988: 62) argues that pronunciation specifically is difficult to master by L2 learners because every other aspect of language, such as syntactic, morphological, and semantic knowledge, is purely mental while pronunciation has a "physical reality" in the sense that it requires timing and sequencing of motor movements and subtle control of the articulators. In contrast, other aspects of grammar only have a "cognitive reality" whereas phonology involves connecting mental knowledge regarding how words are pronounced to physiological speech organs.

The Critical Period Hypothesis, popularized by Lenneberg (1967), is thought to account for why L1 speakers always reach a degree of proficiency that L2 speakers (in other words late learners) do not. The attainment of native proficiency appears to be mediated by a critical period (henceforth CP), a window of time during which language must be acquired by the child somewhere in early childhood. In total absence of language exposure, the child fails to reach L1 proficiency, which is then impossible to attain afterwards, once the CP is over. Evidence for this comes from studies on feral children (Curtiss 1977) and acquisition outcomes in sign

language in the congenitally deaf (Newport 1990). However, the CP is also widely believed to be the reason for L2 speakers' retaining foreign accents (Flege 1999).

Long (1990) reviewed numerous studies investigating L2 acquisition starting age and acquisition outcomes, concluding that L2 speech is accent-free if the onset of acquisition is between ages six and twelve. Regarding the effects of age on L2 learning more broadly, adults and older children tend to have an initial advantage over younger children, but only in specific grammatical areas, namely morphology and syntax. However, learners with earlier age of onset at acquisition (henceforth AOA) tend to reach higher ultimate proficiency than those with later AOAs. The age advantage appears particularly to be a case of phonology in which earlier starting ages lead to better outcomes, though non-phonological L2 acquisition such as syntax and morphology also appears to be influenced by AOA (e.g., Johnson & Newport 1989; DeKeyser 2000).

The finding that older children and adults are at an acquisition disadvantage has been explained by the CP, a time window within which the brain is better equipped to learn a language to such a degree that it becomes an L1. Children raised in multilingual homes may acquire several L1s. Scholars are equivocal regarding the exact neuropsychological mechanisms behind the CP and the exact age of the period's closure. Some believe that, once the age effects begin to fade, learners recruit other, non-linguistic cognitive processes as a compensatory mechanism (Bialystok & Hakuta 1999: 172). One of the manifestations of such compensation would be L1 transfer whereby L2 speakers map L2 forms onto grammatical processes that also exist in the L1 (see 2.2.1). The "Less is More" hypothesis proposed by Newport (1988) suggests that the more limited abilities of children may provide, paradoxically, an advantage for language acquisition and other complex tasks that involve componential analysis. She suggests that complex linguistic stimuli are more readily perceived and remembered as wholes by adults whereas children, due to their limited cognitive capacities, perceive and store component parts of the whole complex stimulus, resulting in a better position to locate the components.

The "Less is More" hypothesis may lend itself particularly to the acquisition of morphemes, in which Newport was specifically interested in the case of ASL, a sign language with complex morphology whereby handshapes and motions indicate morphemes. How the "superior componential analysis" used by children fits into a phonological model is unclear. In fact, Menn (1981) shows evidence of children acquiring their L1s by initially learning words as "whole phonological units", only to decompose them later into individual sounds. Interestingly,

however, infants lose their initial ability to distinguish non-contrastive sounds of the ambient language during the first year of life, an effect which carries onto adulthood (Obler & Gjerlow 1999: 125). To illustrate, Japanese adult monolinguals do not perceive the difference between /r/ and /l/ in English because such a contrast does not exist in Japanese (Buchwald *et al*. 1994; Flege 1988: 327-334).

Further, exceptional individuals appear to not be affected by late starting ages, being sometimes capable of achieving native-like phonology in the L2 (see 2.2.4). Such feats are exceedingly rare, especially in the domain of phonology, and should be unlikely, considering what is known about infant L1 acquisition. Indeed, infants younger than 6 months can initially discriminate all non-native contrasts (e.g., Eimas 1975; Streeter 1976; Trehub 1976), but begin losing the ability to distinguish non-native consonants at approximately 10 months of age (Werker & Tees 1984), and their ability to perceive non-native vocalic contrasts declines even earlier (Kuhl *et al*. 1992). This implies that the brain is relatively closed off to non-native linguistic data starting from an early age, singling out the ambient L1s for efficient acquisition, and rendering the native acquisition of L2s later an unrealistic task. According to Scovel (1988: 181), adult learners who manage to acquire native-like phonology in the L2 are thus "super-exceptional", comprising "about 1 out of 1,000" of any population.

The definitive reasons for varying outcomes in phonological L2 acquisition remain unanswered and the opinions of experts remain equivocal regarding the extent to which various factors, including CP, influence ultimate attainment. Moreover, explanations alternative to CP have been offered. For instance, Bongaerts (1999) identified three main factors contributing to successful native-like acquisition in late L2 learners, namely "high motivation, continued access to massive L2 input, and intensive training in the perception and production of L2 speech sounds", effectively arguing against maturational constraints, or rather, the rigidity of such constraints. Similarly, Flege (1992a, 1992b, 1995) argues that L2 speakers tend to perceive L2 sounds through a filter of already established L1 categories that best correspond to L2 sounds. Flege suggests that learners are not inevitably limited by their L1 categories but may be able to eventually differentiate L2 sounds into categories discrete from L1 categories. According to the view advanced by Flege, whether such differentiation is successful or not depends on the late L2 learners' ability to activate a continuous mode of perception, as opposed to a categorical mode of perception. Young children are said to heavily rely on the continuous mode, in which speech sounds are distinguished by paying attention to minute differences, such as phonetic contexts and realizations by different speakers. However, children will have formed stable L1

phonemic categories by the time they are seven years of age, after which the continuous mode changes into the categorical mode, in which perception is based on contrasts between different categories, such as /b/ and /d/ (Bongaerts 1999; Flege 1992a, 1992b, 1995).

Regardless of the ambiguity and differing views regarding the effects of age, some learners, rare as they may be, manage to evade age-related limitations. The difficulty of accounting for the success of these individuals lies in their scarcity in any given population. However, some research has been carried out on exceptional learners, which will be briefly explored in the next section, as these individuals are important to identify when carrying out investigations on highly advanced L2 speakers.

2.2.4 Language Aptitude and Talented L2 Acquisition

Carroll (1981: 86) defines language aptitude as "an initial state of readiness and capacity for learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity". Carroll developed the Modern Language Aptitude Test, or MLAT, (Carroll & Sapon 1959) which has been the standard of aptitude testing in the literature since its conception. Others have since proposed variants of the Carrollian view of aptitude (see Li 2019 for an overview).

Some adults are exceptionally gifted at L2 learning, but whether this is an inborn trait or a result of conscious learning strategies that lead to more successful acquisition and better retention of the learned language is contested. For instance, Cook (2016: 141-144) places more importance on conscious learning strategies, such as the "Good Language Learner Strategies" compiled by Naiman and colleagues (1995), and other strategies based on cognitive psychology (see O'Malley & Chamot 1990), implying that these form the basis of successful acquisition outcomes in adults. In Cook's view, whatever inborn advantages there may be, conscious strategies may be used as compensation by those who do not have them, effectively minimizing the role that a possible inborn aptitude could have in L2 acquisition. However, work by Skehan (1986) suggests that L1 development in early childhood, as measured by Mean Length of Utterance (MLU), correlates with L2 learning success later in life. This suggests that inborn endowments that facilitate L2 learning after the critical period could be inherited traits. While some believe that L2 learners are simply utilizing the remnants of the aptitude that is used to acquire L1s (e.g., Ganschow & Sparks 1996: 201), the cases of exceptional learners complicate the issue.

Attainment of "native-like" phonology is rare for L2 speakers and is commonly understood to be the most difficult domain to master in L2 acquisition. As such, aptitude differences are likely to be most obviously felt in L2 speaker phonology, at least in the case of advanced learners and experienced L2 speakers. The prevalence of individuals who reach native-like competence post-adolescence has been estimated to range from 5% to 15% in the population (Seliger *et al*. 1975; Birdsong 1999, 2005). However, when discussing exceptionally successful L2 acquisition, phonology is particularly relevant, since many advanced learners have excellent, or native-like, command of lexical semantics, syntax, and morphology, yet retain phonological categories with fossilized L1 phonemes in place of target-like phonemes. It is apparent from highly proficient individuals that phonology tends to be the aspect of grammar that is most notably divergent from target, and in which variation between individuals can be perceived most often. This raises the question of what factors distinguish the few individuals who master second language phonological systems from the rest.

Language aptitude is what adults draw on to learn a second language while children primarily rely on the critical period, the age advantage. It has been found that adult learners will in almost every case retain a foreign accent of varying degrees, and that aptitude as an indicator of acquisition success varies significantly within the adult population (Obler & Gjerlow 1999: 133). For instance, Abrahamsson & Hyltenstam (2008) showed that only those with high aptitude scores and AOAs, at adolescence or after, were rated as "near-native" by native-speaker judges. The same was not found to be true in early L2 learners, implying that to pass for a native speaker in everyday language use, a high degree of aptitude is required for adult learners but not for child learners (ibid.).

What kind of factors, if not metacognitive learning strategies, underpin higher aptitude in some individuals remains unclear. Closely linked with high aptitude is the concept of talented second-language acquisition (Novoa *et al*. 1988), though these are generally considered discrete concepts. Novoa and colleagues (ibid.) offer a possible explanation underlying talented L2 acquisition as stemming from gestational events in which sex hormones shape the brain and, through neuroendocrinological and immunological factors, may endow individuals with specific cognitive abilities, such as being exceptionally talented in foreign language learning.

Talented L2 learners often have seemingly unrelated characteristics in common, such as left-handedness, twinning, homosexuality and being delayed in reading (Novoa *et al*. 1988; Obler

& Gjerlow 1999: 134), as is predicted by the Geschwind-Galaburda hypothesis[4] (Geschwind & Galaburda 1985a, 1985b, 1985c). Although this hypothesis is controversial, traits such as left-handedness have been specifically linked to aberrant language organization of the brain in other studies, lending the gestational hypothesis some credence. For instance, left-handed individuals have stronger functional connectivity between right and left language networks, owing to abnormal hemispheric lateralization in which language ability is not specialized only to the left hemisphere as it is in most of the right-handed population (Wiberg *et al*. 2019).

Furthermore, Schneiderman & Desmarais (1988) suggest that talented L2 learners "possess additional right hemisphere-based flexibility for language processing" and "are utilizing a proportionally greater amount of cortex for language functioning than untalented language learners", differentiating them from the left-lateralized majority of the population. They also cite an unpublished study in which the Geschwind-Galaburda hypothesis was tested, discovering that the so-called Geschwind cluster factors were present more often in the most talented, but also the least talented, L2 learners than the average L2 learners.

Regardless of whether exceptionally talented L2 learners or individuals with above average aptitude emerge due to prenatal circumstances, it is generally held that the ability is mostly independent of general intelligence (Obler & Gjerlow 1999: 135). Since it is not associated with psychometric measurements such as IQ, and since considerations about gestational events are mostly ignored in language research, language aptitude remains the construct which scholars rely upon to explain outcome variation in adult L2 acquisition. Research on language aptitude has accumulated from the past six decades, yielding various theories on its nature. Li (2019) concludes that aptitude is a "domain-specific language learning device", making it different from domain-general processes, such as working memory, that have also been found to facilitate L2 learning (see section 2.3).

2.2.5 The Monitor

The Monitor Hypothesis (Krashen 1981, 1982, 1985; Krashen & Terrell 1983) asserts that, in communicative situations, the L2 speaker modifies the output of the acquired second language

[4] They suggest that certain crucial hormonal events during gestation may cause overdevelopment in some brain regions, while other regions develop poorly. Such development leads to anomalous or unusual organization of cell structures in the brain which may endow individuals with various talents, such as exceptional giftedness in music or linguistic abilities, in addition to exceptionally poor abilities which may or may not be present. One such example is dyslexia, also studied by Geschwind and Galaburda, which they found to be caused by a failure in neuronal migration, leading nerve cells in key brain regions to mature improperly. These regions, in the case of dyslexia, are associated with visual processing of language. (Schneiderman & Desmarais 1988).

based on conscious application of grammatical rules. The Monitor may be thought of as a real-time mental editor of speech, consciously manipulating L2 production. Stephen Krashen introduced the concept as part of the Monitor Model, which contains five basic assumptions regarding language acquisition, one of them being the Monitor. The other four hypotheses will not be discussed here as they are more relevant for second language teaching than second language speech production.

Krashen identifies different types of Monitor users which may be thought of as various mental approaches to performing in communicative situations: over-users, under-users, and optimal users (Krashen 1978, 1982; Stafford & Covitt 1978; Kounin & Krashen 1978). Monitor over-users are self-conscious about their speech to such an extent that it disturbs fluency as the speaker is constantly monitoring their output with consciously internalized knowledge of L2 grammatical rules. Under-users are the opposite in that they do not monitor their production with conscious knowledge but rely entirely on subconscious understanding of the L2 grammar. The optimal user is capable of regulating the use of the Monitor, applying it at appropriate moments so that it does not interfere with communication, as in the case of over-users.

The Monitor framework is not without problems as it contains the controversial notion of an acquired-learned dichotomy, which soon came under heavy criticism for being oversimplified, underdeveloped, and unfalsifiable. However, despite the valid criticisms (McLaughlin 1987: 26-28; Gregg 1984), it is conceptually useful. Importantly, Krashen suggests that personality plays a significant role in how much speakers monitor their output. Indeed, the idea that some speakers are more self-conscious about errors than others is not controversial nor far-fetched. Personality differences could be expected to be seen most clearly in communicative situations with oral production and the question of how personality and the psychological characteristics it entails affect speech production warrants investigation. Indeed, psychological factors become relevant especially in oral production since L2 speakers may be particularly self-conscious about their pronunciation. Because of the phonological dimension, the Monitor, or related concepts, should be afforded some consideration in investigations involving L2 speech performance.

Aside from personality differences, the different types of Monitor users described by Krashen, it could be speculated, may also shift within one individual depending on the context in which the L2 is used. For instance, the type of situation, such as formality, or the type of interlocutor may cause the Monitor to be used differently. Furthermore, how this dovetails with Tarone's

(1979, 1983) account of multiple interlanguages within one individual is a question that warrants further exploration.

2.3 Working Memory

Working memory (WM) was first proposed by Baddeley & Hitch (1974) as an additional mental construct to short-term and long-term memory. At present it is thought to underlie not only language acquisition, but also real-time language production. Real-time language production may be defined as the capacity to rapidly utilize mental linguistic resources at the speaker's disposal to produce coherent speech. WM is involved both in production and comprehension of language (Hawkins 2019: 287-288). Miyake & Shah (1999: 450) define WM as comprising "mechanisms and processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition". This definition is largely agreed upon by authors who have proposed varying models of WM. In Baddeley's (2000, 2007) multi-component WM model, the WM consists of three subsidiary systems regulated by a supervisory system, the central executive: the visuospatial sketchpad, episodic buffer, and the phonological short-term memory (also referred to as the phonological loop). Higher processing capacity of the phonological loop has been shown to correlate positively with L2 acquisition success in various grammatical areas (see Wen 2016), though language aptitude has been shown to be a stronger predictor of language learning than WM (Li 2017).

2.3.1 Optionality and Language Processing Limitations

WM has limited capacity for handling incoming and outgoing information (commonly referred to as processing). In a typical conversation involving speech by two or more participants, WM processes what is formulated in the mind and realized as output, the speech of self, and what is parsed from the input, the speech of others. Due to processing limitations, incoming information may be lost, or grammatical information may fail to be realized in the output, resulting in ungrammatical surface forms. Assuming that the representation is underlyingly correct, ungrammatical speech is a matter of performance errors, unrelated to speaker competence. Indeed, such linguistic breakdowns may also occur in L1 speech.

As regards L2 representations and optionality in the surface forms of learners, whether the underlying representation is impaired or whether the correct grammatical analysis exists but

fails to be realized has been subject to debate. For instance, the Impaired Representation Hypothesis (IRH) claims that inflection is "essentially impaired" owing to a lack of functional categories, features, or feature strength and remains impaired until there is no longer optionality (Eubank 1993/94; Meisel 1997). Alternative views posit that optionality in L2 speech exists due to cognitive demands unrelated to underlying competence. For instance, the attention required by the content of messages on the one hand, and linguistic resources on the other, may exceed the WM's processing capacity (Hawkins 2019: 288). In practice, this means that some grammatical features, such as tense morphology on verbs, may be temporarily inaccessible, though it has been acquired by the learner. In a less demanding speech situation, however, such a feature would be accessible. According to Sharwood Smith & Truscott (2014), L1 and L2 features are not equally accessible to WM, contingent largely on the frequency of their activation. In this sense, the features are "competing" and gradually the grammatical features in the L2 lose their optionality in speech production.

Two accounts that have gained widespread influence and attempt to explain optionality in the use of correct grammatical forms as occurring due to WM processing limitations will be discussed here. These are the Morphological Deficit Hypothesis and the Missing Surface Inflection Hypothesis. The two hypotheses diverge in some respects, such as in the assumption of how morphemes and lexical items are stored in the mental vocabulary, yet both assume that underlying L2 speaker competence may not always be evident from the output.

In proposing the Missing Surface Inflection Hypothesis (MSIH), Haznedar & Schwartz (1997) and Prévost & White (2000) argue that L2 learners, by the time their mental grammars have acquired abstract morphemes, may nevertheless be unable to assign the correct inflection onto the appropriate structures, despite correctly representing tense and agreement. For instance, learners must attach the morpheme *-s* to *speak* to produce *speaks* when the tense category contains the values [finite, non-past, third person, singular]. While *speaks* is specified for these tense values, *speak* has fewer specifications (e.g., person, and number), but occurs in more finite contexts. The key assumption is that learners often have the correct representation of the inflection of verbs such as *speak*, but fail to realize the correct morphology due to processing reasons or communicative pressure. Such temporary access failure results in the selection of underspecified forms in contexts that require specified forms. As a result, L2 speech exhibits overuse of verbs without affixes and insertion of non-finite verbs into syntactic nodes with a [+finite] specification. The MSIH assumes that the syntactic level is correctly represented, even though the inflection of specific lexical items is not, as indicated by the performance of L2

English learners who acquire target-like syntactic properties such as verb movement and Case, contingent on the acquisition of higher functional projections, yet frequently omit past *-ed* and 3rd person singular *-s* (McCarthy 2008).

Work by McCarthy (2008), which resulted in the Morphological Deficit Hypothesis (MDH), attempted to test the assertion that communication pressures on output give rise to optionality in production (Prévost & White 2000: 129). With the assumption that processing demands are mostly manifested in production, she found optionality in comprehension tests in addition to production tests, a finding "particularly problematic for the MSIH". She concludes that optionality is "at least in part a representational issue and does not derive strictly from production-based limitations" and that any possible representational deficits exist in the morphological rather than the syntactic domain. While MSIH assumes that productive use of correctly represented functional morphemes becomes optional due to WM-related reasons, in MDH it is assumed that optionality arises from a partially impaired representation of abstract morphemes. At any rate, both views hold that the most frequent forms that have fewer affixes are accessed when the WM capacity is burdened by demands of communication.

According to McCarthy (2008), morphological competence "need not be considered a different kind of knowledge than syntactic or phonological knowledge" since these domains "consist of organized, structured representations, and L2 research can aim to discover how the competence that consists of these representations is acquired". In the phonological domain, words with target-like phonetic features that are productive, yet optional, could be considered the equivalent of underspecified verb forms in the hypotheses discussed above. The optionality in the phonetic makeup of a word would be a matter of alternation between the target segment and the segment that matches the closest equivalent in the L2 speaker's L1. The capacity of the WM determines how reliably the target segment is accessed, and when communication pressure and possibly other psychological factors overburden that capacity, the non-target L1 segment is accessed and produced instead.

It is plausible that similar computational underpinnings that were presented in the frameworks of MDH and MSIH could account for the optional use of phonological features. For instance, consonants with the feature [+voice] are rare in Finnish phonology, the L1 speakers of which may be expected to optionally use, for instance, [k] in place of [g] in words such as *group*. Both [k] and [g] have the same place and manner of articulation (namely, they are velar plosives) and differ only in the feature [voice]. It will be proposed in this thesis that this feature, for

example, is correctly represented, but cannot be retrieved by the L1 Finnish speaker if the WM capacity is exceeded by other demands, such as mental planning or accessing the appropriate lexical items and morphemes. The closest equivalent, the voiceless velar plosive, in this instance, is transferred from the L1, and its voiced counterpart is the target segment in words such as *group*. Evidence of non-target production occurring due to processing limitations could be obtained by observing L2 speaker performance in situations where the WM may be expected to be burdened (by communicative pressure, for instance), and in situations where production is not accompanied by significant cognitive demands.

In summary, the limitations that WM places on speech are based on the principle that, during on-line production, some information that exists in the speaker's mental grammar is not always made available. The principles with which the linguistic information-to-WM-to-speech output mechanism operates appear to sacrifice certain grammatical properties, allowing some to take precedence over others. This was illustrated in this section by MSIH and MDH with examples of learners defaulting to forms without morphemes that indicate tense. L2 phonological production often exhibits negative transfer, which many observers would attribute to incorrectly represented phonemes. However, if there is any optionality in the realization of phonemes, it is reasonable to assume that that some target features are inaccessible due to external factors related to the communicative situation overburdening the WM. Whether these features (e.g., allophones in specific phonological categories) could be consistently accessed in so-called off-line production, or speech that occurs in relatively short segments and allows for planning is an interesting research idea. Thus far, grammatical optionality present in surface forms and its link to processing limitations has been explored mostly in the domains of morphology and syntax, particularly in L2 research.

3 Methods

3.1 Research Questions

The research questions for the present study are as follows:

Research Question 1: Does the nature of the communicative situation affect phonological production, giving rise to inaccurate representation of L2 phonological proficiency?

Research Question 2: When production is elicited in an experimental situation where the production is not part of a communicative situation, does this affect phonological output in comparison with a communicative situation?

