The Development of Speech Rhythm and Fluency of Advanced English Learners: A Mixed Methods Study of the Correlation Between Native Speaker Evaluations and Acoustic Measures

Emma Salomaa

MA Thesis

English, Degree Programme for Language Learning and Teaching
School of Languages and Translation Studies
Faculty of Humanities
University of Turku

April 2019

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU

School of Languages and Translation Studies / Faculty of Humanities

SALOMAA, EMMA: The Development of Speech Rhythm and Fluency of Advanced English Learners: A Mixed Methods Study of the Correlation between Native Speaker Evaluations and Acoustic Measures

MA Thesis, 65 pages, 12 pages of appendices
English, Degree Programme for Language Learning and Teaching
April 2019

This thesis examined the changes in speech rhythm and fluency of advanced English learners during a pronunciation course. The research strived to answer the following research questions: 'Do speech rhythm and fluency of advanced English learners change after a pronunciation course according to native speaker ratings?', 'if yes, which acoustic measures do these changes correlate with?', and 'what is the correlation between the perceived speech rhythm, fluency, accentedness, and comprehensibility?'. I approached these issues through mixed methods, both quantitative and qualitative. First, 20 advanced Finnish learners of English were selected out of 45 first-year major English students available. The number of the participants was limited keeping in mind the duration of the native-speaker questionnaire. The questionnaire consisted of background information questions and 42 speech samples to be rated on a 9-point Likert scale, cropped from the learner recordings as well as one native speaker. After collecting responses from 31 native speakers of English, I conducted a statistical analysis, whose results were then used for extreme case sampling. The speech of four learners with the biggest changes in their speech rhythm and fluency was then analyzed acoustically to find the contributing factors.

The results showed both positive and negative changes on an individual level, but the differences were not statistically significant on a group level. The acoustic analysis demonstrated higher fluency scores correlating with faster articulation rate, smaller number of unfilled pauses, the location of pauses at phrase or clause boundaries, and fewer repairs. Rhythm measures revealed that pitch and amplitude peaks generally matched better and the use of durational cues as well as vowel reduction and linking increased in the posttest speech. All four rated aspects correlated significantly, particularly speech rhythm and fluency scores in both pre- and posttest samples ($r = 0.98$). Thus, these two can be said to be closely intertwined. Based on the results, speech rhythm should not be neglected in pronunciation instruction as it strongly influences the perceptions of fluency, accentedness, and comprehensibility. It is suggested that further research on rhythm focuses on its nature as both a perceived and produced phenomenon, as well as defining its relationship to fluency.

Keywords: speech rhythm, fluency, prosody, suprasegmentals, pronunciation, L2 English, English as a second language, second language acquisition

Table of contents

List of figures

List of tables

Abbreviations

AmE	American English variety
BrE	British English variety
ESL	English as a second language
F0	Fundamental frequency
L1	First (native) language
L2	Second language (or any foreign language learned after L1)
SLA	Second language acquisition

1 Introduction

Speech rhythm has been a topic of interest since the 18th century, and a subject of ambitious phonetic research since the early 1900s. Nonetheless, only a few decades ago the paradigm shifted from measuring timing parameters to perceived periodicity. More recently, rhythm has been connected to the key measures of fluency, namely pause duration and placement, as well as speech rate. Rhythm and fluency are clearly two separate aspects of spoken language: rhythm refers to the periodicity of beats in speech while fluency describes the smoothness and effortlessness of delivery. Their relationship, however, has not been extensively investigated. In addition, rhythm has been studied acoustically and through listener ratings, but there is barely any research which combines this with auditory analysis and listener evaluation of fluency.

In language teaching, segmentals (individual phonemes) are often emphasized over suprasegmentals (stress, rhythm, and intonation) or even ignored if time is limited. However, research on L2 prosody shows that it often contributes to accentedness and comprehensibility as much as segmental features, or possibly even more (e.g., Derwing, Munro, and Wiebe 1998). Therefore, it would be justified to pay attention to prosody in L2 classrooms. Although most contemporary scholars agree on the L2 pronunciation goal of intelligibility, it can be argued whether advanced language learners and preservice teachers should aim for nativeness. Another issue is whether it makes sense to learn perfectly one native model of English which may not be comprehensible to all L1 and L2 English speakers, the terms *L1* and *L2* referring to the first (native) language and the second language of a person, respectively. It is thus important examining, how the speech rhythm, fluency, accentedness, and comprehensibility correlate with one another.

Based on the previous research on speech rhythm, fluency, and L2 prosody, a research gap can be found: The aim of the present study is to find out whether there is a change in the prosodic features of the speech of advanced major students of English after a pronunciation course, and which acoustic measures contribute to this possible change. First of all, the prosodic development during the pronunciation course is important to examine so that the influence of the instruction can be evaluated. Secondly, determining which features native listeners attend to can help learners to focus on the features that actually make a difference. Thirdly, investigating the correlations between the perceptions of rhythm, fluency, accentedness, and comprehensibility will further our understanding of how connected these aspects are in the mind of native English speakers.

My main focus is on speech rhythm and fluency, but because of their potential influence on the rating scores, I will also consider the roles of accentedness and comprehensibility in the productions. I am also interested in the correction of the aforementioned aspects. The following three were specified as the research questions of this study:

1. Do speech rhythm and fluency of advanced Finnish learners of English change during a pronunciation course according to native speaker ratings?
2. If the native-speaker raters hear changes in the speech rhythm and fluency, which acoustic features do these changes correlate with?
3. Do the four rated aspects, rhythm, fluency, accentedness, and comprehensibility correlate with each other?

I will approach the subject through mixed methods: First, I will have native speakers of English rate recorded pre- and posttest utterances from Finnish students of English. According to these ratings, numerical values for each student will be calculated. Second, I shall select a few students with the most notable differences in their pre- and posttest values and analyze their utterances auditorily and acoustically to find out which features do the native speakers of English seem to react to. It has been noted that acoustic measures can be used to validate listener judgements of L2 speech but should not be used instead of them (Paananen-Porkka 2007; Derwing and Munro 2015). As previous research on the subject highlights, it does not matter if there are measurable changes in the acoustic signal if the listeners do not attend to them (ibid.).

The outline of my thesis is as follows: First, in the theoretical framework section, I will discuss the nature and previous research of speech rhythm, particularly in terms of measuring it. There will also be a short review on fluency as a dimension of spoken language as well as on the relationship of fluency and rhythm. Since the present study examines possible changes in the production of learners after pronunciation instruction, the acquisition of L2 prosody is considered in the theoretical framework as well. Second, we shall look into the methods used in the present study, followed by the report of the results with relevant tables and figures. Then the results are interpreted and their implications in education are considered in the discussion. The present study will also be evaluated and suggestions for further research will be given. Last, I shall conclude the thesis with a summary of the results as well as the main points from the discussion.

2 The study of speech rhythm, fluency, and L2 prosody

In this section, I will present the theoretical framework of the present study. It begins with an overview of speech rhythm and its components, *word* and *sentence stress*, and the issues in measuring rhythm and stress. Then, we take a look at fluency and its relationship with rhythm. Finally, we will move on to the issues related to the learning of L2 prosody: the acquisition processes, L1 transfer, the research on L2 accentedness and comprehensibility, and finally, goals in pronunciation instruction.

2.1. Rhythm as a prosodic element

Wennestrom (2001, 50) describes rhythm as "a universal human trait". This observation has led many phonologists to believe that rhythm is fundamental to stress patterns in human speech. Stress is built on an underlying hierarchical structure provided by speech rhythm. Rhythm can extend across intonation units and even from one speaker to another (ibid.). Speech rhythm is a *prosodic* or *suprasegmental* element, meaning that it is not limited to separate segments in speech but is present throughout the entire utterance (e.g., Cruttenden 2014, 4). In addition to rhythm, Derwing and Munro (2015, 3) list *word* and *sentence stress*, *intonation*, *juncture*, and *tone* into prosody. Lehtonen (1977) notes that the tight connection between rhythm, stress, intonation, and pitch makes it difficult to distinguish these features from one another.

Paananen-Porkka (2007) proposes speech rhythm consisting of the linguistic components illustrated in Table 1. In the present study, sentence stress related features and juncture phenomena are addressed through auditory observation and acoustic measures of rhythm, whereas speech rate and pausing phenomena are examined through the conventional fluency measures.

Table 1 Linguistic components of speech rhythm according to Paananen-Porkka (2007)

Sentence stress	Juncture	Speech rate	Pausing
1) Intensity 2) Pitch 3) Duration 4) Sound quality	1) assimilation 2) dissimilation 3) elision 4) linking	Articulation rate	1) Number of pauses 2) Duration of pauses 3) Location of pauses

One traditional rhythmic phenomenon is *isochrony*, the equal duration of feet and syllables (Arvaniti 2009, 46). English scholars in the 18th century made an observation of English stressed syllables following each other at isochronous intervals, and this was the basis for rhythm research for centuries (Kohler 2009, 30). It was then concluded that all languages can be categorized into either *stress-timed* or *syllable-timed* languages, in which timing refers to duration as a cue of prominence in the acoustic signal. English was examined as the typical example of a stress-timed language. Contrary to the earlier claims, Finnish it not a syllable-timed language (Nieminen and O'Dell 2009). The closest rhythmic pattern would be *mora-timing* usually connected to Japanese but as Nieminen and O'Dell (2009) point out, variation in syllable length can change the timing to resemble any of the three traditional rhythm types.

Grabe and Low (2002, 538) found that the distinction of stress- and syllable-timing is supported by a weak categorization. However, they admit that the distinction cannot be used to categorize all languages, and that the languages in stress-timed, syllable-timed, and unclassified groups overlap considerably. Some scholars prefer using less restrictive terms *stress-based* and *syllable-based* rhythm, such as Deterding (2012, 9-10) who argues that these categories are not only about timing, but also about alternation of stressed and unstressed syllables. Couper-Kuhlen (1993, 13) has discussed the idea of rhythm being a continuum on which languages would be placed predominantly by phonological criteria. On the other hand, Arvaniti (2009, 58) argues that the continuum drawn by rhythmic metrics is not from syllable- to stress-timing but instead from least to most stress-based. It has been proposed that the difference between stress-timing and syllable-timing is actually about "differences in syllable structure, vowel reduction, and the phonetic realization of stress and its influence on the linguistic system" rather than the isochrony of the interstress intervals (Dauer 1983, 51; Ratchke & Smith 2015). Barry (2008, 110), for his part, sees vowel reduction as valid in differentiating languages, but even languages that have this tendency do not all behave the same way. More recently, isochrony has been either abandoned altogether or re-interpreted as a perceptual construct created by the listener compensating for "predictable acoustic regularities" (Ratchke & Smith 2015, 3).

In order to discover the nature and function of rhythm in communication, four perspectives are required in speech rhythm research, according to Kohler (2009, 29): "symbolic representation, production, perception, and communicative function". However, the dominant focus on produced instead of perceived rhythm and particularly on timing has received criticism (Cumming 2010 n.

pag.). Several studies (e.g., Cumming 2010; Dilley, Wallace, and Heffner 2012) provide evidence that perceptual isochrony is not acoustically isochronous, which calls into question the use of metrics or interstress intervals in rhythm analysis. Perceived rhythm is created by multiple integrated acoustic cues, and languages differ in the relative contribution of each cue (Cumming 2010, 209). Therefore, Cumming (2010) calls for more careful consideration of the influence of native language on rhythm perception. Despite their doubts about the existence of rhythm, Molczanow and Wiese (2014, 173) admit that perceived rhythm has an essential role in processing information in speech. Espinosa (2016, 16) concludes that linguists and phoneticians will continue to investigate the nature of speech rhythm; whether it is a universally measurable property of human language or emerges as a result of the different phonetic and phonological structures in languages.

2.1.1 English word and sentence stress

As discussed above, sentence stress is a core component of speech rhythm. It is sometimes referred to as *accent* or *nuclear stress.* The present study uses the term *sentence stress* due to its transparent meaning and because accent is further discussed in the sense of a specific way of pronunciation. Word or lexical stress and sentence stress both express the prominence of one syllable compared to other syllables with the distinction that word stress concerns single words and sentence stress longer utterances (Cutler 2011, 77). However, Barry (2008, 102) argues that word stress is not separable from sentence stress and further, sentence stress cannot be separated from intonation.

In English, multisyllabic words typically have both primary and secondary stress (Cruttenden 2014, 248). *Function* or *grammatical words*, namely auxiliary verbs, prepositions, pronouns, conjunctions, and articles, are usually unstressed and in this position their vowels reduced, whereas *lexical* or *content words* (e.g. main verbs, nouns, adverbs) receive prominence (Cruttenden 2014, 270). Two utterances with a different number of syllables but equal number of stresses are roughly the same length since the unstressed syllables are reduced and compressed (Celce-Murcia, Brinton, and Goodwin 1996, 152). In the present study, the distinction of grammatical and lexical words was used in the process of selecting the utterances for the questionnaire: the more grammatical words and the fewer lexical ones, the more potential vowel reduction and elision.

Wennestrom (2001, 50) describes English rhythm as *trochaic*, which means that the foot begins with a stressed syllable and is followed by an unstressed one. Word and sentence stress together create "the regular, patterned beat of stressed and unstressed syllables and pauses" (Celce-Murcia, Brinton, and Goodwin 1996, 152). However, the same way as speech rhythm is not only about sentence stress, sentence stress is not simply a sum of the lexical stress patterns in the utterance but is influenced by semantic and pragmatic factors as well (Cutler 2011, 82). Cutler (2011, 86-7) notes that sentence stress is used in spoken language to signal information structure, focus, or contrast.

It has been shown that stress patterns in the language production process are not applied by any general rules but are part of the pronunciation information of words in the mental lexicon (Cutler 2011, 78). Occasionally, the urge for more global stress patterns may override lexical prominences (Couper-Kuhlen 1993, 50). For instance, in order to sound prosodically correct, sometimes stresses need to be added to even usually unstressed function words (Couper-Kuhlen 1993, 39). These kinds of rhythmic adjustments to avoid *stress clashes* are quite common in natural speech, when the primary stresses of two adjacent words would otherwise appear immediately after each other (Wennestrom 2001, 56-9). This all matters because in the language comprehension process word segmentation and recognition rely to a large extent on perceptual isochrony and the "identification of lexical stress pattern" (Cutler 2011, 79; Dilley, Wallace, and Heffner 2012, 252). Misplaced stress leads to the loss of important cues assisting speech comprehension and may even cause communication breakdown (Celce-Murcia, Brinton, and Goodwin 1996, 185).

Couper-Kuhlen (1993, 48) found spontaneous English speech not uniformly isochronous but not completely anisochronous either and calls for investigation on whether English isochrony can be generalized to new speech events or even other languages (Couper-Kuhlen 1993, 297). She doubts that "speech rhythm is as fundamental and diversified a cue elsewhere as in English" (1993, 298). Indeed, it seems that there are different levels of importance of prominence in languages. For instance, lexical stress in Finnish is highly predictable, and does not distinguish meanings nor lead to misinterpretations of words per se. In contrast, English lexical stress is an essential part of the knowledge of each word. Hence, Kohler (2009, 31) rejects the idea of defining rhythm based on lexical phonology or sentence stress which is used to emphasize meaning in speech. Kohler argues that rhythm cannot be considered as "a fixed typological prominence pattern for groups of languages" since it varies within each language (2009, 44). Thus, he proposes a new research

paradigm which relies on the listener's perception of rhythmicity instead of measurable isochrony (ibid.). It is also possible to combine the perspectives of perception and acoustic measures, as discussed in the next subsection.

2.1.2 Measuring rhythm and stress

As mentioned above, the earlier rhythm studies were motivated by the urge to find substance for the perceptions of isochrony (Barry 2008, 105-6). The most popular attempts to quantify rhythm have been different kinds of metrics, mathematic formulas, and equations. For instance, in Ramus, Nespor, and Mehler (1999) rhythm in different languages is measured by the proportion of vocalic intervals and the standard deviation of the duration of both vocalic and consonantal intervals, whereas the *Pairwise Variability Index* (e.g., Grabe and Low 2002) uses the vocalic and intervocalic durations in a summation equation. Kohler (2009) claims that these metrics are justified in data sorting but not in constructing an explanatory model of speech rhythm. The problem is that the metrics measure only timing ignoring the other properties of speech rhythm (Arvaniti 2009; Kohler 2009; Arvaniti 2012; Deterding 2012). Based on the results from several studies, other prosodic components such as pitch and speech rate should be included in the analysis (Arvaniti 2012, 89). Arvaniti (2009, 46-53) adds that the metrics are unable to classify all languages and also unreliable due to possible manipulation of results with certain types of utterances. Despite the popularity of the rhythmic measures, Kohler (2009, 30) does not believe that acoustic or articulatory parameters could truly define stress. Arvaniti (2009, 61) proposes a new conception of rhythm which would require the adaption of acoustic measurements to the prosody of each language as well as considering the native speaker perception.