Psycholinguistics literature, particularly that regarding second language acquisition, discussed in chapters 1 and 2, provides the impetus for these questions. This chapter details the procedure that was undertaken to answer them.

3.2 Participants

The participants to the study were recruited from a group among university students enrolled to pedagogical studies and aiming for a subject teacher qualification. A major part of the training program, organized by the faculty of education and culture at Tampere university, involves teacher training in which the students acquire hands-on experience by teaching a series of lessons devised by themselves in their respective subjects. For this study, only the students completing their teacher training as English subject teachers were a target for recruitment. The recruitment campaign was facilitated by teachers responsible for the teacher training, resulting in a total of four participants, from which two were later excluded due to thesis-related limitations. In the English teacher training provided by Tampere university, it is customary to use English as the language of communication during practice lessons, fulfilling the criteria for situations involving organic use of L2 English by competent English speakers. The participants were recruited in Autumn, the beginning of the academic year, immediately before the teacher training was slated to begin.

Aspiring English teachers are an ideal pool of recruitment for multiple reasons. A classroom teaching scenario involves some degree of communicative pressure, i.e., teachers are likely to sacrifice some of their grammatical accuracy in order to communicate the content of their message efficiently to the students. This prioritization of content over grammar is also likely to cause teachers to pay less attention to the well-formedness of their speech due to the need to instruct students on the one hand and to verbally react to what is happening in the classroom or what the students are saying on the other. Accordingly, the classroom teaching scenario creates the circumstances for the types of communicative situations that involve communicative pressure as teachers must form sentences in English quickly, and there is usually minimal time available for mental planning ahead of utterances.

Inherent to classroom teaching is performing in front of an audience, a type of schema which often evokes a well-known psychological phenomenon known as stage fright, a type of performance anxiety during which the affected individual is under heightened stress. Being new to teaching is likely to exacerbate this effect. The stress experienced by the individual restricts the availability of working memory, a concept thought to underpin language production and formulation of longer utterances. Finally, using trainee teachers as participants also eliminates the need to vet for participant competence as the students in teacher training have already completed university-administered tests to prove their competence in their subjects of expertise.

All participants were informed that the goal of the study was to investigate the use of L2 English in different contexts. The participants were not initially informed about the exact details of what was being investigated about their use of English, an important notion considering this might have made them more conscious about their pronunciation in the classroom.

Analysis of data from two participants were included in the final study and their background information were collected in an interview (see 3.7 for more details). The two participants are distinguished from each other by numbers 1 and 2 and they are always referred to in this thesis as “participant 1” and “participant 2”. The most relevant background information is summarized in table 1.

Table 1. Summary of participants' background information

	Age	AOA	Years of study	Reading (hr)	Speaking (hr)	Listening (hr)
Participant 1	23	9	14	5-10	0-5	10-15
Participant 2	25	9	15	10-15	5-10	15-25

Age refers to the age at the time of testing. AOA stands for age of onset at acquisition. The columns for reading, speaking, and listening are the participants' self-reported estimates of weekly engagement with English.

3.3 Classroom Recordings

Approximately five hours of classroom English teaching by each participant was recorded with an audio recording device. Uncompressed audio data was transferred onto computer in the form of Waveform Audio File Format (WAV-files) for repeated and careful listening. Of importance in the recordings were considered recurrent L2 segments that were considered not to be target-like in words and phrases in which the participants produced them. In the audio data spanning multiple hours in duration, much of the captured speech was unusable due to either background noise or speakers' distance from the microphone. Accordingly, only the audio data considered to be of sufficient quality was analyzed for non-target tokens occurring on a segmental level.

The selected tokens were isolated and extracted using freely available audio editing programs and provided to a native English speaker and trained linguist for evaluation and possible disagreements regarding the non-nativeness of the selected tokens. Multiple instances of the same token were provided for evaluation, and the tests for production and perception in the experimental setting were tailored for each participant individually based on the tokens unique to each participant.

Classroom recordings were captured with a Zoom H4n Pro recorder. In order to ensure a maximal signal/noise ratio, the device was placed on a table in the anterior part of the classroom in a directional microphone setting, predicting the teacher's physical location and relative direction from the microphone's direction of capture, for the majority of the lessons' duration. A static recorder was used in lieu of head-mounted or lavalier microphones to minimize invasiveness in addition to making it easier for the participants to forget that they were being recorded, which otherwise might have affected speech performance. In order to keep a written record warranted by a fieldwork investigation, the classroom recordings were accompanied by

a rough transcription of timestamped words and phrases in which non-target phonology could be perceptible.

Following the classroom observation phase, each participant was invited to the phonetics laboratory located at the University of Tampere for a series of tasks involving language production and listening. The participants were additionally interviewed for more details regarding pronunciation strategies and general use of the L2.

3.4 Production Task

The production tasks involved reading out loud phrases and words on a list unique to each participant. The lists contained tokens which the participants had been heard producing in the data from the classroom audio recordings. The tokens were typically embedded in short carrier sentences that were separated by a line break, making it easier to make the appropriate pauses during the reading task. The tokens in the list were dispersed by the addition of arbitrary words and sentences that did not contain segments relevant to the study.

The words and phrases were printed on a sheet of A4 paper which had been placed on a folding music stand adjusted to the reader's eye level. The participants read from their lists while standing in front of a microphone stand adjusted, alongside the music stand, to the participant's height. The microphone used for audio capture in the production task was Audio-Technica AT2035 cardioid condenser. The microphone was placed at a 30 cm distance from the speaker with a pop filter in between capture device and speaker.

To obtain appropriate results, the purpose of the research design was explicitly stated to the participants. It was made clear that the produced oral data was intended to reflect the participants' genuine competence and that the experimental situation was designed to be one of no communicative pressure nor psychological stress. It was paramount to the study and the hypothesis that the participants did not feel nervous about the experiment or their performance during production. As such, the participants were exhorted to attempt to allay any possible anxiety related to either the performance or the situation. The participants were carefully instructed to use pronunciation they felt was most accurate according to their internal grammars, i.e., the aim was that the pronunciation be "natural".

As a pre-emptive measure, the tokens in the reading list were often repeated either in the same word, or as part of another word in case of inadvertent mispronunciations. In the case of some

tokens, such as those that tend to occur in connected speech, this would have been necessary either way. The participants were also encouraged to repeat a word or phrase multiple times in case they would hear themselves produce segments that were, in their understanding, not correct.

3.5 Receptive Task

To gauge competence adequately and tease it apart from potential performance-related distortions as reliably as possible, it is necessary to test performance aurally in addition to oral elicitations. For this purpose, the participants completed a receptive task on a computer following the productive task. For the receptive task, two versions of the words and phrases elicited in the productive tasks had been recorded to be listened by the participant: one version bearing the accurate and thus expected target form and the other bearing the non-target form which the participants themselves had produced in the classroom.

The participants were asked to listen to a list of words and phrases which were sorted into numbered folders with two variations in discrete audio files and then decide which version sounded more correct. The audio files for a particular word or phrase were labeled “a” and “b” and after listening to both alternatives the participants circled or underlined on a sheet of paper the corresponding letter, under the corresponding folder number, indicating the audio file with the more correct sounding pronunciation. The two versions for each word differed only in one segment and the participants were instructed to only pay attention to the segmental level instead of suprasegmental aspects, such as intonation or amplitude. This also applied to recordings containing more than one word, i.e., a phrase.

The participants were encouraged to listen to both versions of each recording multiple times in case that the difference was not clear. It was also made clear that in each folder there necessarily was a difference between the two versions and that the differences were every time a matter of a single segment. Headphones were used in the task to enhance hearing segmental nuances. Adjusting the volume of the computer to which the headphones were plugged was recommended as deemed fit by the participants. The option to select both versions of a particular word or phrase as either correct or incorrect was permitted. In unclear cases, in which a difference was not heard, both versions sounded correct, or both versions sounded incorrect, the participants were encouraged to write the interpretation separately, in words, under the

corresponding number in the answer sheet. The results of the receptive tasks were compared to the findings from the productive tasks.

3.6 Acoustic Analysis

In order to validate the observations from speech events beyond intuition, illustrating the findings with instrumental techniques is in order. The tokens recorded in the production tasks were thus analyzed using sound spectrography. A speech analysis software for phonetics, Praat, was selected for this purpose.

Spectrographs plot frequencies in Hertz along the vertical axis and in milliseconds along the horizontal axis. The frequencies represent bands of energy, commonly known as formants, that correspond with the shape of resonant properties of the articulatory cavities of the sound source (in this case, a human subject). Each speech sound leaves a characteristic pattern on the spectrograph. For instance, all vowels will display a voicing bar at the f_0 level in the region of 100-200 Hz, though this is also influenced by individual characteristics of the speaker. Nevertheless, despite individual differences in the acoustic makeup of the signal, speech sounds are distinguishable from one another by their core features in the spectral signature. Vowels differ from each other by tongue position and lip posture, which are reflected in formant frequencies, primarily in the values of f_1 and f_2. For instance, the vertical position of the tongue attributed to a specific vowel is inversely proportional to the value of f_1, meaning that vowels such as /i/ have lower values. Front vowels on the other hand have higher f_2 values than back vowels, while rounded vowels have lower overall frequencies across formants. (Ball & Rahilly 1999, 162-169.)

As with vowels, consonants also have clear acoustic counterparts on the spectrogram as defined by the features which comprise them. For instance, obstruents have a closure-release pattern, and the place of articulation may also be gleamed from the spectrum. Plosive sounds, a subcategory of obstruents, exhibit a characteristic voiceless closure phase of complete silence. Sonorants on the other hand tend to resemble vowels due to the voicing bar (a diagnostic of vocal fold vibration) which accompanies them. In summary, speech sounds, once the audio data and the diagnostic instrument have been secured, can be clearly distinguished.

3.7 Interview

Due to the small sample in the present study, a semi-structured interview was deemed the most appropriate method for gathering information about the participants' speech strategies (see 3.7.2) and ideological influences (see 3.7.1). The semi-structured interview falls on the direct end of the methodological spectrum and the direct approaches may be summarized as data collection procedures relating to oral or written data (Garrett *et al*. 2003). Interviews, as opposed to questionnaires, produce ample information that is more detailed and can be further elaborated upon (e.g., Wrench 2008).

In order to account for the wide array of possible factors contributing to inter-subject phonological variation, participants were interviewed about the details of the processes behind their L2 speech production and their own beliefs regarding how they sound, or should sound, phonologically. These details, based on answers to open-ended questions, were coded into themes. The questions included production strategies extrapolated from Krashen's Monitor model, general attitudes towards pronunciation and non-target phonology, speaker background and possible inhibitions to sounding like a native speaker of the L2. The purpose of the chosen topics was to obtain further information about the processes behind the participants' language production, as this information could elucidate phonological behaviors. Since the participants were few in number, it was possible to obtain information that was more ample and detailed than it would have been with a larger sample. Due to the open-ended nature of the questions, a thematic analysis was conducted on the interview responses.

3.7.1 Theme 1: Speaker Identity and Attitudes on Pronunciation

The participants' attitudes on foreign accents in both their own pronunciation, and that of English teachers in general, were a target of inquiry in order to account for possible misrepresentation of genuine underlying competence. For instance, many language teachers may be influenced by certain pedagogical perspectives that do not place much value in adherence to native speaker norms (e.g., Seidlhofer 2011), and therefore this type of thinking could have an influence on the teacher's L2 phonology. To name an example, the ELF perspective, advocated by Barbara Seidlhofer among others, promotes the idea that no L2 learner should have L1 proficiency as an end-goal. Recently, the ELF ideology appears to be popular among university-educated language teachers in Finland and should thus be taken into consideration, since focusing solely on achieving "comprehensible English", as L2 learners, may cause them to significantly fall short of target-like pronunciation.

Moreover, L2 speakers may hold beliefs and attitudes that cause them to sound, intentionally, less phonologically proficient. Some L2 speakers may feel that sounding like a native speaker feels "wrong" or "pretentious". Age, too, may be implicated in such cases: adults may be expected to have a stronger sense of self and a more established identity, at the center of which is the language they speak, which in turn may set constraints on acquisition (Schumann 1975). Children may thus be more open-minded and therefore more readily strive for a fully native-like L2 phonology, whereas adult immigrants, for instance, face fears of losing their cultural-personal identities (Christophersen 1973; Snow & Hoefnagel-Höhle 1977). Even in the case of individuals with early AOAs, there is no reason to assume that possible identity-related constraints are not taking effect on pronunciation.

3.7.2 Theme 2: Pronunciation Strategies and Performance Anxiety

Speakers rely on various strategies (see 2.2.5) to monitor their output. Some speakers are more self-conscious when producing utterances than others, which might also affect pronunciation. What strategies each participant used was therefore a relevant question upon which a broader theme was constructed. L2 speech performance in various situations and possible anxiety accompanied with them was merged under the same theme. Regarding pronunciation strategies, the participants were inquired about their tendency to monitor their own pronunciation. The questions regarding performance anxiety concerned the practice lessons taught by both participants and the degrees of psychological stress accompanied by the experience.

3.7.3 Theme 3: Speaker Background

As discussed in chapter 2, speaker background appears to have the most significant impact as to the phonological proficiency that is ultimately reached. AOA and daily habits relating to the use of English were thus considered important questions for the participants since these are important factors in individual variation. Additionally included under the third theme were habits pertaining to daily and weekly use of English, namely the amount of time the participants spend using English in their daily lives by way of speaking, reading, and listening. These are summarized in table 1 in section 3.2.

Included in the speaker background theme was a series of questions regarding individual traits[5] pertaining to talented L2 acquisition (see 2.2.4). However, since neither participant was regarded as an "exceptional language learner" as characterized in section 2.2.4, these questions were not sufficiently detailed and ultimately discarded from analysis. However, in addition to being asked to evaluate their own skills as L2 learners, both participants were asked whether they were right-handed or left-handed, and whether they considered themselves talented in school subjects such as graphic arts, music, sports, math, and languages. Ultimately, however, these questions were considered unimportant information for the reason mentioned above.

[5] The Geschwind Cluster Questionnaire (see Appendix A in Schneiderman & Desmarais 1988) mentioned in section 2.2.4 is much more expansive than the interview questions used in this thesis. Schneiderman and Desmarais suggest that individuals with exceptional talent in learning second languages tend to have life histories and traits that include specific details regarding "birth, prenatal or perinatal difficulties; parental expectations; acquisition of reading; music; childhood diseases; other diseases such as allergies, eczema, asthma, hay fever; family history of schizophrenia; spatial orientation such as reading maps and sense of direction". Additionally, the questionnaire seeks to inquire how study subjects did in school subjects such as gym and athletics, math, science, music, and graphic arts. The view that such factors correlate with talented L2 learning is only partially accepted here, however. At any rate, the evidence is not sufficient to make any strong claims about the issue.

4 Analysis

4.1 Overview of Participant Classroom Performance

4.1.1 Participant 1

For each participant, it was important to first establish the variety of English spoken during their classroom teaching sessions in order to compare the produced segments against the participants' underlying pronunciation paradigm. As was anticipated, virtually all of the participants favored either the GA or the BrE variety. However, it was difficult to pinpoint exactly which variety participant 1 preferred. At times the pronunciation was reminiscent of BrE and other times GA, but in some instances, it appeared to resemble a typical Finnish accent. This variation was carefully documented.

Alternation between BrE and GA norms was most apparent in the use of rhotic consonant /r/. The deletions and non-deletions of the postvocalic /r/ were difficult to predict in cases where surrounding phonetic features, such as vowel quality, would not provide any insight as to what the intended and genuine variety was. For instance, the vowel quality in [wɜːrd] is not an obvious give-away regarding the variety used by the speaker. With the exception of the presence or absence of /r/, then, there does not appear to be an obvious segmental difference in the pronunciation of *word* as far as GA and BrE are concerned, although the deletion or non-deletion of /r/ may be expected to influence features such as lip posture in the preceding vowel. In the case of participant 1, the vowel quality in *word* did not approximate GA any more than BrE, but instead the Finnish mid front rounded vowel /ø/. Indeed, L2 speakers may often be expected to replace L2 vowels with the closest equivalents from the L1 vowel space.[6]

Strictly in terms of rhoticity, participant 1 was heard pronouncing *word* according to BrE norms, yet words such as *bored* and *worked* did not have the deletion observed in *word.* Further, r-colored vowels were heard in words such as *whether* and *wonder,* but not *other.* Words such as *there* and *are* also had typical BrE phonology. The observed r-colored vowels were limited to [ɚ] (for instance, [ɑˠ] was never heard), but a consistent and reliable pattern, i.e., a putative

[6] It bears repeating that even the same vowel phonemes are qualitatively different across different languages and may be seen plotted in slightly different locations in formant diagrams (Zsiga 2021: 40).

phonological rule, was not presumed at any point regarding these features. Pronunciation of vowels followed a similar pattern of being divided between GA and BrE, whereby some words sounded exclusively BrE and others exclusively GA. The nuclei of the stressed syllable in words such as *sorry*, *hospital*, and *common* were of the BrE variety in which the vowel would either be rounded open or rounded open-mid back. The GA equivalents of these are unrounded, [ɑ ~ ʌ]. The reverse was observed in the words *problems* and *not*.

Alternation between GA and BrE was not only observed in different words, but also within words where segments typical to both varieties would occur adjacently. For instance, *quarter* was pronounced with a non-rhotic /r/ and it was followed by an alveolar flap. T-voicing is more common in GA, occurring between sonorants and unstressed vowels, but the manner in which it was used by participant 1 (in the case of *quarter*) was more reminiscent of Australian English ([kwɔtə] with /t/ flapped). Further, the word *notice* was realized with GA vowels and with aspirated /t/, also raising the question whether the underlying phonology reflected a GA system or the L1 system. Incidents such as these, often appearing only once over the course of several hours of classroom recording, made it difficult to determine whether the variety alternation was a question of rule-governed behavior, rather than being individual word-governed.

While it appeared that t-voicing was more common in environments where BrE varieties typically demand an aspirated /t/, in some cases the /t/ was realized as a dentalized [t̪] rather than the typical [tʰ] or [ɾ], making certain words sound as though they were influenced by Finnish phonology. To answer the question of which of the three systems—GA, BrE or Finnish—was dominant in the pronunciation proved challenging at least as far as consonants are concerned. Analysis of five hours of audio recordings resulted in the interpretation that participant 1's interlanguage phonology is divided between at least the three paradigms mentioned above. Which paradigm comes to determine the phonological output appears to occur on a lexical basis, in a rather random distribution. The reasons why this is the case could be explained by what was uncovered in the interview (see 4.4). For this reason, the underlying preferred variety was deemed less relevant for the purposes of this thesis and more focus was placed on target norms that both GA and BrE speakers adhere to.

4.1.2 Participant 2

Participant 2 showed an unmistakable preference for the GA variety. This represented a stark contrast with participant 1. The GA phonology was consistent and never changed to, for instance, BrE pronunciation, making it easier to notice instances where the segments did not

match the target. The non-target segments that could be observed were infrequent. While her adherence to target norms in segmental phonology was remarkable, this could possibly be explained by her careful speech style, characterized by enunciation that was slower than what would be expected from L1 speakers in addition to relatively long pauses between words. The reason for this might be related to how she monitored her output, effectively sacrificing some prosodic aspects of speech in order to be grammatically accurate. Effectively, this strategy defeated the purpose of the research design by forcefully minimizing segmental accidents. Indeed, participant 2 claimed in her interview that she was extremely nervous while speaking in front of a classroom of students (see 4.7.2). The stressful situation might have slowed the speech style, yet in effect allowed extra time for mental planning to produce very convincing American English pronunciation which might otherwise have been riddled with performance errors. While excessive inter-word pauses and overall slower speech may cause the speaker to stand out as a non-native speaker, these qualities are prosodic in nature and outside the scope of this study. Instead, the analysis focuses solely on segmental aspects.

4.2 Production Task Data Analysis – Participant 1

The tokens that were selected and tested for participant 1 are summarized in table 2.

Table 2. Non-target segments observed in the classroom data and selected for the production task

Token	Phoneme	Example	TL-norm	Realized as
Velarization of lateral approximants	/l/	*tell*	[tʰɛɬ]	[tʰɛl]
L-vocalization in syllable rhyme	/l/	*false*	[fɔːls]	[fɔʊs]*
Perseverative assimilation at word boundary	/n ð/	*even though*	[ivɪn̪ðoʊ]	[ivɛntoʊ]
Voicing of sibilants	/z/	*exactly*	[braʊz]	[braʊs]

* The phonetic quality of the vocalized /l/ varies according to spoken variety.

4.2.1 Velarization of Lateral Approximants

RP and GA differ with respect to how /l/ and its velarized counterpart [ɫ] are phonetically distributed and realized. Firstly, it may be generalized that [ɫ] is ubiquitous in GA, making velarization an indicator of which variety the speaker is partial to.[7] Secondly, its realization involves more rules in BrE, meaning that learners inclined towards BrE must establish additional phonological rules concerning the phoneme: in RP, for instance, the secondary articulation does not take place as perceptibly if the consonant is post-vocalic, whereas in GA it does. However, the velarized /l/ is qualitatively different in RP, in that it sounds as though the dorsum of the tongue is not raised as high in the direction of the velum as it would be in GA, and additionally, the lips seem to be more rounded when /l/ is in a syllable rhyme. The rounding becomes apparent in words in which the preceding vowel is unrounded, such as *pill*. For brevity, the non-velarized /l/ will be henceforth called the clear /l/ and the velarized /l/ the dark /l/. The two differ in retraction of the posterior tongue body. Clear /l/ involves forming a single large cavity behind the alveodental constriction, while the dark /l/ involves increased tongue-root retraction, accompanied by degrees of tongue pre-dorsum lowering (Rodrigues *et al*. 2019; Giles & Moll 1975; Browman & Goldstein 1995; Narayanan *et al*. 1997). However, the difference between the clear and dark articulations is graded and not discrete as they both involve apical and dorsal gestures, the only difference being related to timing (Sprout & Fujimura 1993; Slomanson & Newman 2004).