In order to avoid the criticized timing metrics, measuring speech rhythm can also be done by focusing on the acoustic features that determine prominence in speech: fundamental frequency (F0), duration, intensity, and vowel quality. These features are produced and perceived on syllable level (Kohler 2008, 261-2). Although rarely involved in prominence analysis, F0 seems to group syllables more effectively than the other features. Perceived fundamental frequency is called pitch and measured in hertz (Hz). The duration of syllables contributes to global speech rate (Cumming 2010, 192) and is usually measured in milliseconds (ms). The intensity of the sound wave, measured in decibels (dB), increases as the amplitude does. Lehtonen, Sajavaara, and May (1977, 81) claim that intensity is the weakest perceptual correlate of word stress in both English and

Finnish. Wennestrom (2001, 53) interprets the alignment of the amplitude peaks and the pitch peaks representing a regular rhythm. She points out that while both stress and intonation cues include pitch, volume, and duration, they can be examined separately (2001, 47). Also tone and duration have been found to play a significant role in rhythm perception, but to different extents to speakers of different native languages (Cumming 2010, 192). Vowel quality can be observed qualitatively from the spectrogram or measured precisely from the formant values and then compared to the values in native-speaker speech or previous production of the same speaker in a pretest-posttest condition (e.g., Peltola, Lintunen, and Tamminen 2014). In the present experiment, a qualitative analysis was conducted.

The other prosodic features composing rhythm, such as juncture, i.e., assimilation, dissimilation, linking, and elision, can be easy to observe auditorily, but difficult to detect in the acoustic signal. Juncture marks word or morpheme boundaries (Cruttenden 2014, 318) and linking can occur between a word-final consonant, linking 'r', or a glide (/j/, /w/), and a word-initial vowel (Morris-Wilson 1981, 174-5). In the data of the present experiment, juncture, if present, is mostly manifested as linking and elision. It was not in the scope of the study to measure speech rhythm by the common metrics, based on the criticism they have received. Moreover, I was not interested in categorizing the L2 speech rhythm nor examining timing alone. Thus, I combined the perceived rhythm with the acoustic measures in my analysis. It should be borne in mind, however, that the perceived regularity or periodicity may not correspond directly to any measurement of the acoustic signal (Couper-Kuhlen 1993, 14). The next subsection will discuss the other key concept of the present study, fluency.

2.2 Fluency and its relationship to rhythm

The term *fluency* can be used in two different senses: *broad* or *narrow*. Broad definition refers to general language proficiency, whereas narrow definition is objectively measurable. In the present study, fluency is considered only in the narrow sense, which Tavakoli, Campbell, and McCormack (2016, 448) define as "the ability to communicate one's intended meaning effortlessly, smoothly, and with no or little disruption". Ullakonoja (2011, 25), looking at fluency from the phonetic perspective, notes that it is all about prosody. To address the data of the present study, several definitions of fluency can be combined into the following: *read-aloud speech* is fluent when it is spoken at *a regular and appropriate rate*, with *no or a few hesitations* and *a small number of*

relatively short pauses at phrase and clause boundaries. The rhythm is perceived as *relatively regular and natural* (e.g., Segalowitz 2007, 181; Ullakonoja 2011, 38).

Segalowitz (2010, 47-9) distinguishes three aspects of fluency: *Cognitive, perceived*, and *utterance fluency*. Cognitive fluency refers to the way the speaker translates thoughts into speech. As this is difficult to investigate purely from the production, it was not in the scope of the present study. Perceived fluency measures how listeners perceive the production. In the present study, this part is addressed via the native-speaker questionnaire. Utterance fluency is measured objectively from an utterance by acoustic measures. The subsequent acoustic analysis focuses on this aspect. Tavakoli and Skehan (2005) divide utterance fluency further into three sub-dimensions: *breakdown, speed*, and *repair*. Speech fluency is about the rate with which the speech is delivered, breakdown fluency about pausing, and repair fluency refers to the self-correcting phenomena of the speech.

The first aspect of utterance fluency, speed, is manifested in speech rate or articulation rate. Speech rate is measured by the number of units per total utterance time and articulation rate is the speech rate excluding pauses. Thus, pauses affect the speech rate but not articulation rate. Measuring speech rate is not, by any means, straightforward; if measured by syllables per minute, Finnish is spoken faster while English is faster if measured by words per minute (Lehtonen 1979). In general, more advanced learners speak faster than beginners. Interestingly, Segalowitz and Freed (2004) found that L2 speakers' speech rate may even decrease when becoming more fluent, possibly because they develop the cognitive processing skills to monitor their own speech. Since the present study analyzes read-aloud speech instead of spontaneous speech, it is worth considering their differences. Howell and Kadi-Hanifi (1991, 169) found systematic differences in stress positioning between spontaneous and read speech of the same speakers, and majority of the pauses were dropped in read speech. Ullakonoja (2011, 35) notes that the more complex in orthography, vocabulary or sentence structure the text is, the slower it is to read aloud.

The second, breakdown fluency, covers the use of pauses which can be unfilled (i.e., silent) or filled. De Jong (2016, n. pag.) suggests setting a threshold of the pauses around 250–300 ms for the most reliable outcome. She (ibid.) defines silence of 200 ms or less as a micropause, silence of 300 ms to 400 ms as hesitation, and silence of 500 ms or more as an unfilled pause. Pauses at clause or phrase boundaries are usually considered fluent and in the middle of a clause dysfluent (Ullakonoja 2011, 30). Indeed, pause location has been shown to be a stronger factor than pause

time and frequency in distinguishing fluent and disfluent speakers (Paananen-Porkka 2007; Tavakoli 2011). Pausing becomes more native-like when L2 proficiency increases (e.g., Riazantseva 2001), and the earlier the L2 learning starts, the more likely the learner is to acquire appropriate pausing (Trofimovich & Baker 2006). Toivola, Lennes, and Aho (2009) found some support for pausing being culturally determined: L2 Finnish speakers used shorter pauses than L1 Finnish speakers. Sajavaara and Lehtonen (1980) discovered that Finnish speakers have longer pauses, more pauses, and more incorrect pause placement in spontaneous speech than Swedish-speaking Finns. They also tend to use silent pauses when hesitating instead of filled pauses like Swedish speakers (Lehtonen 1981). The third, repair fluency, covers reformulations, repetitions, replacements and false starts (Tavakoli & Skehan 2005). It is often grouped together with breakdown fluency. For instance, Rossiter (2009) uses terms *hesitation phenomena* or *dysfluencies* which include also unfilled and non-lexical filled pauses. Research provides evidence for the negative effect of self-corrections, repetitions and false starts on fluency ratings (ibid.). Witton-Davies (2015, 317) points out that "repair reflects at the same time both a lack of accuracy […] and an interest in accuracy".

The measures should be chosen carefully for they may influence the outcome of the analysis (Tavakoli, Campbell, and McCormack 2016). Witton-Davies (2015) warns about using combined fluency measures, especially when investigating correlations between different measures. He suggests using articulation rate instead of speech rate and measuring length of turn and pause frequency separately instead of mean length of run (MLR). It is also important to consider contextual factors which influence fluency, "such as the topic, mode of interaction, register and setting" (Witton-Davies 2015, 317). Since fluency is interactional, a quantitative analysis of the fluency measures is not enough but needs to be complemented with listener judgments (Paananen-Porkka 2007, 97-103). One popular method of assessing perceived fluency is scalar ratings. People with different backgrounds (e.g. teachers of English as a second language (*ESL*), linguists, naïve listeners) have been shown to give similar fluency ratings, which refers to the fact that they have a similar idea of what is fluent and what is not (e.g., Derwing et al. 2004; Rossiter 2009). Reading tasks are common in data elicitation, because of the ease of administration (Segalowitz 2010, 43). More notably, the participants are free from conceptualization and macroplanning when reading aloud compared to spontaneous speech (ibid.). Although fluency research has been mainly quantitative, Peltonen and Lintunen (2016) note that qualitative analysis

is particularly useful for examining filled pauses and repair phenomena and can shed light on individual differences, too.