Considering the case of Participant 1, it is important to note that the Finnish /l/ is mostly clear, and it could therefore be anticipated that the /l/ phonemes that appear would be closer to BrE, or that the allophone [ɫ] is entirely missing from his segmental feature inventory. While it is true that the clear /l/ predominated, the dark /l/ was not entirely absent. Clear /l/ could be heard in word-initial, word-medial, and word-final positions, such as *learn*, *really*, and *tell*. The dark articulation was rare and only heard in a syllable onset – in the word *longer* – in the classroom recordings. Although there were no other instances of the word *longer* in the data, similar words such as *learn* were produced with a clear /l/. Based on these observations, each of the words mentioned above were incorporated into the experimental production task with the anticipation

[7] According to Sproat & Fujimura (1993), the distinction between light and dark articulations respectively in prevocalic and post-vocalic positions "are widely reported" in both GA and RP. While this view is accepted in this thesis, it is also considered irrelevant due to the articulatory differences in all /l/ positions when comparing RP to GA.

that participant 1 would assume, at least for the duration of the task, a phonology more consistent with a specific variety.

Table 3. Formant frequencies measured for /l/ phonemes in participant 1's production task

Word	f_1 (Hz)	f_2 (Hz)	Classroom	Experiment
learn	381	1440	[l]	[l]
languages	295	1290	[l]	[l]
really	386	1650	[l]	[l]
longer	381	990	[l]	[ɬ]
tell	698	917	[l]	[ɬ]

The columns Classroom and Experiment indicate which variant was produced in both the classroom and the production task.

As indicated in table 3, there was consistency between classroom and experiment in three of the tested items, while one item, *longer*, was inconclusive as it became something of an intermediary between obvious clear and dark variants. However, the /l/ in the coda of *tell* must be either vocalized or velarized in BrE and GA respectively. Participant 1 opted for velarization, much in the same way an American would, realizing the word as [t^hɛɬ] (figure 1).

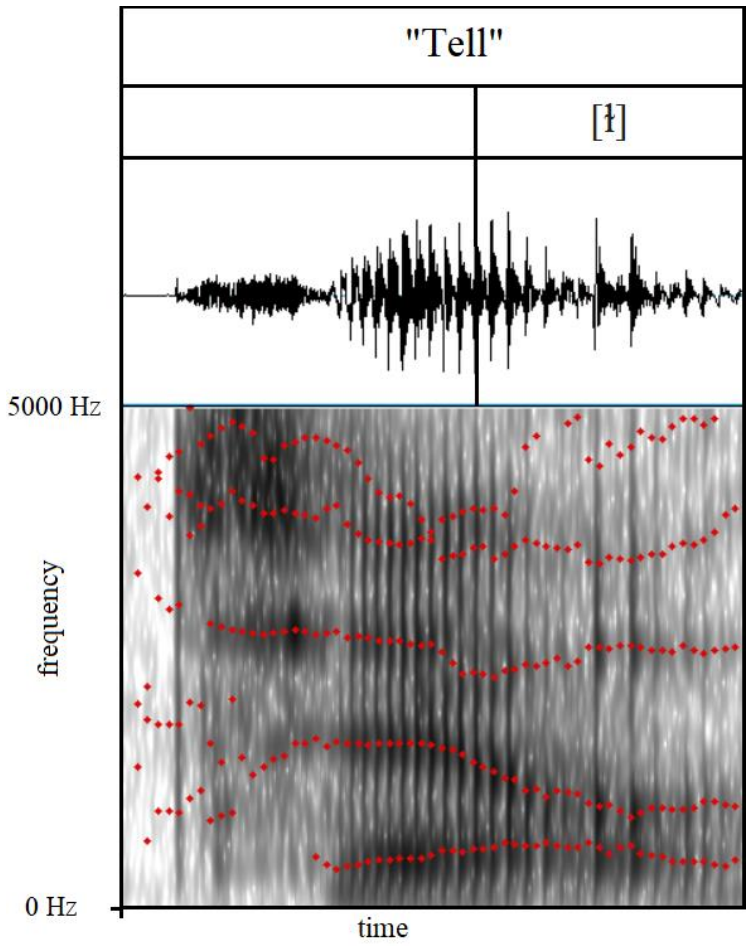

Figure 1. Waveform and spectrogram of *tell* with a dark allophone [ɬ] in the syllable coda position.

It should be noted that all words in table 3 can be said to adhere to target norms in the way they were articulated in the production task, provided that the speaker is presumed to have acquired both varieties and has the capacity to switch between them at will. Participant 1, however, speaks a variety that sounds more like BrE than GA across the board. It could be argued, then, that he should favor l-vocalization in syllable codas, as in *tell*. As seen from figure 1, however, such vocalization does not take place, but instead the /l/ takes on a velarized secondary articulation. The formant pattern in places of f_1 and f_2 indicates that the /l/ is highly velarized. The formant patterns represented by red dots would appear different if the /l/ were vocalized as they would be affected by features such as lip rounding.

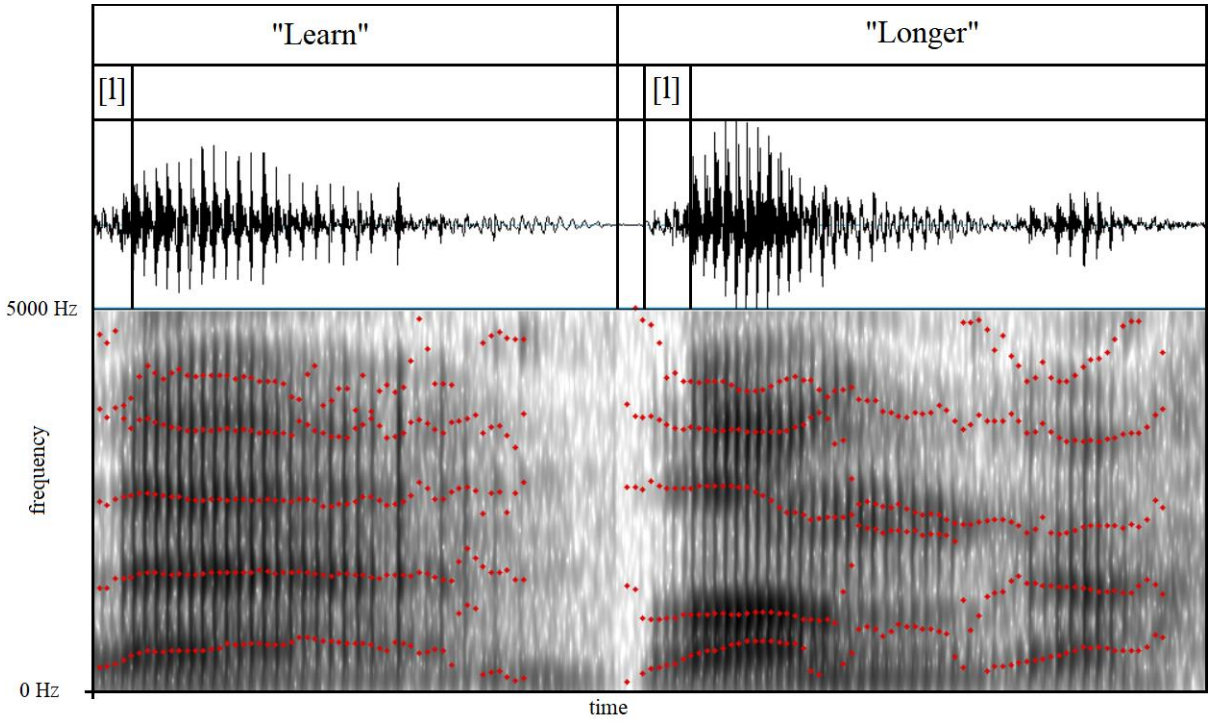

Figure 2. Waveforms and spectrograms showing a clear allophone of /l/ in the onset of *learn* and a darker equivalent in the onset of *longer*.

Figures 1 and 2 illustrate the differences in phonetic quality of the phoneme /l/ across three words. The /l/ in *longer* is significantly darker than it is in *learn*, which may be seen from the tapered distance between the first and second formants. However, the first formant (i.e., the lowest string of red dots in the spectrogram) is virtually equal in *longer* and *learn*. *Longer* was perceptibly dark in the production task, whereas in the classroom it was perceptibly clear, leaving open the question of whether the /l/ actually is dark in figure 2. The velarization of /l/ in *tell* on the other hand is evident from the "velar pinch" occurring between the first and second formants (indicated by the two lines of red dots drawing closer together in figure 1) in transition from the vowel [ɛ]. The /l/ was unambiguously dark in *tell*, confirmed also by the significantly raised f_1 and lowered f_2. The reason for these differences could be a function of initial as opposed to final position of the segment in an utterance[8]: when no speech sounds immediately follow /l/, it does not require apical contact with the alveolar ridge, the absence of which might be compensated by velar raising. In onsets, the apicoalveolar contact is more likely.

[8] Utterance-final segments as opposed to word-final segments is not an arbitrary distinction since the word boundary itself may effectively be non-existent due to the timing of when the last segment of a word and the first segment of the following word commences as these two may overlap. However, at the end of an utterance, be it a single word or a longer phrase, there is necessarily a pause indicating a break between intonation phrases. These breaks cancel out the effects of co-articulation on segments at word boundaries.

4.2.2 L-vocalization in Syllable Rhyme

The most revealing words with respect to the predominating phonology were found to be those in which the syllable rhyme /l/ was clear, a violation of both BrE and GA standards – a word such as *false*, in GA, has an audible [ɬ], but in BrE it may sound as though the /l/ is deleted or becomes vocalized to [ɤ̯] (Sprout & Fujimura 1993), the result of a phonological process called L-vocalization. *False* was pronounced with an audibly clear /l/, sounding as though it is influenced by the L1 more so than either British or American variety. It was determined that this token, rather than an imitation of BrE, was a case of the speaker defaulting to a phonemic L1 equivalent. Regarding vocalization, words such as *all* in BrE usually change to take on the syllable shape V instead of VC.

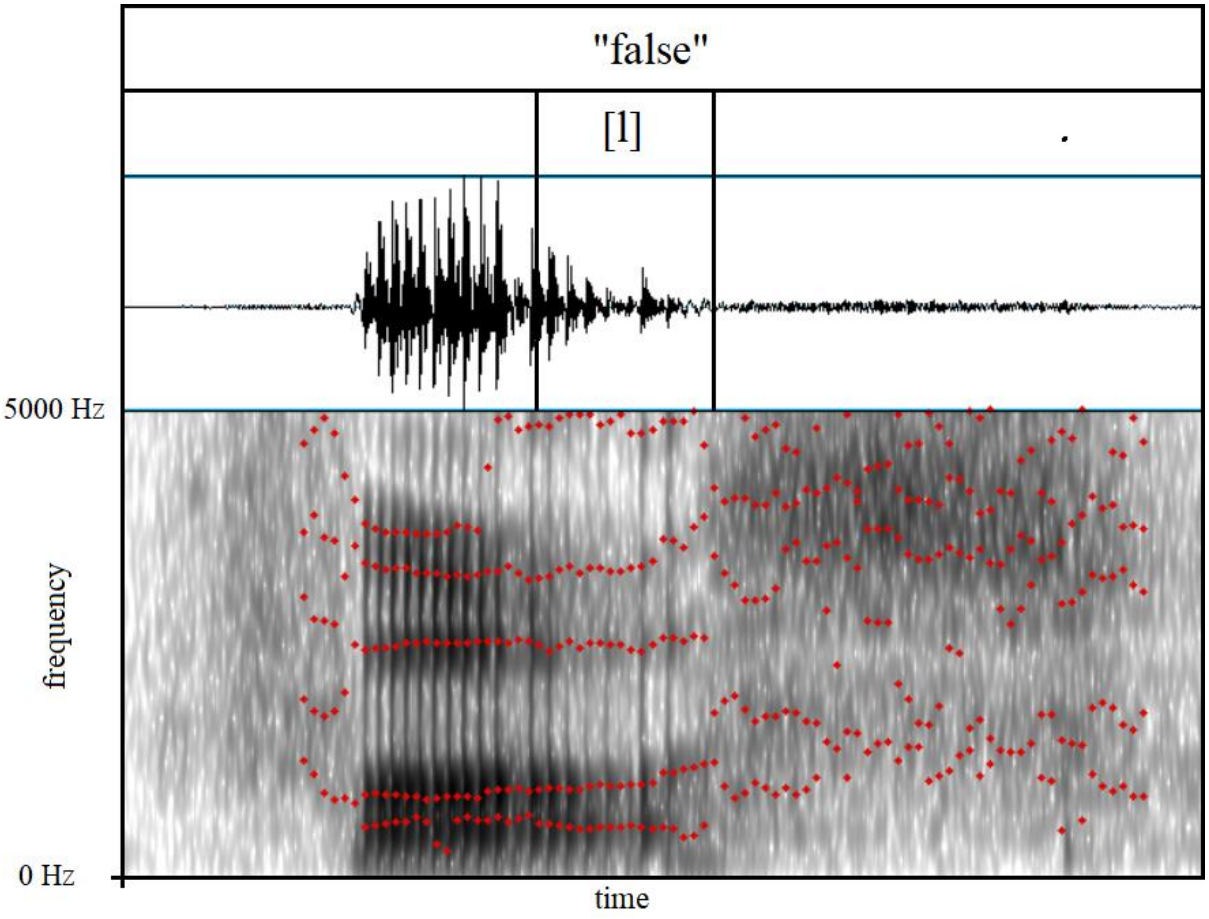

Figure 2. Waveform and spectrogram of *false*.

False is often transcribed as [fɔːls]. However, in BrE (e.g., RP), the vowel quality differs from GA in lip compression, which makes the vowel sound more rounded. Further, in RP, the /l/ is normally vocalized into a compressed version of the preceding vowel, i.e., [ɔ] takes on compressed rounding, which means that the margins of the lips are tense and drawn together

during articulation. In essence, the /l/ disappears in words such as *false* in BrE, while in GA it remains audible. Thus, in narrow transcription, the syllable nucleus of *false* might be written as [ɔᵝ]. In a spectrogram, these rounding features would undoubtedly change the formant structure, but instead in figure 3 the formant pattern seamlessly transitions from the nucleus to the coda, the /l/ being audible and laryngealized.

4.2.3 Perseverative instead of Anticipatory Assimilation at Word Boundary

In words such as *though*, it should be noted, some L1 speakers substitute the normative dental fricative for plosives, such as [d] or [t]. However, this substitution is also attributed to sociolects or ethnolects found in English-speaking countries. This being the case, such a possibility was controlled for in choosing the participants for the study. Indeed, Participant 1 stated that he has not lived nor spent a considerable amount of time in any English-speaking countries, so it is unlikely that he had acquired any local sociolects in one. He was heard substituting [ð] with either [t] or [d] – the features of which are highly similar – in the classroom, only to reproduce the same substitution in the experimental context.

Though the substitution of fricative for plosive was perceptible, it was difficult to determine whether the replaced segment was [t] or [d]. These segments differ primarily in phonation, and the carrier phrase in the production task was *even though*. The word boundary in this phrase consists of a nasal and a fricative, both of which are voiced, and so the presence of voicing, or lack thereof, is not relevant as the substitute to [ð] had voiced segments on either side.

The selection of the phrase *even though* for the production task is not random, but rather selected from the exact oral production in the classroom recordings in which the substitution was perceived. The phoneme /ð/ does not appear to be missing from participant 1's L2 phonology as he had no problem producing it in other common words such as *the*. The reason for the substitution, it is speculated here, is phonologically conditioned. More precisely, it is related to assimilatory processes. In standard English, be it BrE or GA, the nasal consonant in a phrase such as *even though* becomes interdental, a regressive right-to-left movement in assimilation as it is anticipating the place of articulation of the dental fricative that follows. In the case of participant 1, this process appears to have been reversed as the initial segment of *though* is influenced by the place of termination of the segment in [n], the alveolar ridge, in *even*. In short, participant 1 uses anticipatory assimilation in lieu of perseverative assimilation. Why this is the case is a matter of conjecture, but it could be speculated that, in addition to interdental fricatives being phonologically marked, interdental consonants are less natural for

Finns, and Finns could thus be expected to opt for the types of assimilation in which interdentals do not feature, while not having as much difficulty producing them in non-assimilatory forms.

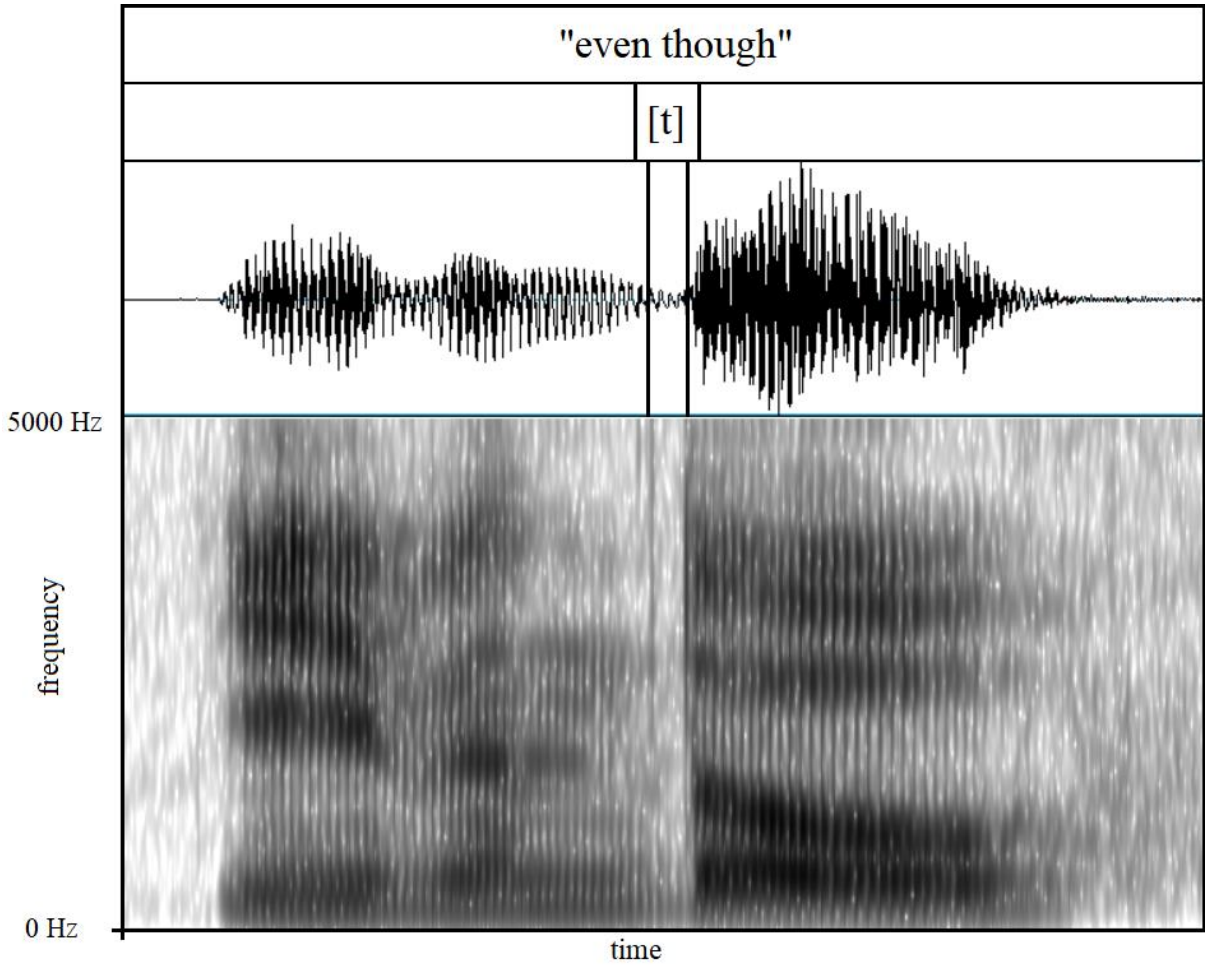

Figure 3. Waveform and spectrogram showing anticipatory assimilation in the carrier phrase *even though*.

In figure 4, The most reliable indicator of the absence of interdental articulation is the "breaking" of the second formant (which in figure 4 commences at approximately 2000 Hz). In perseverative assimilation (where the [n] would change to [n̪]), the second formant would maintain a consistent line throughout the phrase, which it does not do in the spectrogram in figure 4. Instead, *though* begins with an alveolar plosive [t], represented by the silent phase, i.e., absence of acoustic energy in figure 4.

An attempt to elicit the phoneme /ð/ was made in other phonetic environments in the production task, namely in the phrase *she knows that* in which the voicing of the final segment of *knows* would anticipate the voicing of the initial segment in *that*. Contrary to this expectation, the [z]

in [nəʊz] was devoiced and the dental fricative was realized similarly to the [t] in *even though*, though the phonetic quality differed slightly.

4.2.4 Voicing of Sibilants and Plosive-Sibilant Clusters

As anticipated, voicing alternation turned out to be the most salient non-target feature in participant 1's English phonology. This anticipation stems from the fact that Finnish appears to notably lack voiced consonants.[9] In this study, the phoneme /z/ received a lot of attention, vulnerable as it is to substitutions due to spelling pronunciations. Further, /z/ does not exist in Finnish phonology, and orthographically the letter <z> exists primarily in loan words. Thus, in words such as *pizza*, the [z] is normally realized as [ts]. Indeed, it was expected that spelling pronunciations that are non-target would be frequent in the form of devoiced sibilants in the phonology of all participants. In the case of participant 1, numerous instances were observed in classroom production and reproduced, verbatim, for the production task. These instances, it should be noted, were words in which the spelling does not include the grapheme <z>, but instead <x> or <s> – *use*, *easy*, *because*, *exactly*, *exhausted* – each of which normally contain the phone [z], but not the grapheme <z>. Included in this list was also the word *is*, but this is not always realized as [ɪz] by L1 speakers, but sometimes as [ɪs]. The sibilant voicing in *is* appears to be influenced by the phonation of the following segment, at least for some L1 English speakers. Participant 1's performance in the production task in comparison to classroom is summarized in table 4.