As mentioned earlier, the relationship between fluency and speech rhythm is still vague. The core components of fluency, pausing and speech rate, have been found to also contribute to speech rhythm. Hirvonen (1973, 26) noted that "speech rhythm plays a significant role in fluent speech". He argues that it is impossible to produce the correct speech rhythm with stuttering speech, and thus sentence rhythm and fluency are intertwined (ibid.). Adams (1979) and Paananen-Porkka (2007) both found that L2 speakers pause more frequently and at inappropriate places, which does not only affect fluency but also causes incorrect word grouping and disturbs the rhythmical pattern. Paananen-Porkka (2007, 341) discovered that when rating the speech rhythm of the students, the evaluators attended more to the frequency, duration, and location of pauses than erroneous use of weak forms, the lack of F0 variation, or the speech rate. Thus, she (2007, 350) infers that instead of sentence stress, "pausing may be a more important component of English speech rhythm". Consequently, Paananen-Porkka (2007) considers fluency as a quality of speech rhythm.

Another study by Dilley, Wallace, and Heffner (2012) sheds light on the relationship between perceived rhythm and fluency, but they admit that the exact mechanism is unclear. Since both trained and naïve listeners judged the least isochronous speech, the dual task conditioned production, also the most dysfluent, the results can be considered as a valuable example of the listeners' intuitions of these concepts. In addition, they were able to statistically show a strong correlation ($r = 0.92$) between the rhythm and fluency ratings by two distinct listener groups (ibid). Based on their results, Dilley et al (2012, 254) argue that some lexical sequences seem to require perceptual isochrony in order to maintain perceptual fluency. This could mean the appropriate use of vowel reduction, durational cues, and juncture. Tominaga (2011, 53) found rhythm as "the most contributory factor for fluency", similarly to Kormos and Denes (2004, 158) who discovered *pace* ("a temporal variable that also considers [stress]") as a strong predictor for fluency ratings. According to these findings, speech rhythm and fluency clearly have a strong relationship, but the nature of their interaction has not been explicitly defined. The following subsection will consider the concepts of rhythm and fluency from the perspective of second language acquisition (*SLA*).

2.3 Issues in learning L2 prosody

The majority of L2 pronunciation theories and models explain how individual phonemes are learned: for instance, Eckman's (1977) *Markedness Differential Hypothesis* assumes that the more marked (salient or rare) a sound in a speaker's native language is than the sound in L2, the easier it is to learn. Flege's (1992) *Speech Learning Model* posits that L2 sound acquisition is based on the L1 sound categories. Even completely new sounds can be learned fairly quickly if a new category can be established, but the similar sounds may not match the L1 categories fully and therefore be categorized incorrectly, causing errors in L2 pronunciation and perception. Yet, Trofimovich and Baker (2006, 26) suggest that theories explaining the acquisition of L2 segmentals apply to the acquisition of suprasegmentals, too. They found the processes of L2 phonology learning at the sentence level being similar to the ones at segment level. Perhaps, learning and processing prosody, including rhythm, relies on the differences between the smallest prosodic units in the learners' L1 and L2 (ibid.). It seems that the suprasegmentals contributing to speech fluency require "several psycholinguistic mechanisms at multiple levels of processing" (Trofimovich & Baker 2006, 25).

Prosodic differences between languages include average speaking fundamental frequency, variations of fundamental frequency, and speech timing differences (Flege 1981, 446). These influence the perceived speech rhythm and may be transferred from the L1 to L2 (ibid.). Li and Post (2014, 249) conclude that the process of L2 prosodic development is not uniform and depends on the L1. The development of various separate features constituting speech rhythm contribute directly to the L2 rhythm acquisition (ibid.). In addition, Kainada and Lengeris (2015, 269) found that even when learners transferred L1 prosodic features into their L2 speech, it still deviated from the L1 norms. The speech rate of the subjects' interlanguage was slower, pitch span narrower, and pitch level lower than in their L1. Barry (2008, 111-2) argues that all learners, regardless of their L1, have a tendency to overarticulate in comparison to native speakers. Even when the L1 and L2 have similar timing patterns, the learners tend not to differentiate enough between stressed and unstressed syllables. Jenkins (2005, 39) acknowledges that the highly rule-governed word stress in English is difficult to acquire especially for learners whose native language has fixed or relatively fixed word stress patterns such as Finnish, Polish, and Spanish. In the case of read aloud speech, it requires not only command of English sentence rhythm, basic intonation, and

articulatory motor skills, but also mastering the relationship between the spelling and pronunciation (Hirvonen 1973, 36).

As for the learning difficulties of L2 English prosody specifically for Finnish speakers, the most common problems are insufficient and inaccurate variation in F0, insufficient vowel reduction in unstressed syllables, incorrect segment durations interfering with word stress and leading to the disruption of rhythmic recurrence of stressed syllables, and incorrect juncture signaling using glottalization (Hackman 1978; Lehtonen, Sajavaara, and May 1977; Lehtonen 1979; Morris-Wilson 1981; Pihko 1994; Paananen-Porkka 2007). Most of these errors can be accounted for L1 transfer. For example, Eerola et al. (2012, 315) explain that in Finnish vowels and consonants can occur in short or long oppositions, regardless of the preceding or following sound, and the differences in quantity are not related to word stress. Lehtonen (1979, 45-46) argues that the problems in L2 fluency and rhythm are caused not only by transfer, but also monitoring, which makes the speech "too conscious, too careful and therefore not fluent". He sees the higher percentage of pauses in L2 English of Finns as a result of a false strategy of hesitating; Finns tend to be quiet while others use filled pauses (ibid.).

The role of vowel quality as a stress correlate differs greatly between English and Finnish: it is insignificant in Finnish but crucial in English (Lehtonen, Sajavaara, and May 1977, 79). The quality of Finnish vowels remains virtually the same regardless of the degree of prominence whereas English vowels are full in quality only in the stressed positions (ibid.). Morris-Wilson (1981, 200) assures that the flow of the speech, linking and blurring sounds, stress, rhythm, and weak forms are more of a psychological challenge than a physical obstacle for Finns. In fact, Peltola, Lintunen, and Tamminen (2014) demonstrated that it is possible for Finnish learners of English to achieve native-like L2 vowel qualities after explicit pronunciation instruction and that they benefit from the L1 transfer of durational features. Moreover, Ylinen et al. (2010) successfully trained Finnish speakers to use the spectral cues instead of duration in English vowel recognition. Pihko (1994, 85-86) discusses the phenomenon in English comprehension; processing difficulties related to the different timing patterns in Finnish and English are caused most notably by sound reduction. However, she found that Finnish upper secondary school students had become accustomed to English spoken at fast rate by native speakers, and thus could use the rhythm and stress phenomena as comprehension aid (Pihko 1997, 234).

2.3.1 L2 accentedness and comprehensibility

Previous research on perceived fluency often examines its relationship to *accentedness* and *comprehensibility*. The accentedness and comprehensibility of L2 learners have commonly been studied through listener judgements (e.g., Derwing & Munro 1997; Derwing et al. 2004). Derwing, Munro, and Wiebe (1998, 396) define accentedness as "the extent to which [… L2] speech differ[s] from [native speaker] norms". Derwing and Munro (2015, 8) note that perceived accent is the only type of accent. Foreign accent is composed by multiple factors, including segmentals, prosody, voice quality features, speaking rate, and fluency (Major 2007, 541). Comprehensibility, by definition, is the perceived level of understanding an utterance (Derwing, Munro, and Wiebe 1998, 396). Some studies use *intelligibility* instead of comprehensibility. Intelligibility, an objective measure of comprehensibility, was not featured in the questionnaire for methodological reasons: it is typically tested through listening comprehension tasks which would have been complicated and time-consuming to carry out online, as well as unfair to the first speakers rated.