[9] For instance, in the case of the loan word *banaani*, the initial consonant has free variation between /p/ and /b/ in Finnish. In particular, the phonemes /z g b/ seem rare or totally absent among Finnish consonants and are frequently replaced with the phonemes /s k p/, respectively.

Table 4. Words tested for sibilant and cluster voicing

Word	Classroom	Experiment
use	[s]	[s]
because	[s]	[z]
explain	[ks]	[ks]
exactly	[ks]	[gz]
exhausted	[ks]	[gz]

The columns Classroom and Experiment indicate which variant was produced both in the classroom and the production task.

In contrast to laterals, formant frequencies for sibilants were not measured due to the turbulent airstream characteristic of fricative consonant articulation whereby the frequencies on spectrogram tend to be randomly distributed. Moreover, it should be noted that there are various possible articulations for /s/ (e.g., pre-alveolar, alveolar, post-alveolar), which differ in a phonetic, but not phonemic, sense. I.e., the variation is often stylistic or idiolectal, at the very least in Finnish and English, the speakers of which may freely alternate between different realizations of /s/. Since the variable investigated here is one of voicing, the only relevant diagnostic is thus the voicing bar at the f_0 region. This serves as a clear demarcation between sibilant fricatives such as /s z/ and /ʃ ʒ/, and non-sibilant fricatives such as /θ ð/, which otherwise tend to appear similar in spectrographic representations. In this study, the focus was only on the /s z/ contrast.

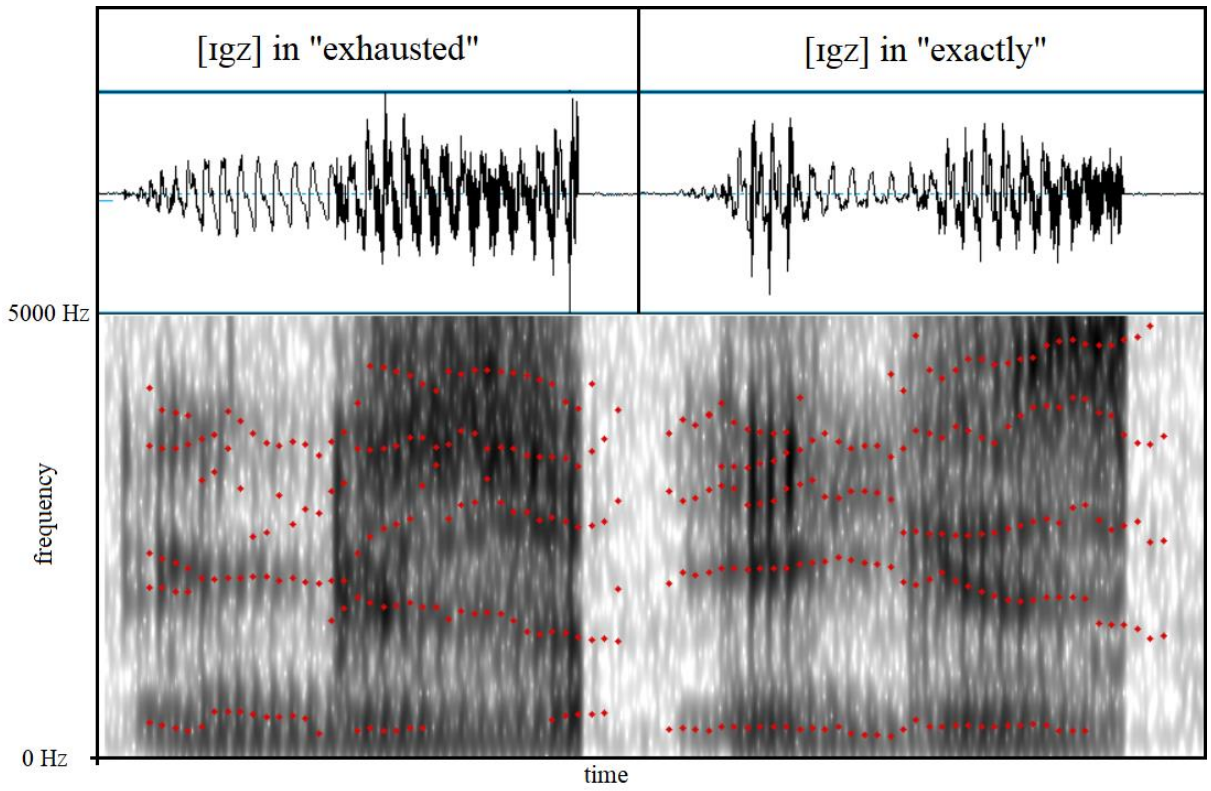

Figure 4. Waveforms and spectrograms showing consonant cluster phonation in the words *exhausted* and *exactly*. The sibilants in each cluster represent the "noisiest" portion in the waveform and spectrogram representations.

The consonant cluster coda of the first syllable in *exhausted* and *exactly* was typically realized by participant 1 as [ks] in the classroom. However, these clusters were correctly voiced in the production task, while in words such as *explain* the cluster was, also correctly, devoiced. This is represented in figures 4 and 5. Participant 1 never voiced the consonants in *explain* while in classroom, but this word was nevertheless included in the production task to test for possible voicing differentiation between words that look orthographically similar. In figures 4 and 5, the difference between [gz] and [ks] is clearly perceptible not only by the respective presence and absence of the seamless voicing band at 0 Hz across the first three segments, but also by the breaking of the first formant (represented by the lowest line of red dots) when the consonants in *explain* are left unvoiced. The first formant in *explain* recovers its consistent pattern only when the articulation of /l/ begins.

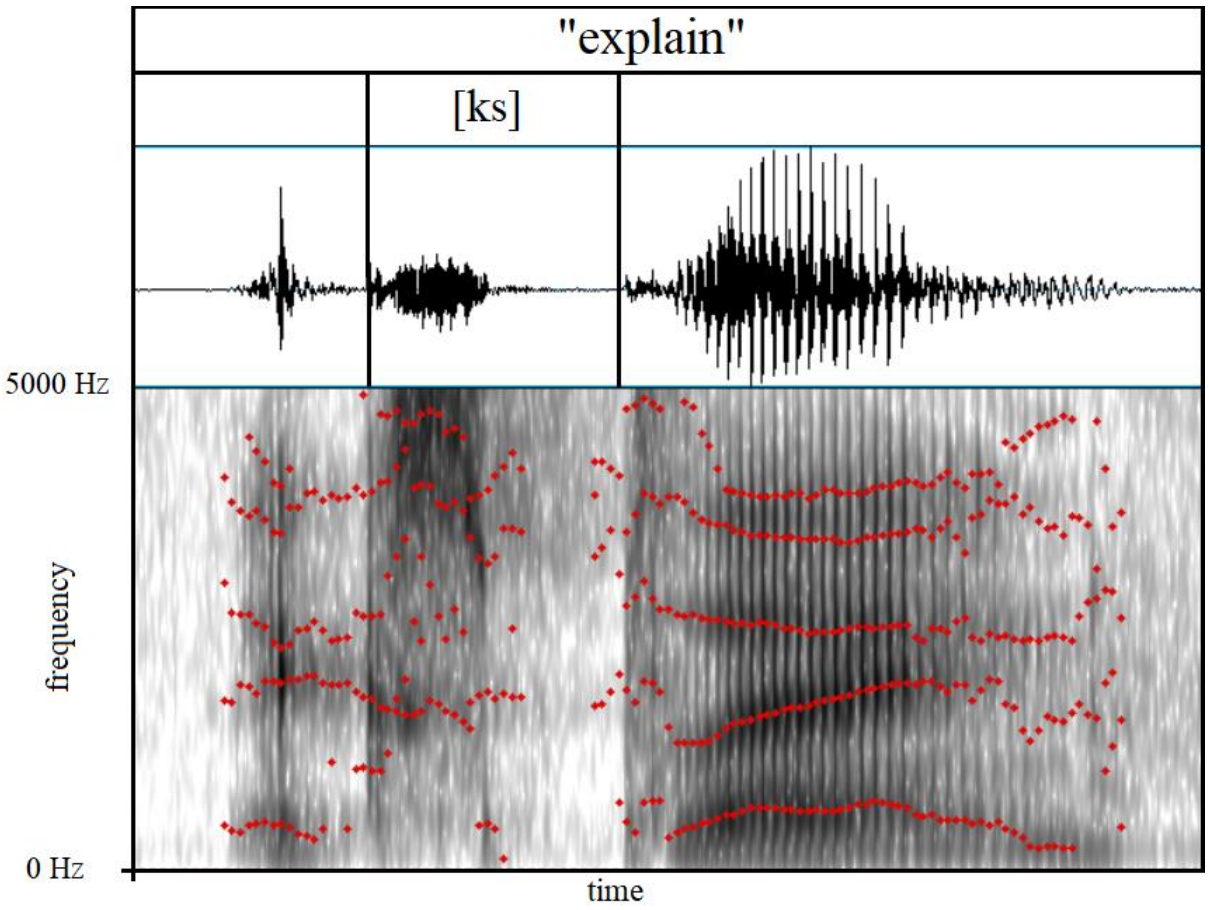

Figure 5. Waveform and spectrogram illustrating the devoiced coda of the first syllable in *explain*.

For L2 speakers, words such as *exhausted* may be difficult to decipher phonologically to produce a target-like output. The challenge is, as far as consonants are concerned, being able to analyze the grapheme <x> as [gz] and not [ks], and additionally being aware that the grapheme <h> has zero realization. The voicing aspect applies also in the word *exactly*. Though the omission of /h/ is not visible in figure 5, participant 1 managed to produce the word in a target-like manner.

Interestingly, sibilants were the only token category in which participant 1 made repetitions during the production task and attempted to, it appears, self-correct. This could mean that the difference between /s/ and /z/ is not entirely clear in the mental grammar, which is therefore causing some hesitation. From the spectrographic representation, it appears that the repeated articulations mostly modified the sibilants in the carrier words that were corrected. For instance, the sibilant in *use* was uttered differently upon repetition while the other segments remained practically indistinguishable (see figure 7). The reason why this word was repeated is not known, but the qualitative difference in [s] suggests that participant 1 had doubts regarding its accuracy. However, *use* was not realized as [ju:z], which is the common target

expectation, but [ju:s] instead – the voicing bar is not present in either iteration of *use* in figure 7, yet the acoustic intensity has been increased in the "corrected" sibilant (represented by the significantly darker areas at f_2 and f_3 regions). This indicates that participant 1 heard the inaccuracy of [s] in the word and attempted to modify it but was not aware that the required modification was [+voice]. The two instances of [s] are audibly distinct. The repetition sounds as though it is articulated with the tongue tip further past the alveolar ridge while the first iteration sounds more retroflexed, which is not uncommon for /s/ in Finnish phonology.

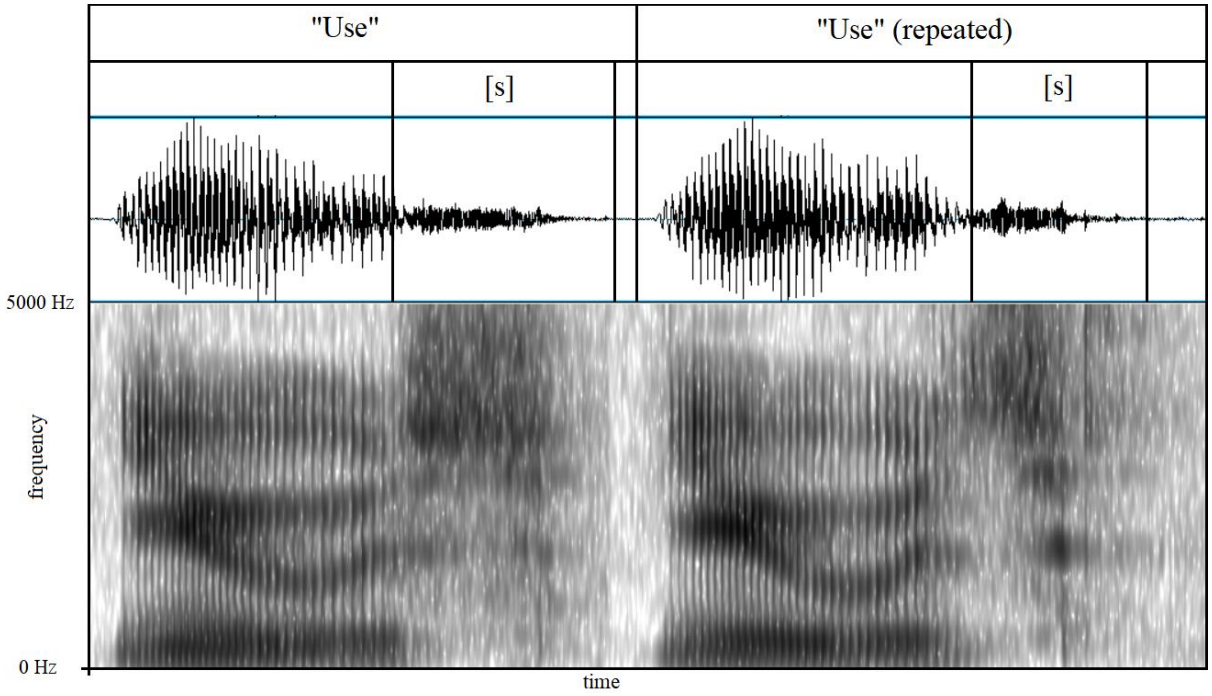

Figure 6. Waveform and spectrogram of the word *use* (left), followed by a "self-correction" (right) in which the word-final sibilant segment is qualitatively different.

In addition to *use*, the word *because* was also repeated. However, the modification was segmental since the feature [+voice] was implemented, thus changing [s] to [z]. The difference may be seen in figure 8 in which the voicing bar is visible at the bottom of the spectrogram in the indicated region of [z], while in the first utterance the indicated region of [s] does not display an obvious voicing bar, and additionally the region near 5000 Hz has more acoustic energy in comparison to the repeated word.

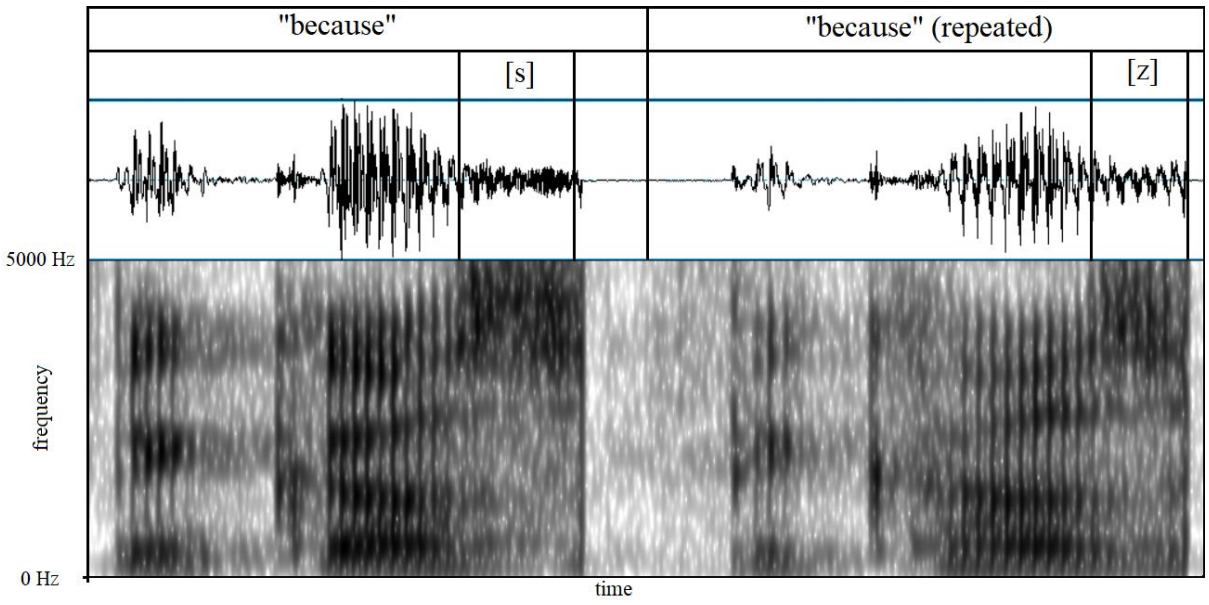

Figure 7. Waveform and spectrogram of the word *because* (left), followed by a self-correction (right) in which the word-final sibilant segment has been changed from [s] to [z].

It is not clear why the self-correction in the case of *because* was a matter of phonation yet not in the case of *use*. However, the acoustic signature in *use* is remarkably similar, save for [s], while in *because* the acoustic energy has less intensity across the entire word as represented by the overall darker tone in the first iteration. Further, *because* was part of a carrier phrase (*I'm sick because of the flu*), while *use* was presented as a singular word in the list of stimuli, raising the suspicion that the voicing was a case of co-articulation due to /s/ being surrounded by voiced segments. The voicing in *because* could have thus been accidentally applied, while in *use* the change from [s] to [z] requires more explicit knowledge of the appropriate sibilant features when its carrier word is presented in isolation, i.e., not part of a larger utterance.

4.3 Receptive Task – Participant 1

The receptive tasks were carried out after the production tasks, making it unnecessary to camouflage the study design. In each phrase there was variation in place of one segment. For instance, in the phrase *I'm zipping my jacket* the variation was in the voicing of the first segment of *zipping*. The variation was in the onset to prevent any perceived differences in other segments, real or imagined, as words are known to be distinguished by the duration of the vowel that precede voiced codas and not necessarily voicing itself. This was also the case in *sip* in the phrase *I sip from this mug*. While these were instances of meaning-changing variation, in other stimuli, such as *football*, there were no meaning-changing segmental differences, but

the participant instead had to choose the "better-sounding" option from velarized and non-velarized alternatives. Clear-dark contrasts were recorded in words in which the clear allophone would not be expected in L1 English speech, regardless of variety, such as *football*, *basically* being the exception as it has a clear /l/ in BrE. Though it was not included in participant 1's production task, *basically* was included in the task to test for sensitivity regarding the accuracy of a specific variety and additionally for hearing the contrast.

As regards *false* and *even though*, these were two discrete tokens that only had one corresponding stimulus in the receptive task while the allophones of /l/ and the voicing contrasts had multiple different stimuli. The reason for this is that the two tokens in question were only elicited once in the production task and therefore only occupied one position: in the coda of *false* for vocalization and in the onset in a lexical boundary for the alveolar nasal-interdental fricative [nð]. Participant 1's performance in the receptive task is summarized in table 5.

Table 5. Summary of participant 1's receptive task results with production task results included for comparison with the same phrases

Contrast	Phrase	Choice	Production Task
[z], *[s]	*I'm zipping my jacket*	[z]	
[z], *[s]	*lots of words*	[z]	
*[z], [s]	*I sip from this mug*	[s]	
[gz], *[ks]	*exactly*	[gz]	[gz]
*[l], [ɬ]	*basically*	[ɬ]	
*[l], [ɬ]	*tell*	[ɬ]	[ɬ]
*[l], [ɬ]	*football*	[ɬ]	
*[l], [ɔː]	*false*	[ɔː]	*[l]
[nð], *[nt]	*even though*	[nð]	*[nt]

In table 4, the contrasts marked with asterisks would not be expected in native speaker phonology in the listed phrases, and as may be seen in the Choice column, participant 1 judged all phrases with the more native sounding contrast as being more accurate. Considering that there are no errors from the perspective of the task design, these results suggest that all of the

tested phoneme contrasts are acquired by the comprehension apparatus and thus acquisition failures may be ruled out as an explanation for the errors in the production task.

4.4 Interview – Participant 1

As was stated in chapter 3, the purpose of the interview was to gain insight into the underlying attitudes, beliefs, and strategies regarding English pronunciation in situations calling for production performance. The findings from the interview could then help in interpreting the production task results.

4.4.1 Attitudes on Pronunciation Inaccuracies

The purpose of the questions was related to pronunciation ideology which was explicitly linked to acting in the role of a language teacher. When the role is assumed, the teacher may be expected to have formed an idealized version of their own pronunciation in the foreign language being taught. The formulated question to make the participants elucidate their own views on pronunciation was how important they felt having native phonology[10] to be in the L2. Participant 1 expressed this to be rather important, though clarifying that speech fluency is much more important for L2 speakers. He further elaborated that while it is not necessary to sound exactly as a native speaker, one[11] should sound natural and exemplary, referring specifically to prosodic features such as word stress and intonation.

To gain similar but more general information on the topic, the participants were further asked to evaluate the claim that L1 segments present in L2 speech equates to erroneous pronunciation. For this, participant 1 gave a nuanced answer, saying that the claim is partially true: retaining some non-target L1 vowels in L2 speech is fine, but the L2 speaker should acquire phonemes that distinguish meaning in some words, providing an example of [t] and [ð] in words such as *third* where the L2 speaker is likely to opt for [t] if the interdental consonant has not been assigned to the word or it is entirely missing from the phonemic inventory.

During the interview, participant 1 made clear that the native speaker phonology, although a standard to be upheld, should not be the end goal of L2 acquisition of English students insofar as all its phonemes and their features are concerned. In order to get the participants reflecting

[10] Native phonology, here, means L2 phonology that approximates native norms as closely as possible.
[11] In this statement, "one" is not referring to L2 speakers in general, but specifically the interviewee himself in the role of a language teacher in a classroom teaching context.

on the topic more deeply, they were asked whether students should be taught to speak with perfect syntactic constructions and other non-phonological areas of grammar if the students' phonological acquisition can reasonably be left in an imperfect state. In his response, participant 1 described that the other areas of grammar, such as syntax, are not as "personal" as pronunciation and therefore it causes no one harm to prescribe target norms in syntax, adding that the English word order is stricter than it is in Finnish and may seriously impede comprehension when used incorrectly.

4.4.2 Pronunciation Strategies and Performance Anxiety

Regarding the theme about psychological stress, the inquiry focused on how stressful the participants experienced speaking English in the role of a teacher in front of students, and moreover, how self-conscious they are about their English pronunciation in general.