The results by Purcell and Suter (1980) revealed L1, aptitude for oral mimicry, length of residence (in a country where L2 is widely spoken), and strength of concern for pronunciation accuracy as predictors for L2 speech accentedness judgements. Gender and years of instruction were not significant (ibid.). Other studies have added motivation, willingness to communicate, and a desire to improve pronunciation to the list of significant predictors (Derwing & Munro 2015, 44-46). The consistency of contact with native speakers, L2 use in different contexts, as well as positive attitudes towards the target culture have been revealed important as well (Moyer 2013, 18-19). In the present study, the learners had applied to and been accepted to study English in the university so it can be assumed that they are highly motivated and that their language aptitude is relatively high.

Several studies have investigated the linguistic features which affect the judgements of accentedness and comprehensibility: for instance, Kang (2010, 301) found suprasegmentals contributing independently to the ratings. The best predictors in accentedness scores are pitch range and word stress, rhythm, as well as vowel and consonant errors (Kang 2010; Trofimovich & Isaacs 2012; Saito, Trofimovich, and Isaacs 2017). Slow speech rate has made listeners judge the utterance more accented and less comprehensible, yet listeners also tend to give lower ratings to speech which is too fast (Major 2007). Prosodic features, such as word stress and rhythm, can account for as much as 50% of the variance in accentedness ratings for L2 speakers with diverse

L1 backgrounds (Trofimovich & Isaacs 2012). Comprehensibility scores, on the other hand, are affected by several linguistic domains, most significantly speech rate, word stress, fluency, vowel and consonant errors, lexis, and grammar (Hahn 2004; Kang 2010; Trofimovich & Isaacs 2012; Crowther et al. 2015; Saito, Trofimovich, and Isaacs 2017). Amongst novice raters, word stress has been found to influence both accentedness and comprehensibility scores (Trofimovich & Isaacs 2012). Saito, Trofimovich, and Isaacs (2017, 459) conclude that comprehensibility and accentedness "constitute two overlapping but distinct goals of L2 oral skill development".

Saito, Trofimovich, and Isaacs (2017, 459) confirm that while listeners with linguistic background or teaching experience can consistently evaluate complex linguistic categories that are less intuitive, also naïve listeners are able to reliably rate phonological, lexical, grammatical, and discourse structure aspects on a scale. Accentedness ratings may not be significantly different between native and non-native listeners (Major 2007), but judgements of accentedness are influenced by the listeners' residence in a country where their L1 is not widely spoken (Major 2010). Familiar accented speech seems to be the easiest to understand (Major et al. 2002, 175). Curiously, Major (2007, 541) found that even listeners who were unfamiliar with the language spoken were able to correctly judge the accent native or non-native. If the listeners are unfamiliar with the language they are evaluating, they will most likely focus on fluency (Major 2007, 549). Furthermore, Major et al. (2002) found that non-native rhythm, even when dissimilar to the listener's native language rhythm, can have an assisting or impeding effect on the listener comprehension.

As for non-native English teachers, Kim (2008) found that ESL students may hold unjustified negative attitudes against non-native English teachers. Some of them judge the teacher's speech difficult to understand only on the basis of their foreign accent. Levis et al. (2016, 918) see no reason to believe that pronunciation could not be taught equally well by non-native English-speaking teachers as native English-speaking teachers. However, there is a risk that learners lose their confidence in the non-native teacher due to obvious errors whereas careless pronunciation from a native-speaking teacher is possibly ignored altogether (Levis et al. 2016, 917). The next subsection discusses different approaches to teach L2 pronunciation.

2.3.2 Pronunciation instruction goals

In the research and instruction of L2 pronunciation, two main paradigms have been competing for the default status: the *nativeness* and the *intelligibility principle* (Levis 2005, 370-71). The nativeness principle was the dominant paradigm before the 1960s. Native-like pronunciation was considered attainable and desirable until research had consistently showed that the pursuit of nativeness of pronunciation is unrealistic for the majority of learners and teachers. The intelligibility principle focuses on successfully delivering the message: simply put, learners need to be understood. It acknowledges that the success of communication does not depend on accentedness, no matter how strong (ibid.). The application of intelligibility principle increased along World Englishes movement and the growing need for mutual intelligibility (Levis 2016, 429-430). Further research is required to determine the primary factors which influence the intelligibility and comprehensibility the most and apply this knowledge into pronunciation instruction (ibid.).

Today, published pronunciation materials still follow the model of General American (GA) and Received Pronunciation (RP) to a great extent although these accents are not spoken by the large majority of native speakers (Levis 2005, 371-72). For instance, Singapore English speakers who had RP as a pronunciation model, found it difficult to understand Estuary English, which is much more widely spoken in England. Deterding (2012, 10) points out that even the rhythmic structures differ between the conventional stress-based inner-circle varieties of English (e.g. UK and US) and syllable-based outer-circle varieties (post-colonial societies). Thus, preferring only prestige models can be counterproductive for L2 learners' spoken language comprehension.

Tergujeff (2014), too, demands intelligibility as a goal instead of nativeness. She claims that pronunciation instruction in Finnish comprehensive school and upper-secondary school focuses on single segments rather than rhythm, stress, and intonation. Tergujeff (ibid.) argues that the teaching of suprasegmentals is often neglected altogether, particularly if prosody is absent in school books. She found it positive that some teachers report using global English in their teaching in addition to British and American pronunciation models. Foreign accent is indubitably more acceptable in international contexts today, especially when often it is the lingua franca, not a native language of anyone involved (ibid.). As for non-native stress, it seems that even L1 speakers can

be moderately flexible with anomalous stress due to the differences and change of acceptable stress patterns over time in global English varieties (Jenkins 2005, 40).

Another distinction in pronunciation instruction can be made between *bottom-up* and *top-down* models. Bottom-up model begins with single phonemes and proceeds to larger units whereas top-down model means starting with the more global features, such as intonation and advancing to smaller units over time. Lintunen (2014, 183) recommends top-down approach for communicative instruction because the instruction of suprasegmentals has been shown to benefit English learners' comprehensibility compared to segmental instruction (e.g., Derwing, Munro, and Wiebe 1998). Regardless, Levis (2016, 432-433) notes that emphasizing suprasegmentals over segmentals is problematic for two reasons: first, segmentals and suprasegmentals depend on and influence each other, and second, not all suprasegmentals are equally important or learnable. Morris-Wilson (1981, 196) demonstrates the problem of separating the two levels in teaching coarticulation: first the learners practice pronouncing separate phonemes accurately, and later they are told to ignore precise articulation and blur sounds together. Barry (2008) recommends pronunciation practice focused on fundamentally segmental properties for all language learners because together they constitute the appropriate prosodic patterns of speech. In other words, pronunciation instruction which is essentially segmentally focused can still contribute to the learning of correct speech rhythm (Barry 2008, 115). Therefore, Barry rejects the need to introduce the concept of stress-timing additionally (ibid.).

In this theoretical section I have aimed to introduce and explain the key concepts and previous research findings as well as methods, which are required to understand the context and the methodology of the present study. Next, we shall move on to the empirical part of the thesis, beginning with the methodology.

3 Methodology

This section will introduce the methods and data of the present study. Before discussing these in detail, it is worthwhile to view my hypotheses for each research question:

1) My first research question asks whether the speech rhythm and fluency of advanced English learners changes during a pronunciation course according to native speaker ratings. I hypothesize that on an individual level there are changes in speech rhythm and fluency of the learners but on average they are not significant. The students will most likely pay more attention to accuracy in segments than to fluency and speech rhythm in the posttest utterances because separate phonemes are practiced more during the course and emphasized in assessment. I expect non-native intonation and accentedness to distract some evaluators and cause them to give lower ratings on fluency and rhythm as these aspects have been found to influence one another (Kang 2010; Trofimovich & Isaacs 2012; Saito, Trofimovich, and Isaacs 2017). The most advanced students are likely to have no notable changes in their speech, or a negative change due to the increased focus on segmentals as well as nervousness in the testing situation.
2) The second question was 'if there are changes, which acoustic measures do they correlate with?' Previous studies on L2 English rhythm of Finns have shown that their most common problem is vowel reduction (Hackman 1978; Lehtonen, Sajavaara, and May 1977; Lehtonen 1979; Morris-Wilson 1981; Pihko 1994; Paananen-Porkka 2007). Thus, my hypothesis is that the students with biggest changes in their pre- and posttest ratings have learned to use weak forms (i.e., to reduce the vowel quality) and how to alternate stressed and unstressed syllables. They will probably also speak at faster rate, as previous fluency studies predict.
3) My third research question asks whether the perceived speech rhythm, fluency, accentedness, and comprehensibility correlate with each other. Similar to the results of Dilley, Wallace, and Heffner (2012) and Tominaga (2011), I expect a strong correlation between speech rhythm and fluency. Since it may be difficult for the raters to distinguish rhythm and fluency features from accentedness, these scores will most likely correlate as well.

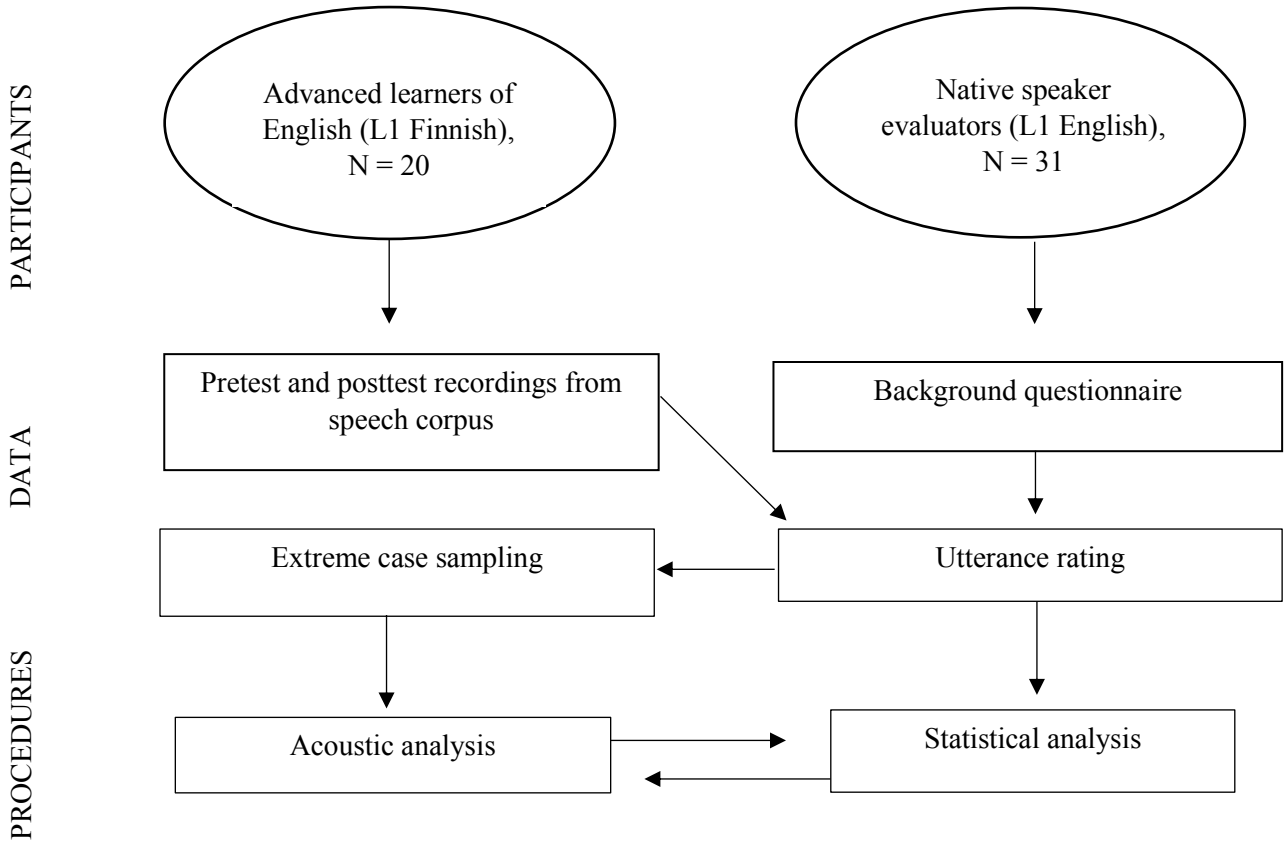

Figure 1 The research design of the present study

My mixed methods approach employs both quantitative and qualitative methods in order to achieve the most comprehensive results. It is based on previous studies on both rhythm and fluency. For instance, Derwing and Munro (2015, 9) find the comparison of rating data with acoustic measures much more useful than the information extracted from these approaches separately. In the present study, the data for the native-speaker evaluation were first selected and edited into short audio samples. Then the questionnaire was created and conducted to elicit how native English listeners perceive the production of advanced learners of English. These ratings were used to determine the extreme cases for the subsequent acoustic analysis. Finally, eight samples, two from each of the four learners, were analyzed auditorily and acoustically in order to reveal the features which the native-speaker evaluators attended to. The research design is visualized above in Figure 1.

3.1 Participants

The research design of the present study included two groups of participants: 20 advanced learners of English and 31 native-speaker evaluators. As for the evaluators, Derwing and Munro (2013, 181) acknowledge that L2 learners can be involved in accentedness rating but it is yet unknown which level of L2 knowledge is required to evaluate comprehensibility or other features of speech. Therefore, I chose to recruit only native speakers as evaluators. The following subsections discuss the subject groups in more detail.

3.1.1 Advanced English learners

In the present study, the English students were selected by convenient or opportunity sampling: partially for their availability and willingness to volunteer, but also for their role as advanced L2 students (Dörnyei 2007, 98-99). I was able to obtain pronunciation recordings from a speech corpus collected by the English Department. To be consistent, all recordings were from the fall 2017. These 45 first-year major English students had consented on a separate form to have their productions used in research. The students were tested in two separate sessions: the first, pretest, was recorded at the beginning of the academic year and the second, posttest, after the pronunciation course. Thus, I had 90 recordings in total to choose from.

To protect the confidentiality of the students, their names were removed from the recordings and each was given a code (for example, F12: F = female, 12 = running student number). The only shown background information therefore was gender. In addition, all students speak Finnish as their native language and are assumed to have completed the upper secondary school according to the Finnish national core curriculum. Instruction of English typically begins in the third grade and lasts until the graduation after twelve years, usually taught as the advanced syllabus language. The target CEFR (Common European Framework for Languages) level is B2. A learner on this level "[c]an produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses" (Council of Europe 2018).

At the time of recording, the language laboratory at the department was unavailable, and therefore the acoustic environment was not optimal. Based on the time limitations concerning the evaluation, I selected 20 students with the least background noise and echo. Students who spoke very quietly were also excluded to avoid additional volume adjustments, as well as students with

a voice quality that might have distracted listeners too much from the task (e.g., creaky voice). Fifteen of the selected students were female (75%) and the rest (25%) male. Eleven (55%) of these students participated in the American English (AmE) pronunciation course, and nine (45%) took the British English (BrE) course. These distributions were close to the original proportions of the 45 students.

Spoken English is a mandatory pronunciation course in the first-year basic studies for the English language major students. The students are tested before and at the end of the course by recording them reading aloud a certain text. The course contains 90 minutes of instruction per week for one semester. According to the course teachers (both AmE and BrE), one whole lesson is dedicated to weak forms and stress, and another to rhythm, linking, and intonation. In addition, fluency-related issues (stress, linking, etc.) are practiced throughout the course in different contexts, integrated into phoneme exercises. Students' proficiency of prosody has some influence in the assessment process but mostly in situations where students' performance is in between two grades, in which case fluency can raise it. According to the University Study Guide, Spoken English proficiency goals are that "students will improve their English pronunciation as well as the *rhythm* and *fluency* of their speech *compared to their baseline*. Students will master the standard English phonology and transcription. They will be able to analyze and improve their pronunciation. Students will learn to receive feedback on their pronunciation and to give feedback to other students. Students will be able to identify and produce the sounds of standard English".

3.1.2 Native-speaker evaluators

Both native and non-native English speakers, as well as ESL teachers, linguists and novice listeners have been recruited as evaluators in previous studies. Tominaga (2011) found that non-native English speakers could be stricter in their evaluation than native speakers, possibly because they are more aware of different factors having had to learn the language themselves. I wanted the evaluators in the present study to represent typical native speakers of English, which the learners could have to interact with, preferably without any foreign linguistic evaluation experience (i.e. *novice raters*). I followed Dörnyei's (2007, 99) suggestion of the sample size for a correlational research and aimed for at least 30 participants. Next, I will illustrate the background of the evaluators based on their questionnaire responses.