Participant 1 admitted to feeling nervous about teaching English, partly due to feelings of self-doubt. The anxiety was present, in his words, particularly at the beginning of lessons, gradually abating towards the end, but lasting throughout the entire training period, i.e., the first practice lessons were not any less stressful than the later lessons. When asked whether he had found, either generally or specifically during teacher training, such anxiety to affect his verbal abilities, he responded positively. During and immediately before the practice lessons, participant 1 described pacing back and forth and occasional jumbled speech as typical signs of nervousness, later adding that stuttering and repetition of words tend to also occur while speaking in a state of heightened psychological stress.

The monitor concept was briefly described to the participants as a simultaneous thought process whereby the speaker thinks about the well-formedness and grammaticality of their speech. Participant 1 described a tendency to monitor his linguistic output "to a fault", suspecting that this also might hinder speech production itself. According to him, the monitoring pertains specifically to pronunciation whereas other grammatical features of speech are "natural" and automatic. The pronunciation, he added, remains a preoccupation because how second language speech sounds may not match reality from the speaker's own perspective.

4.5 Production Task – Participant 2

The tokens that were selected and tested for participant 2 are summarized in table 6.

Table 6. Non-target segments observed in the classroom data and selected for the production task

Token	Phoneme	Example	Target norm (GA)	Realized as
Velarization in Lateral Approximants	/l/	*please*	[pʰɬiːz]	[pʰliːz]
Voicing of Sibilants and Plosive-Sibilant Clusters	/s/	*wins*	[wɪnz]	[wɪns]
Audible Release of Word-final Plosives	/t/	*bit*	[bɪt̚]	[bɪt]

4.5.1 Velarization in Lateral Approximants

When the spoken variety is of the North American persuasion, the clear-dark /l/ contrast emerges as an important distinguishing phonological element. Indeed, the /l/ that is heard in GA is particularly dark when compared to BrE varieties (see 4.2.1 for a brief overview) and was therefore anticipated to be ubiquitous in participant 2's phonological surface forms. For the most part, this turned out to be the case, save for a few instances in which the clear allophone was produced, seemingly due to inattention (underlyingly working memory limitations, it is presumed here). The instances in which the /l/ was perceptibly clear were conspicuous due to how generally consistent the GA norms were in participant 2's phonology. One such instance occurred in the first recorded lesson, at the beginning of which she is heard saying *please* with a clear articulation of the /l/ and similarly, though not as obviously, three minutes later in *please listen*. Both *please* and *listen* should have a dark articulation of the /l/, though to varying degrees – in *please* the /l/ is typically darker by comparison, but the /l/ in *listen* is nevertheless dark when contrasted with the same phoneme in Finnish. However, the /l/ phonemes that were uttered in this case were reminiscent of the Finnish phonemic equivalent.

The words tested for velarization are summarized in table 7. Generally, the clear articulation is extremely rare in participant 2's phonology. For instance, a word-final clear /l/ was never observed. For a possible contrast to a clear articulation, included in the production task were

words that participant 2 repeatedly produced with an obvious dark articulation, namely *football* and *astrology*.

Table 7. Formant frequencies measured for the /l/ phoneme in words from participant 2's production task

Word	f_1 (Hz)	f_2 (Hz)	Classroom	Experiment
please	439	1330	[l]	[ɬ]
basically	568	1377	[l]	[ɬ]
hopefully	435	1103	[l]	[ɬ]
astrology	532	1047	[ɬ]	[ɬ]
football	481	917	[ɬ]	[ɬ]

As indicated by table 7, all of the words had a perceptibly dark articulation in the experiment, though the degree of velarization was not the same. Unsurprisingly, *astrology* and *football* are the most velarized, as the f_1 and f_2 frequencies have the least separation, and they were always perceptibly dark in the classroom data. The reason as to why these words are more velarized than the rest is likely related to the vowel features of the segment preceding the /l/ in both words, namely that of tongue retraction characteristic of back vowels. This is not the case for the words that had perceptibly clear articulations in the classroom data – the /l/ in *please* and *basically* are preceded by bilabial and velar plosives respectively, while in *hopefully* it is preceded by a labiodental fricative (the vowel /u/ was not realized between /f/ and /l/). However, all the articulations in the investigated words were perceptibly dark (see figures 8, 9 and 10) and there is a possibility that the measured f_1-f_2 values are not entirely accurate. Indeed, it is challenging to reliably measure the velarization of rapidly articulated laterals since velarization is a secondary articulation whose onset may overlap with a preceding segment. For example, in the case of *please*, the posterior raising of the tongue towards the velum is likely to commence during the aspiration phase of the plosive, before the lateral is released. Nevertheless, there is nothing inherent about aspiration that would anticipate the secondary articulation of the /l/ and must be specifically applied by the speaker. In other words, it is not any less likely to be realized as [ɬ] than it is as [l] in onset clusters such as /pl/ as an effect of co-articulation, while the surrounding vowels may exert an influence.

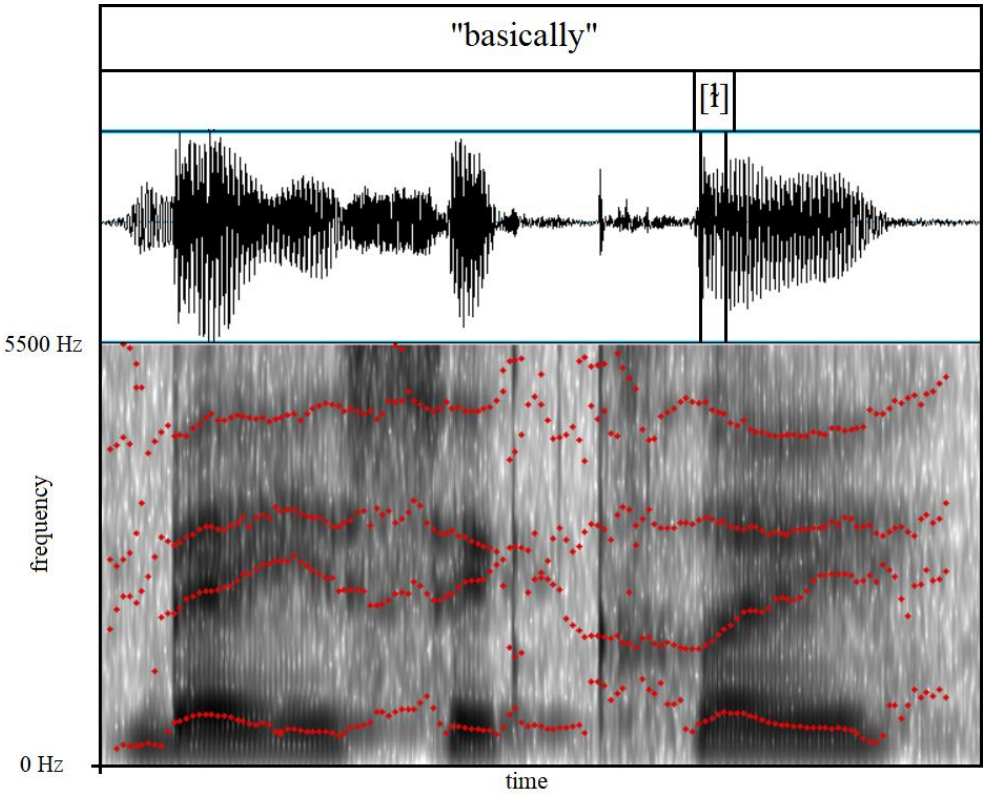

Figure 8. Waveform and spectrogram of *basically*.

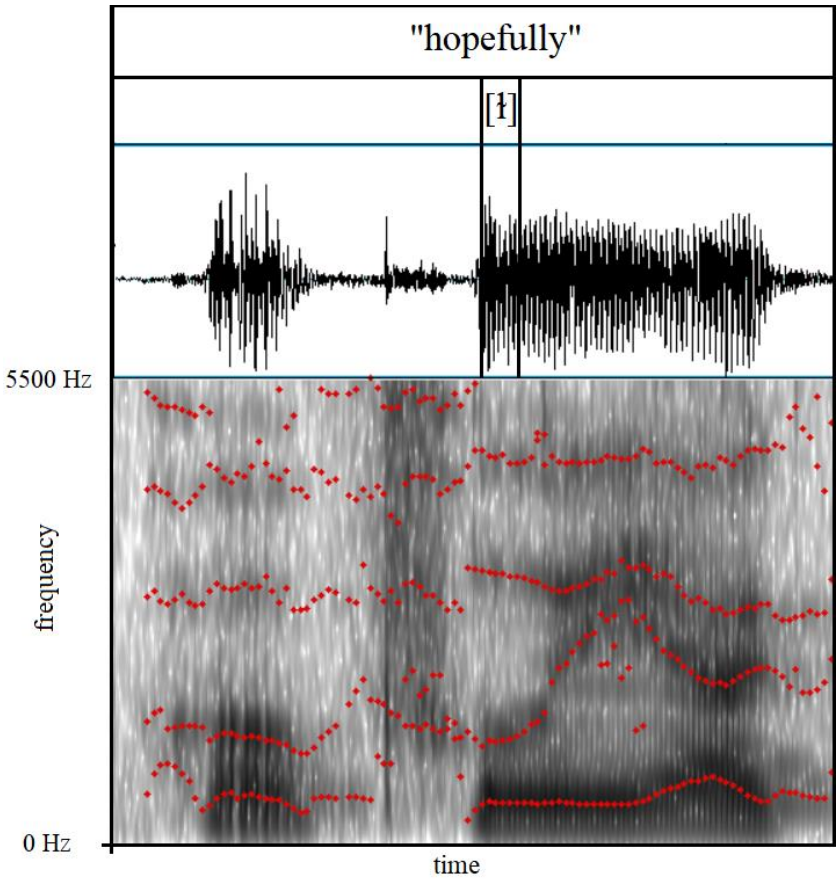

Figure 9. Waveform and spectrogram of *hopefully*.

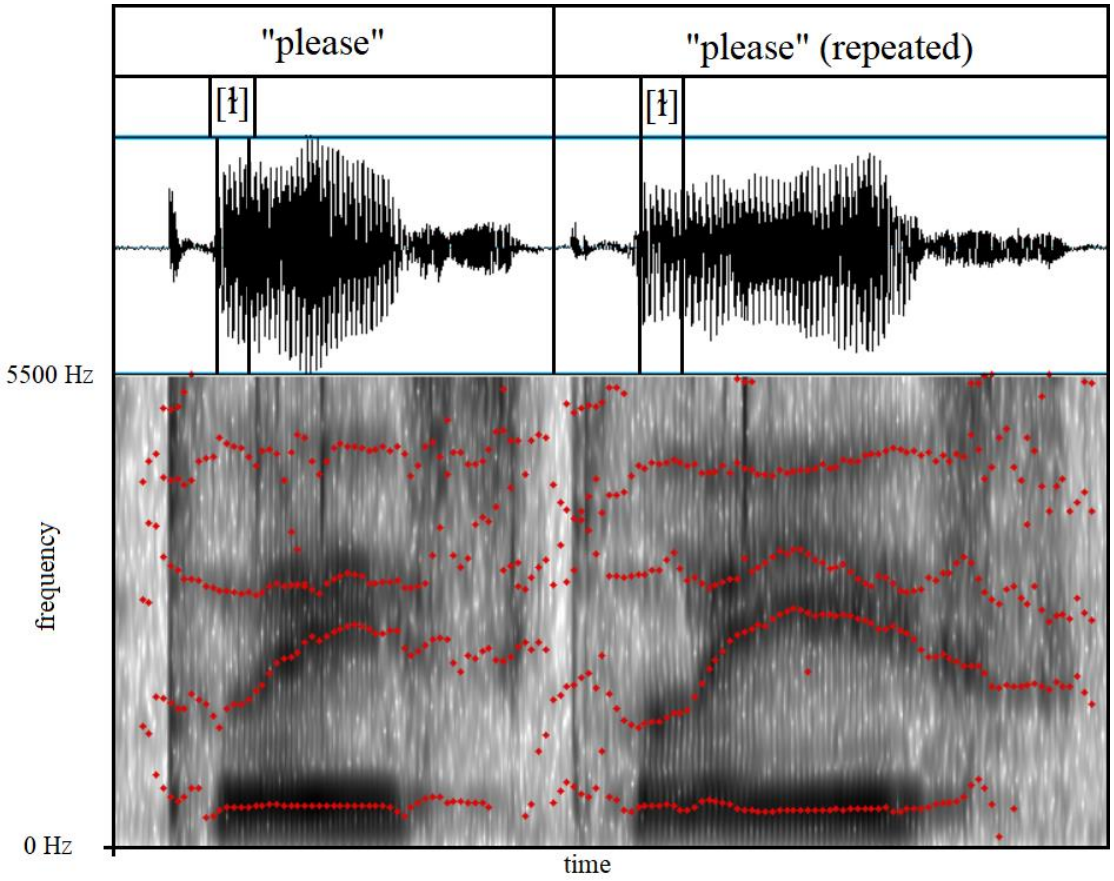

Figure 10. Waveforms and spectrograms of *please* and its repetition which shows increased velarization.

The experimental results show that the velarized allophone is favored in most, if not all, articulations. Meanwhile, the clear allophone, which in this case is interpreted as being the L1 equivalent of the normally used L2 /l/ phonemes, is only realized as an error of performance, not an error of competence.

4.5.2 Voicing of Sibilants and Plosive-Sibilant Clusters

While teaching her English lessons, participant 2 was observed producing multiple words in which the sibilant, which normally would be voiced in L1 English speech, was not voiced. For the production task, some of these words were included in phrases that would make voicing more likely. It is important to note that some words change meaning if the sibilant is left unvoiced, creating minimal pairs conditioned on voicing. Such minimal pairs are numerous in English: *close* (verb) and *close* (adverb); *ice* and *eyes*; *face* and *phase*; *juice* and *Jews*. Voicing in these lexical pairs tends to also affect pre-sibilant vowel length, which is often used as a cue for identifying /z/ as opposed to its unvoiced counterpart (Cole & Cooper 1976; Derr & Massaro 1980; Jansen 2004).

The words tested for sibilant phonation are summarized in table 8. As was the case for participant 1, these were selected based on the words heard in the recorded classroom data and judged as having unvoiced sibilants contrary to target norms.

Table 8. Words tested for sibilant and cluster voicing

Word	Classroom	Experiment
wins	[s]	[s]
example	[s]	[z]
words	[s]	[s]
things	[s]	[z]

The columns Classroom and Experiment indicate which variant was produced both in the classroom and the production task.

A particular word of interest that participant 2 repeatedly used in the classroom was the word *wins* (in phrases such as *so we know who wins*). Whether used as a pluralized noun or as a verb in second person singular present tense, the -s suffix in *wins* ([wɪnz]) should be voiced as its unvoiced counterpart corresponds phonologically to the verb *wince* ([wɪns]). It is debatable whether the nasal consonant is lengthened by the voicing of the following sibilant, though that is not relevant here. Be that as it may, *wins* was considered appropriate for testing the /z s/ contrast.

Among participant 2's stimuli, *wins* was included as an isolated word, i.e., not as part of a longer phrase. Whether analyzed as a plural noun (e.g., *lots of wins*) or a present tense verb (e.g., *she wins*), the sibilant in this word is expected to be voiced by L1 speakers.[12] The acoustic signature of participant 2's production of this word is represented in figure 12. It appears that her phonological representation of *win*, once the inflectional suffix -s is added, is underspecified for voicing in the sibilant segment.

[12] It is debatable whether L1 speakers necessarily differentiate phonological minimal pairs such as *wince* and *wins* at all, but rather analyze them as homophonous, relying mostly on pragmatic and semantic cues for disambiguation. Much of the processing behind similar words is likely subconscious, and it is moreover likely that the phonetic makeup of *wince* does not occur as a possibility, especially for L2 speakers, when reading a different, but phonetically similar, word out loud from a paper.

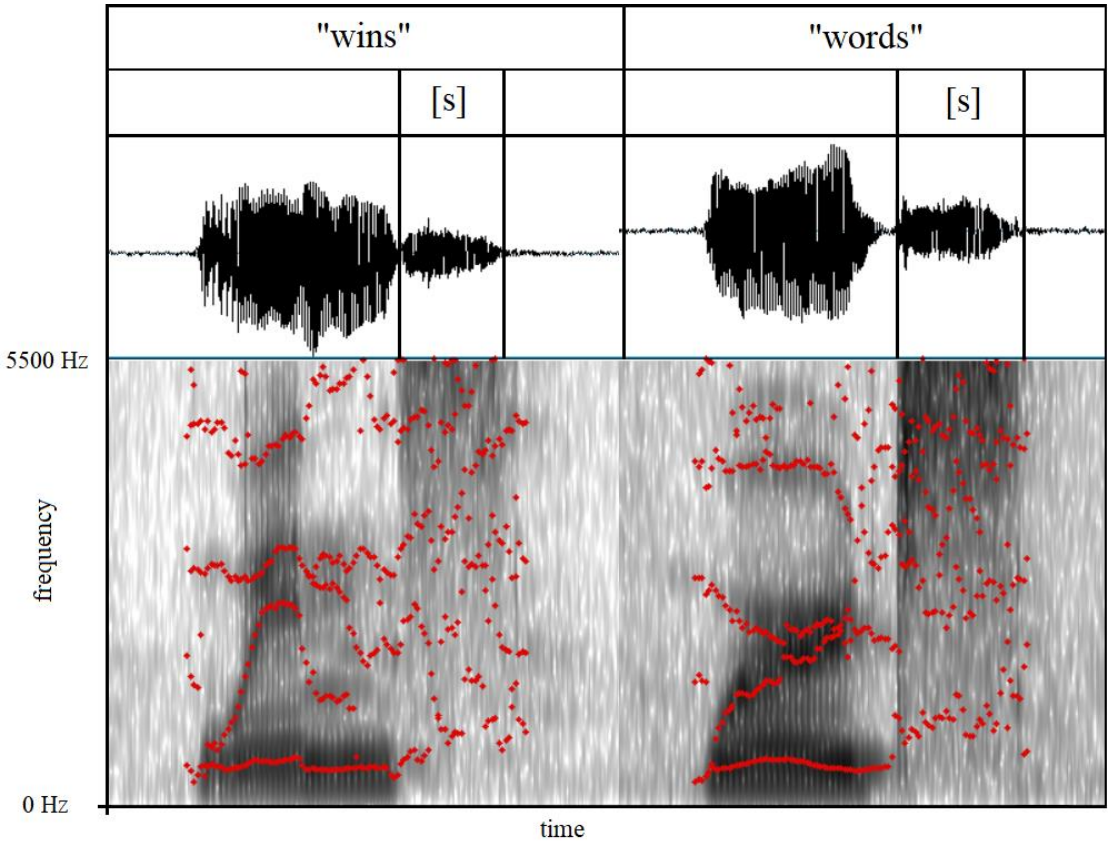

Figure 11. Waveforms and spectrograms showing devoiced word-final sibilants in the words *wins* (left) and *words* (right).

Lack of voicing in figure 12 is evident first and foremost from the absence of energy at the f_0 region, but also from the "broken" formant structure at the last segment (represented by the scattered red dots). This is particularly true of the word *wins*, which shows a total dispersion, while in *words* the f_1, f_2, f_3 and f_4 groupings remain more intact, reflecting lesser airflow turbulence. In voiced sibilants, the formants are usually less scattered and show a consistent pattern, though both modes of phonation exhibit a markedly aperiodic waveform pattern.

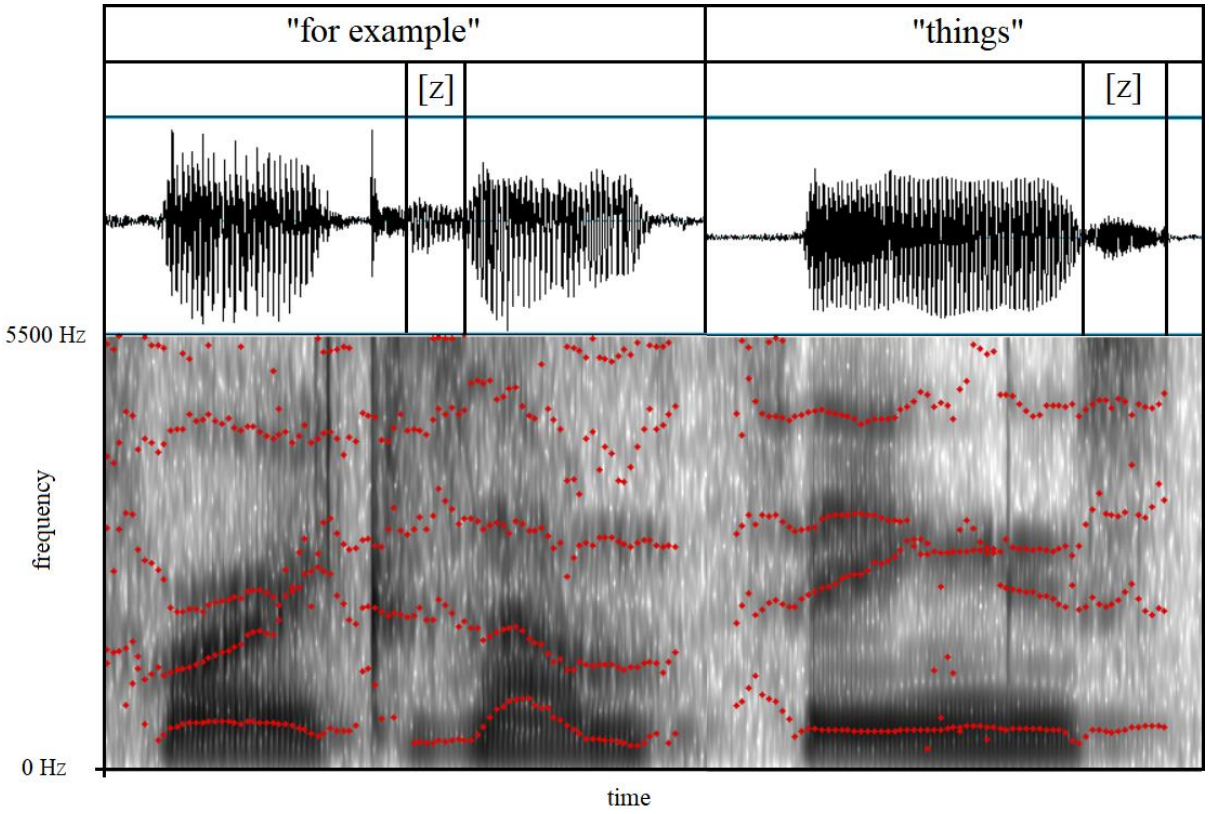

Figure 12. Waveforms and spectrograms showing voiced clusters in codas of words *for example* and *things*.