Table 2 Age and gender distribution of the evaluators

		Age					
		20-29	30-39	40-49	50-59	60-89	Total
Gender	Female	5	6	2	1	2	16
	Male	1	11	1	2	0	15
Total		6	17	3	3	2	31

Table 2 shows the age and gender distribution of the participants. Sixteen females (51.6%) and fifteen males (48.4%) completed the questionnaire. The average age was 37.9 (range 20-88, median 35, mode 39), the majority of the participants (N =17) being between 30 and 39 years. As for the accents (Figure 2), the majority spoke American (67.7%, N = 21), six British (19.4%), three New Zealand (9.7%), and one Australian English (3.2%). If the minority accents are combined into non-rhotic accents, they account for 32. 3% (N = 10) of the total. Three of the native speakers were bi- or multilingual (9.7%): one person's other native language was Arabic, another's Spanish, and the third one's Urdu and Punjabi.

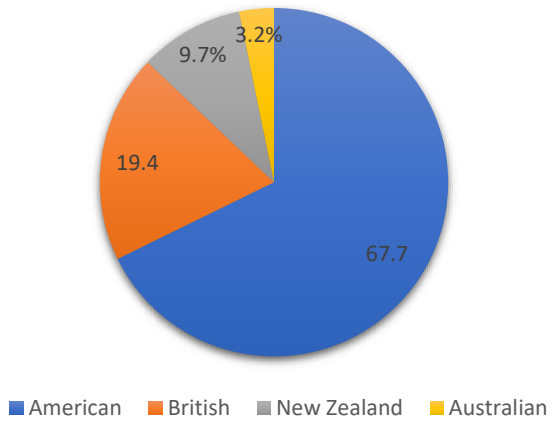

Figure 2 Accent distribution of the evaluators, N = 31

More than a third of the respondents had a master's degree or higher education (35.5%, n = 11), or bachelor's degree 35.5% (n = 11). Three of them had an Associate degree (9.7%), and six high school or equivalent (19.4%). Thus, the group is generally well-educated.

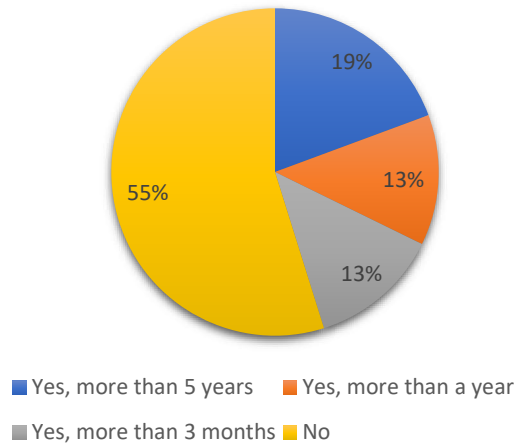

Figure 3 Residence in non-English speaking countries, N = 31 ("Have you lived in a non-English speaking country for a longer period?")

Fourteen participants responded having lived in a non-English-speaking country (Finland, El Salvador, Guatemala, France, Nepal, Japan, Turkey, Thailand, The Netherlands, Colombia, Spain, and India). Seventeen evaluators had not lived in a non-English-speaking country (Figure 3). A clear majority had studied a foreign language (90.3%, N = 28), which included both Germanic and non-Germanic languages, as well as languages traditionally divided into stress-timed and syllable-timed languages. Three participants had studied basic or intermediate Finnish. Approximately one third (32.3%, N = 10) had been engaged in language teaching, most of them English (N = 9) out of which one ESL informal tuition, and one Spanish.

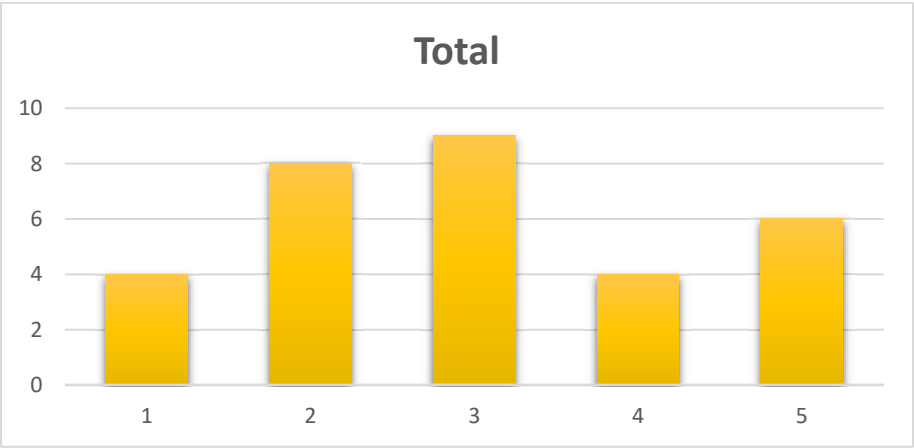

Figure 4 Familiarity with the accent produced by Finnish speakers of English (1 = 'I have never heard a Finnish person speaking English', 5 = 'I hear Finnish people speaking English very often')

At the end of the questionnaire the evaluators were asked how familiar they are with the Finnish accent of English (see Figure 4 above). Most of them answered '3' (N = 9) and '2' (N = 8). Six answered hearing it very often ('5') and four had never heard a Finnish person speaking English ('1'). Finally, none of the participants reported having hearing impairments or using hearing aids. As a conclusion, most of the evaluators are globally-oriented, well-educated people who know foreign languages, interact with people from other countries, and hear foreign English accents.

3.2 Data

Next, I will present the data of the present study, which were elicited in two different ways: the recorded audio samples were collected earlier by the English Department in a pretest-posttest condition, and the native-speaker ratings were compiled through an online questionnaire.

3.2.1 Pretest and posttest recordings

As mentioned earlier, the audio samples had been recorded in the fall semester 2017. The students had some time to familiarize themselves with the text before both pre- and posttest recordings. They were allowed to read it through and make notes. The productions were recorded individually with a digital audio recorder. The testing situation was similar regarding the room and equipment. However, the students did not know exactly what was tested in the pretest, but after the course they knew very well what they were expected to have learned. This awareness may be heard in the posttest recordings as nervousness: shaking of the voice, more self-corrections, focusing on separate words and segments, or faster speech rate. All of these can make the speech less fluent. It is also important to acknowledge that the phonetic qualities of read aloud speech do not equate the ones of spontaneous speech. Hirose and Kawanami (2002, 97) found that "dialogue speech generally shows wider dynamic ranges in its prosodic features, such as in tone and rhythm, as well as a higher speech rate than read speech". On the other hand, Munro and Derwing (1994) found no evidence for an advantage in spontaneous speech condition when analyzing accentedness ratings.

The pretest text was a text which has been designed to include every English vowel and consonant in various environments, often referred to as 'Wolf passage' (Appendix 1). There are some problematic parts in the text concerning rhythm. Following the example of Deterding (2012, 14-15), I only focused on the following three phrases (the first two in the same sentence): 'as soon

as they heard him', 'full of concern for his safety', and 'that had just escaped from the zoo'. This was due to the multiple function words in which vowels may be reduced (*as, as, of, for, that, had, from, the*), and the content words in which the first syllable would be unstressed (*concern, escaped*) (ibid.). Since vowel reduction and stress are core components of rhythm, these parts should reveal how rhythmic the speech of the learners is. The posttest text was a modified article from BBC website, and thus authentic and more complex in grammar, vocabulary and sentence structure (Appendix 2). One of the selected students was absent at the time of the posttest and therefore read a different text, also a modification of a BBC article (Appendix 3). In the posttest texts, the selected sentences and phrases included fewer proper nouns and more grammatical words with potential weak forms, as well as words with varying stress patterns (*decision, transition etc.*).