Contrary to expectations, participant 2 alternated between voiced and unvoiced sibilants, devoicing the final segments in *wins* and *words*, while voicing the corresponding segments in *example* and *things*. The reason for this pattern is not obvious since devoicing is rather counterintuitive in words such as *words* and *wins*, which have a voiced alveolar plosive and voiced alveolar nasal, respectively, preceding the word-final sibilant. Further, *wins* finds a minimal pair in *wince*, contingent on the voicing of the final segment. Similarly, in *things* the final segment is preceded by a voiced velar nasal, yet in this word the sibilant was voiced. In words such as *example* on the other hand, the grapheme <x> should be intuitively devoiced by L1 Finnish speakers, yet the segment was clearly realized as [gz] instead of [ks] (see figure 13).

4.5.3 Audible Release of Word-final Plosives

It can be difficult to ascertain whether released word-final plosives are an indicator of absence (of the feature [+unreleased]) in the underlying representation of L2 speakers, or rather a by-product of stilted or emphatic speech styles. However, participant 2's consistent American

English phonology caused the inaudible release to be a feature of interest. Indeed, deletions[13] of consonants, especially that of the phoneme /t/, are common particularly in GA (Odden 2005: 32). Participant 2 was heard optionally applying this feature correctly, and other times releasing the word final /t/ in utterances in which it did not sound natural. One such example is releasing the /t/ at utterance-final function words such as *it*. With the exception of cases where the pronoun is emphasized for some specific purpose, L1 speakers rarely have a reason to say [ɪt] in connected speech. Instead, the surface representation of *it*, in utterances such as *I see it*, is most likely [ɪt̚], arguably even in situations in which the speaker is nervous, while the underlying representation retains the form /ɪt/. However, unreleased or inaudibly released plosives are not limited to function words but are all commonly present in content words.

The words selected for the production task were ones with a high degree of optionality in the release of the word-final plosive, mostly /t/, in the classroom. These words, and an example of an utterance in which they appeared, are presented in table 9.

Table 9. Words that were observed as having most optional use of unreleased plosives

Word	Phrase	Classroom	Experiment
translate	*I had to translate**	[t]	[t̚], [t]
planet	*save the planet*	[t], [t͡s]	[t͡s]
bit (adverb)	*quite a bit*	[t]	[t]
start	*whoever wins gets to start*	[t̚], [t]	[t]

* The full phrase in the production task was *I had to translate, for example, save the planet*. This phrase, it should be noted, has three instances in which the word-final plosive is likely to be unreleased or assimilated: /d/ followed by [tʰ], /t/ followed by [f] and /t/ as utterance-final.

As indicated by table 9, the investigation of unreleased consonants focused mainly on the phoneme /t/. In contrast to other plosives, such as /p/, /t/ is subject to more phonological processes, allowing for more allophonic realizations such as t-voicing, t-glottalization and dentalization. Due to labiality, /p/ is restricted to fewer alternations, most often [p̚], [pʰ] or [p].

The phrases *I had to translate* and *save the planet* were both part of the same utterance. Interestingly, this utterance was repeated by participant 2 with the modification that the

[13] Technically speaking, inaudible plosives are not deleted, though many listeners would judge them to be. Evidence for this is provided by imaging studies showing overlapping articulation of a plosive and the segment following it whereby the obstruent closure phase is initiated but the release phase is never completed, hence making it inaudible. For an X-ray study, see Browman & Goldstein (1990), and for an electropalatographic study, see Nolan (1992).

unreleased allophone of /t/ in *translate* was present in the first iteration and in the repetition the /t/ was released. *Translate* with unreleased final /t/ and the repetition with released final /t/ are presented in figures 14 and 15 respectively.

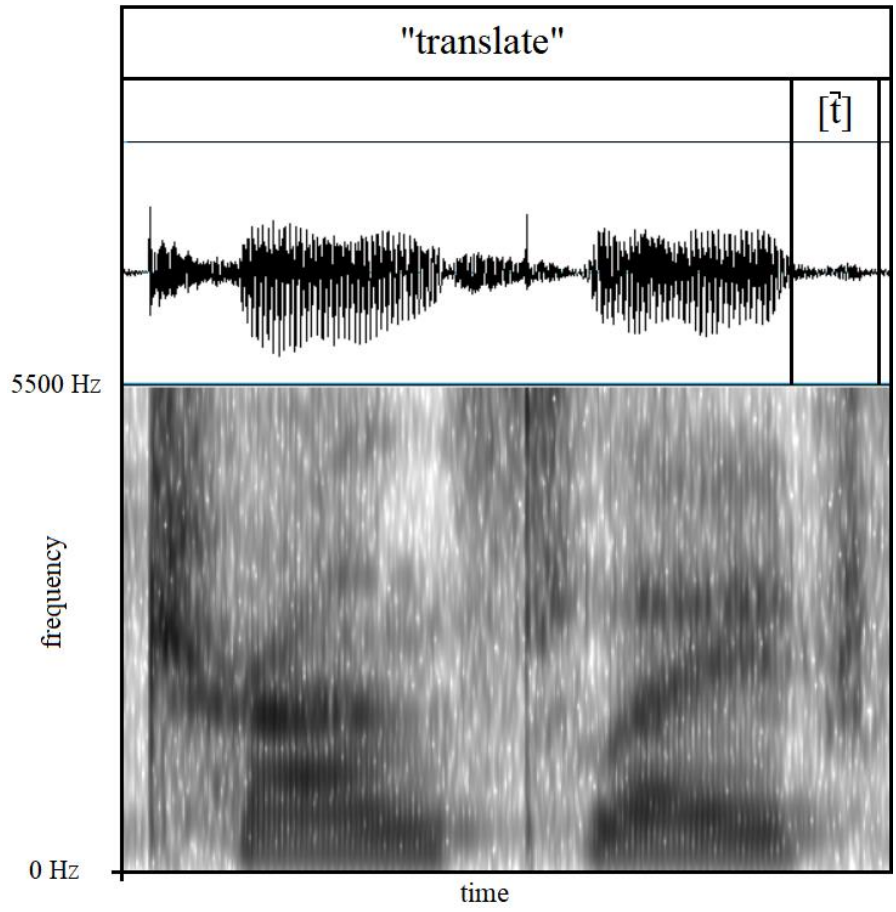

Figure 13. Waveform and spectrogram of *translate*, showing inaudible, or unreleased /t/, in the coda.

There are three acoustic phases in the articulation of plosive consonants that are distinguishable in a spectrogram: closure phase, silent phase, and release phase. The closure phase involves creating an obstruction in the oral cavity, causing a build-up of air behind the articulators, characteristic of obstruent consonants. When the air is released, the plosive can be said to have an audible release, and plosives become perceptible only upon the completion of the release burst. The three phases are all visible in figure 15, but not figure 14 in which the silent phase may be seen in the indicated margin, followed by aperiodic sound. In figure 15, on the other hand, there is an obvious release signature, represented by energy across the entire frequency band (0-5500 Hz). These are typically accompanied by strong aspiration when the release phase is completed. In figure 14, the release phase has been deleted likely due to an effect of connected speech.

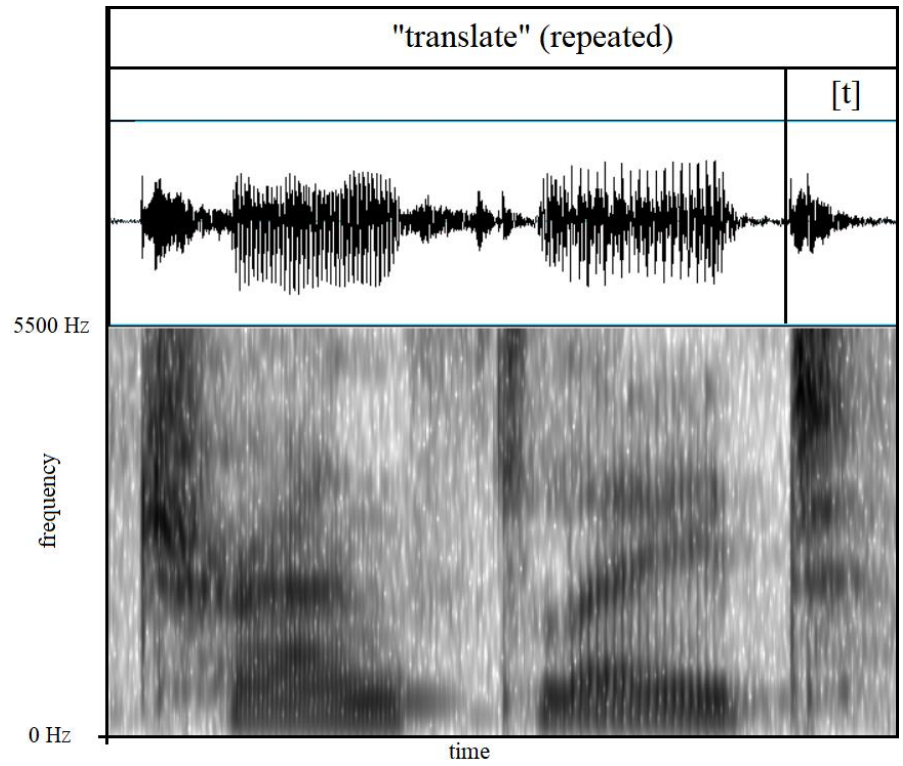

Figure 14. Waveform and spectrogram of the repetition of *translate*, showing audible [t] in the coda.

The reason for the utterance's repetition may lie in its length, or perhaps the unreleased consonants themselves. Indeed, participant 2 may have conscientiously repeated some of the words in the utterance in order to increase their clarity. In other words, she may have opted for hypercorrect forms. However, that is not relevant as the first produced version of the utterance implies acquisition of the phonological category in addition to its correct application in surface forms and appropriate phonetic environments. Further, in a separate phrase, *can you translate it*, participant 2 realizes the /t/ in *translate* appropriately as [ɾ], demonstrating elaborate acquisition in the phonemic category in question.

Planet on the other hand did not have any modification upon repetition, featuring audible bursts in both iterations, in which the /t/ was also qualitatively similar, exhibiting some affrication to produce a word-final sound akin to [t͡s]. This is illustrated in figure 16.

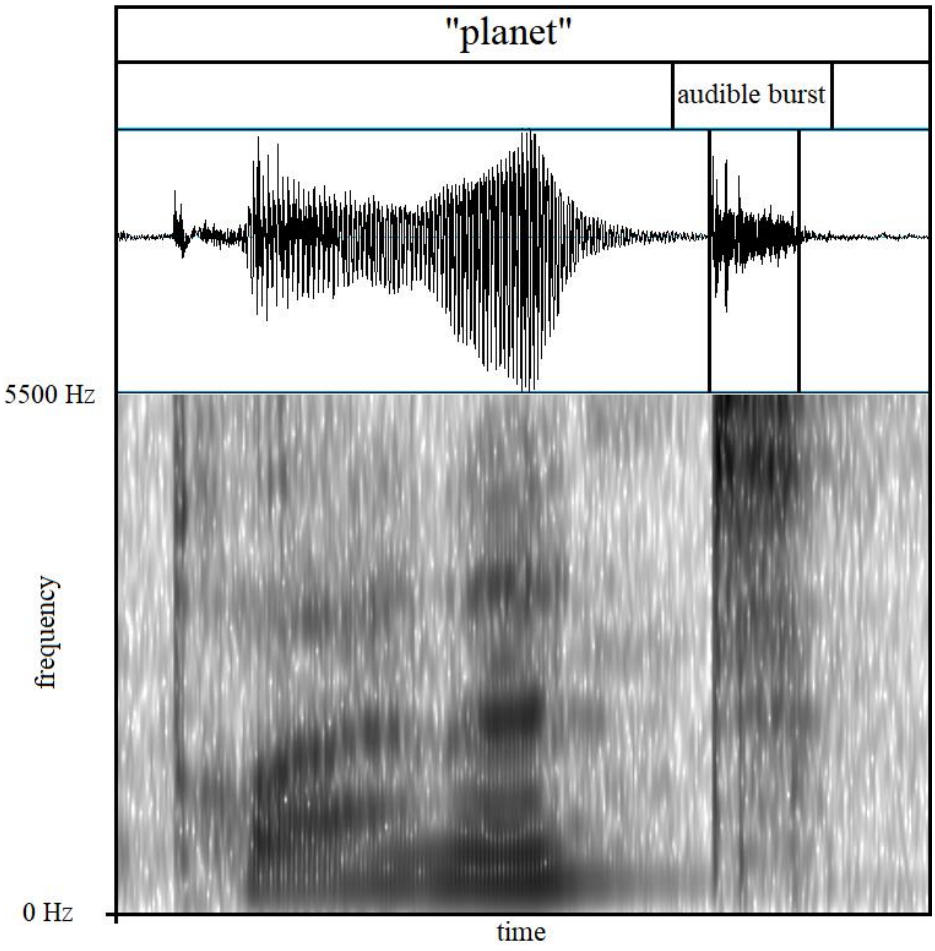

Figure 15. Waveform and spectrogram of *planet* with a release burst at the last segment [t].

While the inaudible allophone of /t/ would be expected in the word in question, especially for a speaker of the North American varieties, an alveolar affricate is also heard as occurring allophonically, both in onsets and codas, in free variation with an aspirated /t/ (Wells 1982: 515). In codas, /t/ produced as [t͡s] may be heard in, for instance, *tea* – [t͡sɪi̯] – in some varieties.

It is unclear why the phoneme /t/ was always audible in *planet* while *translate* shows allophonic variation. However, it is possible that participant 2 has internalized a phonological rule that determines utterance-final plosives, particularly /t/, to always be audible. As regards *bit* and *start*, the coda segment was quite clearly released as a strongly aspirated alveolar plosive, as represented in figure 17.

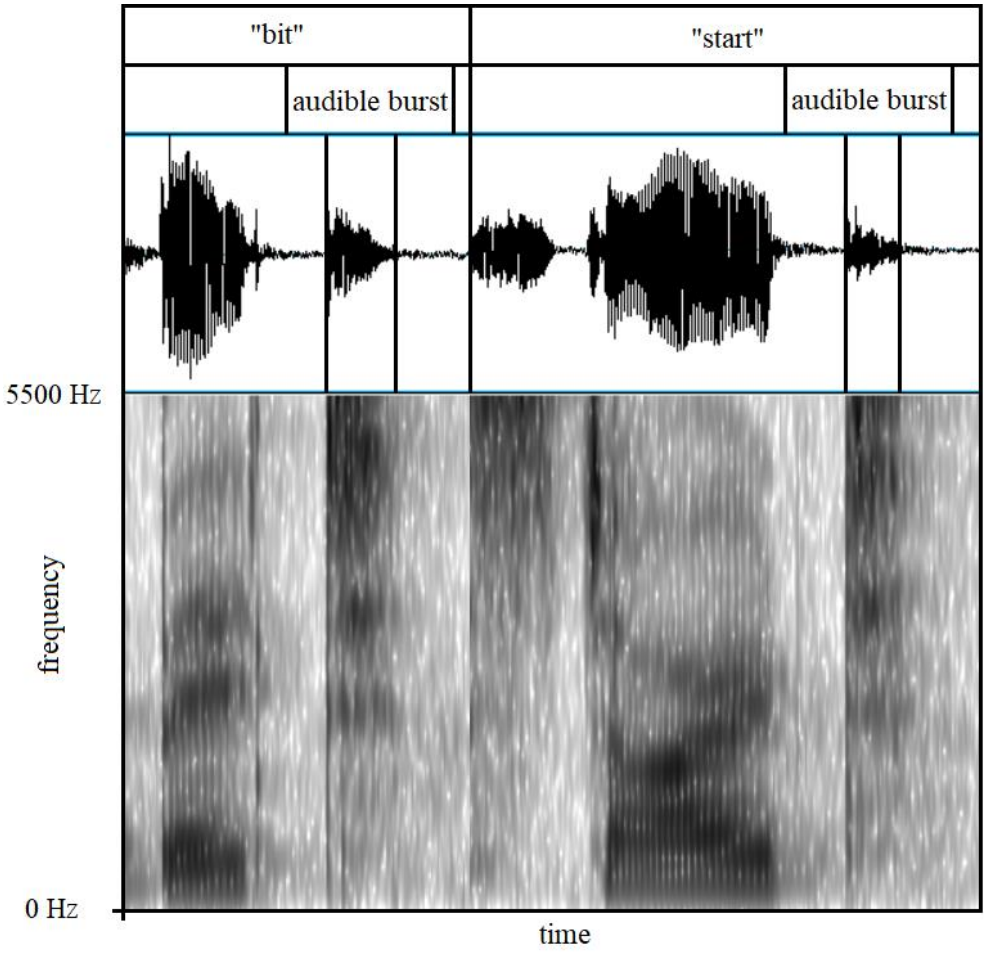

Figure 16. Waveforms and spectrograms of *bit* and *start*.

A problem for interpretation arises regarding whether to judge the final releases as non-target. Though audible releases in utterance-final positions as a systematic feature in American L1 English phonologies could be judged as somewhat peculiar, it cannot be ruled out that some L1 speakers might systematically release plosives in final positions of either specific words or utterances in general. However, this would be purely idiolectal and not relevant in the framework of this thesis in which it is presumed that failure to realize allophones in appropriate environments is a function of processing limitations.

4.6 Receptive Task – Participant 2

Between Participant 1 and 2, there was overlap in two of three tested phonological categories, namely those of voicing and velarization. The same contrasts in the same phrases were thus included in the listening tasks for both participants, the only difference being l-vocalization and lexical boundary assimilation for participant 1 and word-final plosives for participant 2.

Details of participant 2's performance are summarized in table 10 with production task results for comparison with the same phrases.

Table 10. Summary of participant 2's receptive task results with production task results included for comparison with the same phrases

Contrasts	Phrase	Choice	Production Task
[z], *[s]	*I'm zipping my jacket*	[z]	
[z], *[s]	*lots of words*	*[s]	*[s]
*[z], [s]	*I'm sipping from this mug*	[s]	
[z], *[s]	*things that I like*	*[s]	*[s]
*[l], [ɬ]	*basically*	[ɬ]	[ɬ]
*[l], [ɬ]	*tell*	[ɬ]	[ɬ]
*[l], [ɬ]	*football*	[ɬ]	[ɬ]
*[t], [t̻]	*I have studied quite a bit*	[t̻]	*[t]
*[t], [t̻]	*that's great, thank you*	[t̻]	*[t]

The Contrasts column indicates the contrasts in each phrase and the "erroneous" phone or allophone is marked with an asterisk.[14] Similarly, the columns for Choice and Production Task indicate what was chosen or produced with the erroneous choices marked with asterisks. In the recording of the phrase *that's great, thank you*, an intentional pause was included between *great* and *thank you* in order to make it obvious that there is no co-articulation occurring at the lexical boundary.

[14] Not all the choices are necessarily erroneous but are nevertheless classified as such because they would not be expected in regular L1 speech. The allophonic contrasts [t] and [t̻] in the listed phrases are both possible in L1 speech, but the former can hardly be said to feature in these phrases unless there are specific emphases or speech styles. In the audio recordings, the speaker does not emphasize any elements except for syllables that are normally stressed. In other words, the recordings are spoken in natural and casual sounding style in order to make the choice for the erroneous allophone as counterintuitive as possible.

4.7 Interview – Participant 2

The same battery of questions used to interview participant 1 (see 4.4) was presented to participant 2 whose responses are paraphrased below.

4.7.1 Attitudes on Pronunciation Inaccuracies

She stated that approximating native speaker phonology is not important for L2 learners and does not strive to achieve such a standard herself. Regarding English teachers, she stated that teachers should master certain features, citing aspirated plosives as an example, but so long as comprehensibility is not hindered, retaining some L1 vowels is acceptable. She added that, to a degree, L1 transfer in pronunciation is more acceptable than in other areas of L2 production. When asked to assess the claim that L1 segments in L2 speech equates with erroneous pronunciation, she strongly disagreed. Whether the same principle could be applied to other areas of grammar, such as syntax, was difficult to answer and required much reflection. In her mind, phonology and other areas of grammar are not comparable because, in her words, syntax is more about maintaining well-formed and comprehensible structures, which become beneficial, for instance, in professional life.

As regards speaker identity, participant 2 did not consider sounding like a native speaker to be problematic. This question was particularly germane as participant 2 can sound indistinguishable from a North American native speaker over multiple consecutive utterances. Her reasoning was that people have a tendency to acquire a speech style that they hear consistently, and it is therefore only natural to speak whatever variety one has been exposed to, regardless of proficiency. Regarding her own spoken variety, she explained that the GA phonology is easier and "fits in her mouth better", adding that if she were to speak like a British person, it would feel superficial and fake. She further elaborated that she spoke an RP variety of English while in high school, but later acclimated to American English phonology after prolonged exposure to American media.

4.7.2 Pronunciation Strategies and Performance Anxiety

When asked about performance anxiety before and during practice lessons, participant 2 stated that she had been extremely nervous and that she no longer remembered most of them, with the implication that the stress of anticipating and teaching the lessons has led to weakened recall of their contents. Her speech capacity is affected by stressful situations in the form of delayed word retrieval and "speaking more than one should", she explained. She reported

feeling less nervous about teaching the more lessons she had taught, contrary to participant 1's experience.

Regarding the question of the monitor, participant 2's response suggests that she hardly monitors her own production. She described the fact that the phonetic information of most words in her vocabulary is fully acquired and thus largely subconscious: citing the word *internet* as an example, only a few words require conscious reflection regarding their pronunciation.

4.8 Summary

4.8.1 Participant 1

The interview did not reveal any barriers to approximating native-like L2 phonology, as participant 1 stated not having any reasons, ideological or otherwise, to purposely misrepresent his phonological competence. It may be concluded then that the pronunciation that was observed was representative of "genuine performance". Further, sounding as native as possible is rather important for participant 1, though fluency is his first priority. He reported having felt nervous about teaching practice lessons in English.

Participant 1's tokens for the production and perception tasks were the voicing in sibilants and consonant clusters containing either sibilants or voiced plosives and the velarization of laterals, or the clear-dark contrast of /l/, in addition to vocalization of /l/ and assimilation. Regarding the allophones of /l/, participant 1's performance was consistent when comparing classroom and experimental data. This was particularly true of word-initial positions. However, one exception was the word *tell*, which participant 1, correctly, articulated with [ɬ] in the syllable coda in the experiment task, contrary to what was observed in the classroom. In the receptive task, participant 1 selected the more accurate option in all stimuli with the contrast in /l/.

Regarding the vocalized /l/ in *false*, participant 1 produced a clear and audible [l], contrary to target norms, not contradicting what was observed in the classroom. However, he judged vocalized /l/ to sound more target-like than the clear /l/ he consistently produced.

The assimilatory process at word boundary of *even* and *though* did not differ between classroom and production task, in which participant 1 produced the non-target substitution of

a fricative for plosive ([ivɛntoʊ] instead of [ivɪn̪ðoʊ]). However, after listening to the contrast, participant 1 selected the target, as opposed to the non-target option in the receptive task.

In the case of sibilant and cluster phonation, the words *exactly*, *exhausted* and *because* (respectively, [gz], [gz] and [z]) were the most significant items. It was interpreted that the phonation in these words differed between classroom and experiment data with production task yielding target-like production. This is considered the most significant example of context-dependent variation, matching expectations. Likewise, all stimuli bearing the contrast between /z/ and /s/ were judged correctly in the receptive task.

In summary, participant 1 has acquired all the contrasts as far as comprehension is concerned, but was unable, in place of some features, to produce them consistently in a target-like manner. The production results showed, however, that the experimental situation may help to produce the target contrast more often. The reason for the apparent helpfulness of the experimental situation is most likely the absence of communicative pressure or nervousness, or both.

4.8.2 Participant 2

Participant 2's interview indicates that she does not consider native-like pronunciation as an important goal for herself nor for L2 learners, but does not object to students trying to sound as native as possible, which is important when attempting to capture genuine L2 competence. The experience of teaching English in a classroom was highly stressful for her.

Participant 2's tokens for the production and perception tasks were the voicing in sibilants and consonant clusters containing either sibilants or voiced plosives and the velarization of laterals, or the clear-dark contrast of /l/, as with participant 1. Additionally, word-final audible releases of plosives were chosen for participant 2 based on classroom performance.

In the case of the laterals, articulation of /l/ in all five words (see table 7) was dark, as was anticipated. Three of the five words, namely *please*, *basically* and *hopefully*, were heard as having clear articulations at times in the classroom, i.e., there was optionality in their realization. Two words in the task, *astrology* and *football*, were always heard as having a dark articulation in the classroom and were unsurprisingly articulated in the same way in the production task, but showed also a stronger velarization compared with the words which had an optional realization of /l/. This was interpreted by the differences in f_2 values, which was especially pronounced in the word *football*. The receptive task results fall in line with the

entirely target-like production of participant 2's laterals: she perceives the contrasts in this token and judges the velarized /l/ as being more correct, consistent with her own production.

The results from the production of the voicing token were more mixed: the codas in *words* and *wins* were articulated as [ds] and [ns] instead of [dz] and [nz], respectively, falling in line with the non-target classroom output. Since there was optionality in [ds] and [dz], and [ns] and [nz] in the classroom, this result was unexpected as it was anticipated that sibilants in such clusters would be voiced by the participant if given time to plan and self-evaluate the output. In contrast, the clusters in the words *example* and *things* were articulated as [gz] and [ŋz] respectively, supporting the hypothesis that, if optionality exists, the target form would be opted for over the non-target form in the experimental situation. As regards the receptive task, participant 2 demonstrated a capacity to perceive the /s z/ contrast, but nevertheless judged [ds] to be more correct than [dz] in the phrase *lots of words*, as well as [ŋs] over [ŋz] in *things* (the full phrase was *things that I like*) as they were also produced in the experiment, suggesting that this is how the phonetic segments in the analyzed words are genuinely represented.

Unreleased word-final plosives, a feature which typically finds its target and non-target realizations in a speech chain rather than singular words, were included in four carrier phrases (see table 9). Here participant 2 almost exclusively opted for released allophones, some of which were strongly aspirated to the point of exhibiting /t/ affrication. However, one instance of an utterance shows the [t̚] allophone in *translate*, after which the phrase was repeated with a [t] in the same word. The findings regarding this token may be flawed because of the study design, which will be discussed in the following chapter. Interestingly, in the phrases *I have studied quite a bit* and *that's great, thank you,* she selected the unreleased plosives (in the words *bit* and *great*) as the better options in the receptive task, which is contrary to what she herself produced in the production task.

5 Discussion

The purpose of this study was to investigate the possibility of performance errors, rather than "impaired representation", underlying non-target phonological surface forms. If the error was at the level of representation, this would implicate L2 competence instead of performance. The reasoning is that if there is optionality in how surface forms are realized, especially in a communicative situation as opposed to the experimental one used in this study, specifically the pressure to communicate might be the likely explanation for the inconsistent production of target-like L2 speech. In cognitive terms, the component of the mind that is responsible for applying grammatical rules into L2 speech is the working memory (WM), whose processing capacity may be exceeded when constructing grammatically complex utterances without enough time for pre-planning. In more proficient and experienced L2 speakers, the production of utterances alone may not strain the WM's capacity most of the time, but external factors, such as psychological stress, may temporarily impair or otherwise limit that capacity. Accordingly, the study design was aimed at creating an experimental situation in which the speakers may plan their utterances and produce them in a way that best reflects their own conception of how correct the utterances' phonetic structures are. What was produced in the experiment matched the non-target pronunciation observed in the classroom. The results from this experiment were then compared with the context in which communicative and psychological pressure were present (i.e., the classroom context), with the presumption that possible differences might provide evidence of a WM-based "misrepresentation" of the L2 speaker's genuine competence. The study design's purpose was to provide answers to the following research questions:

1. Does the nature of the communicative situation affect phonological production, giving rise to inaccurate representation of L2 phonological proficiency?

2. When production is elicited in an experimental situation where the production is not part of a communicative situation, does this affect phonological output in comparison with a communicative situation?

To answer the research questions, several hours of oral L2 English production by the selected participants were recorded while they were teaching English while using English as the language of communication in a Finnish high school. Based on the analysis of these recordings, a production task was designed with a list of specific words whose phonetic realization were interpreted to be non-target in the classroom recordings. In other words, the non-target forms observed in the classroom were to be repeated by the participants in a context not involving communication in English. Indeed, the participants completed the production task in a phonetics laboratory by reading out loud from a list with permission to pre-plan production and repeat the utterances, if they so wished. These elements, in addition to the unlimited time to complete the task and the instructions to produce utterances as target-like as possible, were crucial for the experimental situation. The participants' performance in the production task was recorded with a high-quality microphone and analyzed using Praat, a phonetic speech analysis program.

A receptive task with the same tokens was designed to confirm how the participants themselves represented the forms that they produced. The recordings played back to the participants included a target-like pronunciation of specific words in addition to a second version of same word with a non-target pronunciation that mimicked the participants' own pronunciation as observed in the classroom recordings. In order to account for possible ideological constraints that might have influenced phonological performance in the classroom, the participants were interviewed about their beliefs regarding how target-like English teacher pronunciation ought to be, and how they expect themselves, as teachers, to ideally perform in terms of L2 phonological production.

The classroom observations and the production task in conjunction with the receptive task were designed to answer the first research question. The interview was designed to provide additional information regarding "the nature of the communicative situation". In other words, how stressful the situation was and to what extent such situations generally affect pronunciation in the participants' opinions. The answer to the second research question is answered more concretely by comparing the classroom observations to the acoustic analysis of the production task. The next section focuses on how the results from these analyses were interpreted.

5.1 Interpretation of the Results

Considering the results of both participant 1 and 2, it should be noted that performance in the production task was not consistent across tokens. Instead, some tokens differed in how they were produced in the experiment as opposed to classroom, while other tokens showed an inverse pattern. Further, performance in the case of certain categories of tokens, such as the voicing of consonants for participant 2, was inconsistent within that category. This inconsistency is an important finding as it could reveal that the hypothesis related to processing demands may not equally apply in all of the tokens investigated. Crucially, however, participant 2 judged the <s> to sound more correct when it is [s] than [z], when presented with the contrast in two carrier words, *things* and *words*, in the receptive task. Many native speakers would represent the last segment in the inflected nouns as having [+voice], though this is not necessarily the case for all native speakers. The consistency between production and judgment of these words suggests that this is how participant 2 genuinely represents them. Judging by the classroom data, it appears that the phonetic quality in inflectional morphemes as represented by the grapheme <s> is underspecified for voicing and may be in free variation between [z] and [s].

The results of the production task in the case of some tokens selected for this study may not be equally reliable from the standpoint of the theoretical framework, since the participants' production may have been influenced by other factors such as flaws in the task design or the testing situation itself (see 5.2.2 and 5.2.3). For instance, the testing situation may increase the formality of the test subject's speech, causing the articulation of words such as *walkin*' to become *walking* (Fischer 1958: 49). The deletion of the release phase in word-final /t/ may have been an instance where the perceived formality of the situation influenced the articulations, since the release of a plosive consonant is a feature whose presence or absence is linked to connected speech, much like the *-ing* suffix reported by Fischer (ibid.). Nevertheless, it would follow from this interpretation that the classroom situation was also affected by formality, considering its consistency between classroom performance and experimental performance by participant 2. At any rate, the variation is a case of careful connected speech as opposed to natural connected speech. To some degree, judgments by the participants in receptive tasks may ameliorate the misrepresentations that potentially arise from speech production by the participants themselves.

The tendency to consistently produce one allophone over another in an experimental situation as opposed to optional production in a situation involving communicative pressure does seem to support the study hypothesis. For instance, participant 2 only produced the highly velarized /l/ in the experimental situation while the classroom data suggests that there was some optionality in the articulation of /l/. In contrast, the failure to produce a voiced dental fricative in a specific phonetic environment appeared to be consistent regardless of situation, as was the case for participant 1, who additionally judged the same token as non-target when uttered by a different person in the receptive task. This indicates that L2 speakers may consistently realize certain phonemes incorrectly despite having the correct underlying representation thereof.

The comparison of the receptive and production task performance suggests that perception of the L2 is more accurate than production, which was true for both participants. The receptive tasks may then be interpreted as providing a more accurate picture of how participants underlyingly represent their own phonology, as opposed to surface forms.

In summary, some tokens worked better than others in lending or decreasing support for the study hypothesis (see 5.2.6 for a more detailed discussion regarding the investigated tokens), which should be taken into consideration when devising similar studies in the future. For both participants, the tokens involving the allophones of /l/ and voiced segments which were often unvoiced in the classroom context are interpreted as the strongest indications of context-dependent variation in phonological production. Taken together, however, the evidence provided by the results are somewhat inconclusive with respect to the hypothesis. Further research on more phonemes is required in order to establish a clear picture of their representation and production in different contexts.

5.2 Study Limitations and Reliability of the Results

The first limitation of this pilot study that should be addressed is a theoretical one. It should be noted that the exact manner the mental grammar interacts with WM and how grammatical information is relayed through different networks before it is produced as speech is difficult to demonstrate and thus remains poorly understood. Thus far, aside from neurological research, all that has been produced in the linguistic literature are theoretical models as to how this system might operate. For instance, Hawkins (2019: 287) offers a model based on previous authors' contributions, in which WM is a domain-specific process which interacts, as

appropriate, with UG and other linguistic elements of the mind that are simultaneously in bidirectional interaction with short-term and long-term memory. Since empirical studies in linguistics have been of little help in demonstrating how this system works, researchers are at present restricted to the level of theory and hypothesis. This limitation should be acknowledged with respect to the investigation that was carried out for this thesis.

5.2.1 Reporting and Analysis of Audio Data

A major limitation of this study was in the disparity of documentation between the classroom production and the experiment production. No comparable acoustic analyses were conducted on the classroom production due to limited availability of recording equipment. This deficit was compensated for by selecting only the most unmistakably non-target segments for the experimental production task, in other words, segments that could be confirmed to be inaccurate in terms of target language norms by listening to recordings from the classroom English lessons multiple times. Secondly, the segments and the words in which they featured were judged to be inaccurate in multiple instances with some optionality. Thirdly, more than one person judged the recorded segments to be inaccurate, one an L1 English speaker. Taken together, it may be stated with confidence that what was perceived to be instances of non-target speech were indeed non-target.

5.2.2 The Elicitation Technique

Regarding the production task, asking the question whether the elicitation yielded surface forms that genuinely represent underlying competence reveals a potential flaw in the design. Though the participants were informed that they should try to produce the list in the most representative and natural manner possible, inattention to accuracy, and thus the production of inaccurate forms, remains a possibility. A possible work-around to this would be to have participants listen to their own production and judge its accuracy instead of listening to a native speaker produce the contrasts, which was done in this study. However, this is not to say that the receptive task in this study was inappropriate, but quite the opposite: its sole purpose was to reveal possible contradictions between what the participants produced on the one hand and judged to be accurate on the other.

Additionally, the elicitation technique used in the study, namely the reading list, could also be open to criticism. For instance, Ladefoged (2003: 23) argues that speakers are “apt to read with a different pronunciation from that in their normal speech” and that a more natural

pronunciation may be obtained "by giving a prompt in English, or an equivalent in the contact language being used, and then having the speakers respond by saying the required word in their own language". Instead of using reading lists, Ladefoged recommends other elicitation techniques such as naming objects in pictures. However, this study was designed to elicit a product of the underlying phonological representation from the participants and not "natural speech" necessarily. Further, according to Chambers (2008: 6), people tend to "concentrate on their pronunciation almost completely, especially when the reading is being recorded by someone who is admittedly studying language". This could be interpreted as an elicitation of a product of the correct representation, as opposed to performance errors. However, Colantoni and colleagues (2015: 110) advise against using stimuli in written form in language that differs from the participant's L1's grapheme-to-phoneme correspondences, further adding that error-free reading aloud can be difficult even for L1 speakers. Such correspondences should be considered as valid concerns regarding the present study, as some of the tokens included words with consonant clusters whose pronunciation does differ between Finnish and English in features, such as voicing (for instance, *exhausted*). Moreover, Finnish orthography is almost entirely phonemic, whereas English orthography is not.

5.2.3 The Testing Situation

The experimental situation may be criticized in at least two ways: firstly, testing situations distort "normal" pronunciation. Secondly, the study subjects' awareness of being tested may itself cause psychological stress, which is a variable that was claimed to be eliminated in the experimental situation of this study.

Regarding the first point, the way speech is incorrectly represented in testing situations is commonly understood to be associated with the apparent formality of the situation. This may manifest in the addition of segments in words whose surface forms lack them in casual speech. For instance, Fischer (1958: 49) discovered that Boston schoolchildren pronounced the *-ing* suffix in participles such as *walking* as [ɪn] in informal communication, but in a testing situation switched to using [ɪŋ]. Bearing this in mind, a similar distorting effect could be expected to have occurred in participant 2's production of the unreleased plosives, though an attempt was made to pre-empt this by including the tokens in longer carrier phrases instead of singular words. Aside from unreleased plosives, however, it is unlikely that the formality of the testing situation significantly distorted the other tokens.

As for the second point, the awareness of performing a task in the presence of a language researcher is likely to cause pressure to perform, which could be interpreted as potentially causing psychological stress. The awareness of having the performance recorded may also exacerbate that stress. To reduce the stressfulness of the situation, the participants were encouraged to not feel nervous about the situation. Furthermore, they were explicitly informed that the intention behind the testing situation is, specifically, that the participants should not experience psychological stress.

5.2.4 Optionality and Psychological Factors

This study is centered around the idea of psychological pressure affecting phonological processes. Here psychological pressure may mean psychological stress, but may also be taken to mean the obligation to communicate using L2 in situations which call for spontaneous L2 speech. Furthermore, unless the command of the L2 is highly advanced and the speaker is highly confident about his or her speaking competence, one may expect that one or more of the automatic processes are processed at the conscious level. Such conscious, or non-automatic, processing may arise from self-consciousness regarding pronunciation or doubts about one's own proficiency. At any rate, conscious processing of this kind may be expected to be processed by working memory.

As predicted by the Missing Surface Inflection and Morphological Deficit hypotheses, when WM runs out of capacity, "breakdowns" occur in speech. If breakdowns occur on the phonological level, it could be expected that L2 speaker phonologies differ in situations where communicative pressure is present and in situations without any such pressures. Optionality, or the failure to apply a given phonological rule at one time but not the other, could be interpreted as an indication of this phenomenon.

It was predicted that phonological optionality would be demonstrated in speech samples recorded in classroom situations. It was further predicted that, if such optionality were caused by psychological factors such as stress, the pronunciation would show a stabilization effect, i.e., optionality decreases in direct proportion to psychological stress. Psychological stress presents a problem, however, in the difficulty to quantify it. On the one hand, it is possible to quantify it as a physiological response by measuring heart rate or cortisol levels in saliva. On the other hand, subjects may be interviewed about their subjective experience of psychological stress and how it is experienced on a physiological or cognitive level. The limitation in this study was that no reliable measures of the physiological stress response were used.

Additionally, the notion of psychological stress and communicative pressure affecting speech production is presumed to be true *prima facie*. However, how stress affects individuals is likely highly subjective and varies from one individual to the next. Thus, even if the stress response could be measured physiologically, there is no guarantee that this will affect specific cognitive processes, such as language, in all individuals equally. Indeed, it is reasonable to assume that, in this sense, some are more resilient than others. Stress as a dependent variable, as poorly as it is defined and quantified, is a significant weakness in this study.

5.2.5 Selection of Participants

English teacher students were recruited for this study because their training was considered beneficial for the study methods, namely the provision of an elicitation situation with communicative pressure, but also due to their presumed proficiency: to be admitted in the teacher training program, the students must have demonstrated a highly proficient level in English by taking a test provided by their university. Thus, the participants' admission to teacher training was taken as evidence of aptitude in the recruitment process, eliminating the need to separately test for proficiency. However, testing for proficiency is not the same as testing for aptitude. Accordingly, the lack of testing may be considered a limitation here as it was inferred that the participants had above average aptitude by virtue of being enrolled in the language teacher training program, in addition to being university graduate students in English. However, this is not a reliable indicator of aptitude as majoring in English does not automatically imply linguistic giftedness in L2 learning.

In addition to L2 speaker participants, the inclusion of a native speaker control group would have been beneficial, as it could have provided information about how robust the testing methods and tools were. In the case of this study, these would have been the production and receptive tasks. Indeed, in studies investigating L2 speakers, L1 speaker controls are normative.

5.2.6 The Tokens

The analysis of the phoneme types in this study suggests that some are more suitable than others for the investigation that was conducted. For instance, the categories whose realization is largely conditional on the speech chain, namely the inaudible release of plosives and assimilatory processes at word boundaries, may not be as reliable as phonemes whose realization is mostly unaffected by connected speech, such as the fricative consonants that were

used as stimuli in this thesis. Further, not all tokens were tested in equal measure, which should be corrected in the future. For instance, the lateral approximant category had several stimuli in the production and receptive tasks, yet L-vocalization only had one stimulus in both tasks.

The inherent differences between speech sounds should also be taken into consideration in testing, which was neglected in this study. For instance, both voiceless and voiced dental fricatives, [ð] and [θ], are notoriously difficult for L2 speakers, as attested by non-target realizations by L2 speakers with various L1 backgrounds. Furthermore, it is known that liquid consonants are acquired later than other L1 sound categories by normally developing children, and that they pose difficulties for children with delayed speech development (Jakobson 1968; Strange & Broen 1980). Thus, inclusion of liquids as opposed to more "natural" sounds, in studies requiring oral production, could be questioned. Liquids include lateral approximants, which this study included as a token tested on both participants, and it is not known whether liquids have been particularly difficult for this study's participants to acquire as either L1 or L2 learners. In future studies, either individual learning difficulties in the case of specific speech sounds should be accounted for, or the more "universal" consonant and vowel phonemes, such as nasals or the close front unrounded vowel, should be favored over the ones that are more inherently difficult.

6 Conclusion

The hypothesis of this study was conceptualized primarily within the framework of working memory models whose operating principles were assumed to be perceptible in practice, specifically in the research design. The hypothesis is based on theories regarding the way in which language is represented by L2 speakers: being able to apply acquired grammatical markers in speech may not be a question of either having a correct or incorrect underlying representation (UR), but the correctly represented forms may temporarily be unavailable to the speaker, resulting in speakers' defaulting to underspecified[15] forms in the output. The hypothesis of this study is specifically related to phonology, and it is assumed that the phonological representation of specific phonemes or their realization in specific words may also not be a question of either having a correct or incorrect UR. Instead, it is presumed that regardless of whether the UR is impaired or target-like, it may not always be realized similarly in phonological surface forms, but may instead be conditioned by factors such as processing limitations of the working memory.

The key methodological features of this study included tasks for both production and perception, individually tailored to each participant based on several hours of perceived production in communicative situations, namely classroom English teaching. Based on the hypothesis and study design, it was anticipated that in certain situations, L2 speakers produce segments that are not accurate, but the same segmental errors in the same words and phrases will not be reproduced in situations in which the speaker is allowed to reflect on accuracy and possibly correct their production by way of repetition. The latter situation is thought to reflect genuine competence, while the former is thought to reflect performance errors. Cases in which differences arise between these two situations are mediated by working memory, provided that the WM hypothesis is accepted. However, even if the theoretical foundations of this study are

[15] The term underspecified may be understood here as either defaulting to a corresponding segment from the L1 inventory, or the failure to apply the appropriate phonological rule, such as a specific segmental feature in specific words, since phonological rules often must be acquired on a word-by-word basis in languages whose orthographies are phonetically unpredictable, English being an example.

not accepted by those appraising it, the reported findings are nevertheless interesting and may fit into alternative frameworks.

The most important findings of this study may be summarized as follows: L2 speakers do not produce language in the same way in all situations, and the perceived approximation to target norms varies according to context. This, it is interpreted here, occurs mostly due to lack of time to plan or focus on how well-formed the produced utterances are. This explanation also fits into Tarone's theory of multiple interlanguages (see 2.2.2).

6.1 Future Directions

One of the merits of this study is that it is easy to replicate with different participants as everyone has different phonologies, and the methodology can be applied to each individual's phonological system. However, a repeated study should include a native speaker control group with which production obtained from L2 speakers could be compared. It should also be mentioned that the segments that were analyzed were entirely dependent on the individual surface phonology of the participants that were observed while they were teaching English lessons. While numerous non-target segments were observed, many tokens had to be left out due to the considerable depth of the analysis conducted on individual segments produced by the participants. For instance, all vowel tokens were left out of the analysis, and the focus remained on the features of consonant phonemes. However, investigating the articulation of vowel phonemes is equally valid within the theoretical framework of this thesis.

As regards the study design and methods, the most important modification to the methodology used in this thesis is the way classroom segments are recorded. A replication of this study should be able to capture repeated instances of tokens (e.g., a specific word) in a situation that is equivalent to the classroom scenario in this study, and the captured audio must be of high quality. This may require a type of microphone that is attached to the speaking participant who may move while speaking. At any rate, a methodologically improved replication of this study is in the realm of possibility, though it would require some specific solutions of a technical nature, such as selecting the appropriate audio capture system or device. In addition to everything mentioned above, the potential limitations discussed in more detail in section 5.2 should be taken into consideration in future research with similar methodology.

6.2 Value and Contribution

This study involved both observational and experimental methods. Not only were the segments analyzed carefully, but several segments were analyzed, often including multiple stimuli per token. Moreover, the analysis accounted for not only the participants' output, but also the participants' perception and evaluation of the same segments uttered by others. Interviewing the participants about their L2 English production further helped to clarify possible factors behind their phonological performance. For instance, the participants were asked to explain what kind of monitor users they are and how gifted they are at L2 learning and other skills in their own experience.

Contrasting L2 speaker production with underlying L2 representations and investigating whether there is a possible disparity between how language is represented in the mind and how it is produced in specific communicative situations are issues that have been largely ignored in linguistic research. Instead, it is often either taken for granted or left undiscussed that what has been acquired in the mental grammar and what is produced in speech mostly works seamlessly. Alternatively, it may be assumed that acquisition is paramount and even though performance errors occur thereafter, they are inconsequential. However, this may not always be the case, as speakers may be judged based on their production in situations that indeed are of consequence. For instance, teachers, especially language teachers, may evaluate and grade students based on their L2 performance errors that do not reflect their competence.

In the practical context of language education which this thesis focuses on, teachers may be mistakenly judged as being less competent in the language they are teaching based on performance errors. Furthermore, perceived L2 competence in a professional context may be extended to a broader perception of competence. In other words, colleagues and supervisors may form unfounded perceptions: some employees are seen as more competent than others due to their L2 performance in workplaces in which a link between work performance and language skills may be insignificant or not exist. These issues will become increasingly pertinent in an increasingly globalized world in which English is the lingua franca and in which L2 English is likely to assume an important role in workplaces.

Research on L2 phonology and foreign accents generally are important topics to investigate as they are sociologically important. For instance, Toivola and colleagues (2008) claim that non-target pronunciation can stop otherwise linguistically proficient immigrants from gaining employment that matches their education or professional abilities, while adding that

“immigrant speech may disturb or irritate the listener regardless of whether the immigrant’s Finnish is comprehensible”. Thus, the research presented here has far-reaching implications that go beyond linguistics. Indeed, such research could potentially benefit recruiters and persons tasked with employee assessment, covering a wide variety of disciplines. Considering its implications, the importance of research on accents and other phenomena related to L2 oral production cannot be understated.

References

Aboitiz, F. 2018. Voice, gesture and working memory in the emergence of speech. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 19: 70–85.

Aboitiz, F. 1995. Working memory networks and the origin of language areas in the human brain. *Medical hypotheses*, 44: 504–506.

Aboitiz, F., García, R., Brunetti, E. & Bosman, C. 2006. The origin of Broca's area from an ancestral working memory network. In Y. Grodzinsky & K. Amunts (eds), *Broca's Region*, 3–16, Oxford: Oxford University Press.

Aboitiz, F., Aboitiz, S. & García, R. 2010. The Phonological Loop: A Key Innovation in Human Evolution. *Current Anthropology*, 51(S1): S55–S65.

Abrahamsson, N. & Hyltenstam, K. 2008. The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30: 481–509.

Acheson, D. J. & MacDonald, M. C. 2009. Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, 60: 329–350.

Acheson, D. J., Hamidi, M., Binder, J. R. & Postle, B. R. 2011. A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23: 1358–1367.

Ahmadian, M. J. 2012a. The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*, 16: 129–149.

Ahmadian, M. J. 2012b. The relationship between working memory and oral production under task-based careful online planning condition. *TESOL Quarterly*, 46: 165–175.

Baddeley, A. D. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4: 417–423.

Baddeley, A. D. 2007. *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A. D. & Hitch, G. J. 1974. Working memory. In G. Bower (ed.), *The psychology of learning and motivation*, 47–90, New York: Academic Press.

Ball, M. J. & Rahilly, J. 1999. *Phonetics: The Science of Speech*. New York: Oxford University Press.

Bialystok, E. & Hakuta, K. 1999. Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (ed.), *Second language acquisition and the Critical Period Hypothesis*, 161–181, Lawrence Erlbaum Associates Publishers.

Birdsong, D. 1999. Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, 1–22, Mahwah, NJ: Erlbaum.

Birdsong, D. 2005. Interpreting age effects in second language acquisition. In J. Kroll & A. De Groot (eds), *Handbook of bilingualism: Psycholinguistic perspectives*, 109–127, New York: Cambridge University Press.

Boersma, P. 2012. Modeling phonological category learning. In A.C. Cohn, C. Fougeron & M. K. Huffman (eds), *The Oxford handbook of laboratory phonology*, 207–218, Oxford University Press.

Bongaerts, T. 1999. Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners. In D. Birdsong (ed.), *Second language acquisition and the Critical Period Hypothesis*, 133–159, Lawrence Erlbaum Associates Publishers.

Broselow, E. 2018. Syllable structure in the Arabic dialects. In E. Benmamoun & R. Bassiouney (eds), *The Routledge Handbook of Arabic Linguistics*, 32–47, London and New York: Routledge.

Broselow, E. 1983. Non-obvious transfer: on predicting epenthesis errors. In S. Gass & L. Selinker (eds), *Language transfer in language learning*, 269–280, Rowley, MA: Newbury House.

Broselow, E. 1988. Prosodic phonology and the acquisition of a second language. In S. Flynn & W. O'Neil (eds), *Linguistic theory and second language acquisition*, 295–308, Dordrecht: Kluwer.

Broselow, E., Chen, S-I. & Wang, C. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, 20: 261–280.

Browman, C. P. & Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (eds), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376, New York: Cambridge University Press.

Browman, C. P. & Goldstein, L. 1995. Gestural syllable position effects in American English. In F. Bell-Berti & L. J. Raphael (eds), *Producing Speech: Contemporary Issues*, 19–33, New York: AIP Press.

Buchwald, J., Guthrie, D., Schwafel, J., Erwin, R. & Van Lancker, D. 1994. Influence of language structure on brain-behavior development. *Brain and Language*, 46: 607–619.

Carlisle, R. 1997. The modification of onsets in a markedness relationship: testing the interlanguage structural conformity hypothesis. *Language Learning*, 47: 327–361.

Carroll, J. & Sapon, S. 1959. *Modern language aptitude test*. New York, NY: The Psychological Corporation/Harcourt, Brace Jovanovich.

Carroll, J. 1981. Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.), *Individual differences and universals in language learning aptitude*, 83–118, Rowley, MA: Newbury House.

Chambers, J. K. 2008. *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Maiden, MA. Blackwell Publishers.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts; London, England: The MIT Press.

Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Chomsky, N. 1975. *Reflections on language*. New York: Pantheon Books.

Chomsky, N. 1980. *Rules and representations*. Oxford: Blackwell.

Chomsky, N. 1986. *Barriers*. Cambridge, Massachusetts: The MIT Press.

Christophersen, P. 1973. *Second-language learning: Myth and reality*. Baltimore: Penguin.

Colantoni, L., Steele, J. & Escudero, P. 2015. *Second Language Speech: Theory and Practice*. Cambridge University Press.

Cole, R. A. & Cooper, W. E. 1976. Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 58: 1280–1287.

Cook, V. 2016. *Second Language Learning and Language Teaching*. Fifth Edition. New York: Routledge.

Corder, S. P. 1967. The significance of learner's errors. *International Review of Applied Linguistics*, 5: 161–170.

Curtiss, S. 1977. *Genie: A psycholinguistic study of a modern day "wild child"*. New York: Academic Press.

DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22: 499–533.

Derr, M. A. & Massaro, D. W. 1980. The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/–/jus/ distinction. *Perception & Psychophysics*, 27: 51–59.

Dulay, H.C. & Burt, M.K. 1972. Goofing: an indication of children's second language learning strategies. *Language Learning*, 22: 235–252.

Dulay, H.C. & Burt, M.K. 1974. Errors and strategies in child second language acquisition, *TESOL quarterly*, 8: 129–136.

Eimas, P. D. 1975. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-1] distinction by young infants. *Perception and Psychophysics*, 18: 341–347.

Eubank, L. 1993/94. On the transfer of parametric values in L2 development. *Language Acquisition*, 3: 183–208.

Fischer, J. L. 1958. Social influences on the choice of linguistic variant. *Word*, 14: 47–56.

Flege, J. 1988. The Production and Perception of Foreign Language Speech Sounds. In H. Winitz (ed.), *Human Communication and Its Disorders, A Review 1988*, 224–401, Norwood, N.J.: Ablex.

Flege, J. 1992a. The intelligibility of English vowels spoken by British and Dutch talkers. In R. Kent (ed.), *Intelligibility in speech disorders*, 157–232, Amsterdam: John Benjamins.

Flege, J. 1992b. Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (eds), *Phonological development: Models, research, implications*, 565–604, Timonium, MD: York Press.

Flege, J. 1995. Second language speech learning. Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233–277, Timonium, MD: York Press.

Flege, J. 1999. Age of Learning and Second Language Speech. In D. Birdsong (ed.), *Second language acquisition and the Critical Period Hypothesis*, 101–133, Lawrence Erlbaum Associates Publishers.

Ganschow, L. & Sparks, R. L. 1996. Anxiety about Foreign language learning among high school women. *Modern Language Journal*, 80: 199–212.

Garrett, P., Williams, A. & Coupland, N. 2003. *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. Cardiff: University of Wales Press.

George, H.V. 1972. *Common errors in language learning*. Rowley, MA: MIT Press.

Geschwind, N. & Galaburda, A.M. 1985a. Cerebral lateralization. Biological mechanisms, associations and pathology: I. *Archives of Neurology*, 42: 428–459.

Geschwind, N. & Galaburda, A.M. 1985b. Cerebral lateralization. Biological mechanisms, associations and pathology: II. A hypothesis and a program for research. *Archives of Neurology*, 42: 521–552.

Geschwind, N. & Galaburda, A.M. 1985c. Cerebral lateralization. Biological mechanisms, associations and pathology: III. *Archives of Neurology*, 42: 634–654.

Giles, S. & Moll, K. 1975. Cinefluorographic Study of Selected Allophones of English /I/. *Phonetica*, 31: 206–227.

Gregg, K. R. 1984. Krashen's Monitor and Occam's Razor. *Applied Linguistics*, 5: 79–100.

Hawkins, R. 2019. *How Second Languages are Learned: An Introduction*. Cambridge: Cambridge University Press.

Haznedar, B. & Schwartz, B. 1997. Are there optional infinitives in child L2 acquisition? In Hughes, E. Hughes, M. and Greenhill, A. (eds), *Proceedings of the 21st Annual Boston University Conference on Language Development (BUCLD)*, 293–306, Somerville, MA: Cascadilla Press.

Haznedar, B. 2001. The acquisition of the IP system in child L2 acquisition. *Studies in Second Language Acquisition*, 23: 1–39.

Haznedar, B. 2003. Missing Surface Inflection in Adult and Child L2 Acquisition. In J. M. Liceras, H. Zobl & H. Goodluck (eds), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002): L2 Links*, Cascadilla Proceedings Project Somerville, MA 2003.

Jakobson, R. 1968. *Child language, aphasia, and phonological universals*. The Hague: Mouton.

Jansen, W. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. Groningen Dissertations in Linguistics 74. Groningen: University of Groningen.

Johnson, J. & Newport, E. 1989. Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21: 60–99.

Juffs, A. & Harrington, M. W. 2011. Aspects of working memory in L2 learning. Language teaching: Reviews and studies. *Language Teaching*, 42: 137–166.

Kounin, T. & Krashen, S. 1978. Approaching Native Speaker Competence from Two Different Directions. In C. Blatchford & J. Schachter (eds), *On TESOL '78: EFL Policies, Programs, Practices*, 205–212, Washington: TESOL.

Krashen, S. 1978. Individual Variation in the Use of the Monitor. In W. Ritchie (ed.), *Principles of Second Language Learning*, 175–183, New York: Academic Press.

Krashen, S. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

Krashen, S. 1982. *Principles and Practices of Second Language Acquisition*. Oxford: Pergamon Press.

Krashen, S. 1985. *The Input Hypothesis: Issues and Implications*. London: Longman.

Krashen, S. & Terrell, T. 1983. *The Natural Approach: Language Acquisition in the Classroom*. Hayward, CA: Alemany Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. 1992. Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, 255: 606–608.

Ladefoged, P. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA & Oxford: Blackwell.

Lahiri, A. & Reetz, H. 2002. Underspecified recognition. In C. Gussenhoven & N. Warner (eds), *Laboratory Phonology VII*, 637–677, Mouton de Gruyter: Berlin.

Lahiri, A. & Reetz, H. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38: 44–59.

Lance, D. 1969. *A Brief Study of Spanish-English Bilingualism: Final Report. Research Project ORR-Liberal Arts – 15504*. College Station, TA: Texas A & M.

Lardiere, D. 1998a. Case and Tense in the 'fossilized' steady state. *Second Language Research*, 14: 1–26.

Lardiere, D. 1998b. Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research*, 14: 359–375.

Lass, R. 1984. *Phonology: an introduction to basic concepts*. Cambridge University Press.

Lenneberg, E. H. 1967. *Biological Foundations of Language*. New York: Wiley.

Li, S. 2017. The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. In L. Gurzynski-Weiss (ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors*, 41–70, Amsterdam: John Benjamins Publishing Company.

Li, S. 2019. Six decades of language aptitude research: a comprehensive and critical review. In Z. Wen, P. Skehan, A. Biedron, S. Li & R. L. Sparks (eds), *Language aptitude: advancing theory, testing, research and practice*, 78–96, Routledge.

Long, M. H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition*, 12: 251–285.

McCarthy, C. 2008. Morphological variability in the comprehension of agreement: an argument for representation over computation. *Second Language Research*, 24: 227–263.

McLaughlin, B. 1987. *Theories of second language learning*. London: Edward Arnold.

Meisel, J. 1997. The acquisition of the syntax of negation in French and German: contrasting first and second language development. *Second Language Research*, 13: 227–263.

Menn, L. 1981. Theories of phonological development. In H. Winitz (ed.), *Native language and foreign language acquisition*. New York: New York Academy of Sciences.

Miyake, A. & Shah, P. 1999. *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Naiman, N., Fröhlich, M., Stern, H. & Todesco, A. 1995. *The Good Language Learner*. Clevedon: Multilingual Matters.

Narayanan, S. S., Alwan, A. A. & Haker, K. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *The Journal of the Acoustical Society of America*, 101: 1064–1077.

Nemser, W. 1971. Approximative systems of foreign language learners. *International Review of Applied Linguistics*, 9: 115–123.

Newport, E. 1988. Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences*, 10: 147–172.

Newport, E. 1990. Maturational Constraints on Language Learning. *Cognitive Science*, 14: 11–28.

Nolan, F. 1992. The descriptive role of segments: Evidence from assimilation. In G. J. Docherty & D. R. Ladd (eds), *Papers in laboratory phonology II: Gesture, segment, prosody*, 261–280, New York: Cambridge University Press.

Novoa, L., Fein, D. & Obler, L. K. 1988. Talent in foreign languages: A case study. In L. K. Obler & D. Fein (eds), *The Exceptional Brain: Neuropsychology of Talent and Special Abilities*, 294–302, New York, NY: Guilford.

Obler, L. 1978. *A unitary phonological production system in balanced bilinguals*. Paper presented at the Linguistic Society of America Meeting, Champaign-Urbana, III.

Obler, L. K. 1982. The Parsimonious Bilingual. In L. K. Obler & L. Menn (eds), *Exceptional Language and Linguistics. Perspectives in Neurolinguistics, Neuropsychology, and Psycholinguistics*, 339–347, New York: Academic Press.

Obler, L. K. & Gjerlow, K. 1999. *Language and the Brain*. Cambridge University Press.

Odden, D. 2005. *Introducing Phonology*. Cambridge: Cambridge University Press.

O'Malley, J. M. & Chamot, A. U. 1990. *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

Prévost, P. & White, L. 2000. Missing surface inflection or impairment in second language? Evidence from Tense and Agreement. *Second Language Research*, 16: 103–33.

Schneiderman, E. & Desmarais, C. 1988. The talented language learner: some preliminary findings. *Second Language Research*, 4: 91–109.

Schumann, J. H. 1975. Affective factors and the problem of age in second language acquisition. *Language Learning*, 25: 209–235.

Scovel, T. 1988. *A time to speak. A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*. Oxford University Press.

Seliger, H. W., Krashen, S. & Ladefoged, P. 1975. Maturational constraints in the acquisition of a native-like accent in second language learning. *Language Sciences*, 36: 20–22.

Selinker, L. 1969. Language Transfer. *General Linguistics*, 9: 69–92.

Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10: 209–231.

Selinker, L., Swain, M. & Dumas, G. 1975: the interlanguage hypothesis extended to children. *Language Learning*, 25: 139–191.

Sharwood Smith, M. & Truscott, J. 2014. *The multilingual mind: a modular processing perspective*. Cambridge: Cambridge University Press.

Skehan, P. 1986. Cluster analysis and the identification of learner types. In V. Cook (ed.), *Experimental approaches to second language learning*. Oxford: Pergamon.

Slomanson, P. & Newman, M. 2004. Peer group identification and variation in New York Latino English laterals. *English World-Wide*, 25: 199–216.

Snow, C. & Hoefnagel-Höhle, M. 1977. Age differences in the pronunciation of foreign sounds. *Language and Speech*, 20: 357–365.

Sommerstein, A. 1977. *Modern Phonology*. Edward Arnold Publishers.

Spencer, A. 1996. *Phonology*. Oxford & Cambridge, Massachusetts: Blackwell Publishers.

Spencer, A. 1991. *Morphological Theory*. Oxford & Cambridge, Massachusetts: Basil Blackwell.

Sproat, R. & Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21: 291–311.

Stafford, C. & Covitt, G. 1978. Monitor use in adult second language production. *ITL: Review of Applied Linguistics*, 39–40: 103–125.

Strange, W. & Broen, P. 1980. Perception and production of approximant consonants by 3-year-olds: A first study. In G. Yeni-Komshian, J. Kavanagh & C. Ferguson (eds), *Child Phonology*, *Volume 2. Perception*, 117–154, New York: Academic Press.

Streeter, L. A. 1976. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 259: 39–41.

Tarone, E. 1979. Interlanguage as chameleon. *Language Learning*, 29: 181–191.

Tarone, E. 1983. On the variability of interlanguage systems. *Applied Linguistics*, 4: 142–163.

Tarone, E. 2018. Interlanguage. *The Encyclopedia of Applied Linguistics*, 1–7.

Toivola, M., Lennes, M. & Aho, E. 2008. Onko maahanmuuttajien vieraalla aksentilla väliä? In M. O'Dell & T. Nieminen (eds), *Fonetiikan päivät 2008*, 73–78, Tampere Studies in Language, Translation and Culture, Series B 3. Tampere: Tampere University Press.

Trehub, S. E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47: 466–472.

Wells, J. C. 1982. *Accents of English*. Cambridge University Press.

Wen, Z. 2016. *Working memory and second language learning*. Bristol: Multilingual Matters.

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

Westergaard, M. 2021. Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research*, 37: 379–407.

White, L. 1991. The verb-movement parameter in second language acquisition. *Language Acquisition*, 1: 337–360.

White, L. 1992. Long and short verb movement in second language acquisition. *Canadian Journal of Linguistics*, 37: 273–286.

White, L. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Wiberg, A., Ng, M., Al Omran, Y., Alfaro-Almagro, F., McCarthy, P., Marchini, J., Bennett, D. L., Smith, S., Douaud, G. & Furniss, D. 2019. Handedness, language areas and neuropsychiatric diseases: insights from brain imaging and genetics. *Brain: a journal of neurology*, 142: 2938–2947.

Wrench, J. S. 2008. *Quantitative Research Methods for Communication: A Hands-on Approach*. New York: Oxford University Press.

Zsiga, E. C. 2021. *The Phonology/Phonetics Interface*. Edinburgh Advanced Textbooks in Linguistics.