There were several criteria for choosing which utterance from each recording to feature in the questionnaire: I wanted there to be long enough of a pause before and after the phrase, so that cropping it would be easier and that the evaluators would not feel like something was missing from the sentence. I tried to avoid puffing sounds and coughing, but not errors, repetitions, and self-corrections. Audacity audio software was used for cropping the utterances, the length of which varied between 3 and 12 seconds (average length 6.5 s). The number of syllables varied from 14 to 26. Derwing et al. (2009) argued that short audio samples are better for pronunciation and fluency evaluation because the longer the recordings, the longer the total duration of the experiment. A long experiment can cause listener fatigue, and the large number of samples had to be taken into account in the present study. Another problem with longer utterances is that the evaluators may only remember the first or last parts of the speech.

For control purposes, two samples from a native speaker of English (American male, 23 years old) were added: one phrase from the pretest text and one from the posttest text. He received the same instructions as the students and read both texts entirely. Since some of the recordings of the students included background noise, the recording of the native speaker was conducted imitating this acoustic environment so that it would not stand out from the learner samples.

3.2.2 Native-speaker questionnaire

Derwing and Munro (2015, 9) consider comprehensibility ratings or intelligibility measures as "the gold standard for measuring pronunciation improvement", since acoustic measures cannot always tell which changes listeners actually react to. That is why the questionnaire in the present

study included also evaluation of the utterance comprehensibility, in addition to rhythm, fluency, and accentedness. The questionnaire formatting was based on Dörnyei (2007, 109-110): After the questionnaire title, there was a short introduction which stated the general purpose of the study (see Appendix 4). To avoid listener bias, it was not revealed to the evaluators that almost all of the speakers were non-native English learners and native Finnish speakers. I reminded the participants that there are no right or wrong answers, guaranteed confidentiality, and requested honest answers. Specific instructions followed after the background information section, then the questionnaire items, here the 42 audio samples. Finally, there were two more questions and a space for additional comments. After submitting their responses, the listeners were thanked for their contribution, and offered the researcher's email address for any questions and for entering the gift card drawing.

The specific instructions guided the evaluators to pay attention to the following aspects: speech rhythm ("A steady, regular pattern of beats in the utterance created by the alternation of stressed and unstressed syllables"), accentedness ("The degree to which the pronunciation of an utterance sounds different from a native English speaker norm"), fluency ("How smoothly and rapidly an utterance is spoken; sounds are connected and fluid, the duration and placement of pauses are appropriate"), and comprehensibility ("How easy it is to understand what is being said"). A methodologically similar study was carried out by Kennedy and Trofimovich (2010), in which the students' accentedness, comprehensibility, and fluency were evaluated from pretest and posttest recordings. The definitions were adapted from Munro and Derwing (1999) and Derwing et al. (2004). The questionnaire items consisted of 42 utterances in sets of ten to twelve, which were each to be rated according to:

1. Speech rhythm (1 = very unnatural; the stress is placed and marked incorrectly - 9 = very natural; the stress is placed and marked correctly)
2. Accentedness (1 = heavily accented; the speaker sounds like a non-native speaker of English with a strong foreign accent - 9 = not accented at all; the speaker sounds like a native speaker of English)
3. Overall fluency (1 = not fluent at all; sounds are disconnected, speech rate is too slow or too fast, pauses disrupt the flow of speech - 9 = very fluent; sounds are connected, delivery is smooth and effortless)
4. Comprehensibility (1 = very difficult to understand - 9 = very easy to understand)

Accentedness, fluency, and comprehensibility were operationalized based on previous research (e.g., Derwing & Munro 2013; Munro & Derwing 1994; Munro & Derwing 1999; Levis et al. 2016), and rhythm was set on the same 9-point Likert scale. The wording of the opposite ends of rhythm was not straightforward: since the evaluation in the present study leans strongly on the listener intuition of their native language, I assumed that the word 'natural' was appropriate, even though ambiguous or loaded words should usually be avoided (Dörnyei 2007, 108). It was also speculated whether a 5- or even 7-point scale would have resulted in sufficient variability in the learners scores. Isaacs and Thomson (2013) found no difference between experienced and novice raters in 5-point or 9-point rating scale conditions when judging L2 speech samples for comprehensibility, accentedness, and fluency. However, raters experienced more difficulties differentiating between mid-scale steps in the 9-point scale condition.

It was important to keep the length within 30 minutes to avoid participant fatigue (Dörnyei 2007, 110). I did not write down the phrases which the listeners were about to hear and rate in order to avoid influencing the comprehensibility scores. Also, I did not want the raters to compare the learner productions to a certain spoken model, and therefore did not add native-speaker produced utterances as an example. To keep the questionnaire from exceeding 30 minutes time limit, there was no practice round. Instead, after testing the questionnaire, the order of the audio samples was changed slightly so that the first production of each phrase was as comprehensible as possible. For piloting the questionnaire, I asked a few people to test it in case for any malfunctions, unclear terms, or missing response options. I also wanted to know if it was possible to complete the questionnaire in the estimated 30 minutes, which was confirmed by all testers, myself included.

3.3 Procedures

The main procedures of the present study were the creation, distribution, and analysis of the questionnaire, as well as the subsequent acoustic analysis based on the results of the statistical analysis. Furthermore, the rating scores and acoustic measures were compared to find answers to the research questions.

3.3.1 Distribution and analysis of the questionnaire

The online questionnaire was created on Google Forms, a free service provided by Google. The questionnaire distribution was conducted as non-probability or snowball sampling (Dörnyei 2007,

98): I contacted a few people who are native English speakers and asked them to complete the questionnaire as well as to share it further. The link to the questionnaire was also shared on social media platforms with a description of an appropriate participant. Several people shared the link on their own accounts. The reason for this type of distribution technique was that access to native English speakers with varying backgrounds can be rather difficult in Finland. I wanted to recruit a diverse group of raters to get a more global perspective. A known problem of respondent self-selection is that when anybody can choose to participate or opt out, the sample may not represent the target population and thus the generalizability have to be carefully considered (Dörnyei 2007, 100-101). To make participating more appealing to the less likely evaluators, the participants were offered a chance to win a $25 Amazon gift card, which was then sent to the winner after the questionnaire had been closed. Many of those who chose to complete the questionnaire did not enter their name into the drawing, so assumingly they were not all motivated by the prize.

By this method, 31 responses were collected from native speakers of English. The listening sessions were not time-restricted: the evaluators were allowed to listen to the recordings and change their responses as many times as they needed, similar to Kennedy and Trofimovich (2010). Statistical analyses, which consisted of basic descriptive statistics, Paired samples t-tests, Pearson's correlation coefficient, Guttman Split Half Reliability coefficient, and Cronbach's alpha, were conducted on SPSS and Microsoft Excel. The rater comments on the questionnaire were considered for additional insight into the listener experience. The rating scores from the questionnaire can be found in the results section.

3.3.2 Extreme case sampling and acoustic analysis

Extreme case sampling was used to select the learners for the acoustic analysis. Dörnyei (2007, 153) views extreme case sampling as a valuable tool which can single out cases that reveal new information of the subject and can lead to new conceptualizations and propositions. By extreme case sampling it was possible to select only a couple speakers, and therefore have more resources to conduct a post-hoc analysis as well as a deeper qualitative analysis to investigate the reasons for the deviance. The subsequent acoustic analysis was conducted on Praat, a computer software for phonetic analysis. The eight selected samples were first analyzed for the fluency measures which were articulation rate, number of silent and filled pauses, mean duration of unfilled pauses, location of pauses and number of repairs. Following the example in Cumming (2010, 166) the utterances

were syllabified according to the pronunciation of the speakers. It is possible that native speakers of English would pronounce the same words with less syllables due to sound elision or coarticulation.

After fluency, speech rhythm was considered. The analysis focused on the variation in pitch and amplitude, and the alignment of their peaks, vowel reduction, duration of stressed and unstressed syllables, as well as linking sounds. Due to the background noise, the analysis was conducted mostly manually from waveform and wideband spectrograms. There were certain limitations in the spectrographic analysis, also discussed by Low (2006) and Deterding (2012): for instance, the segment or syllable boundaries were not always discernible. This was one of the reasons for excluding some acoustic measures of isochrony from the present study (e.g., interstress intervals). The next section will present the analyses with the most important findings.

4 Results

In this section I shall present the findings from the quantitative and qualitative analyses. First, the statistical analysis of the questionnaire responses will be discussed and then the auditory and acoustic features of the eight samples of the four selected speakers are analyzed in detail.

4.1 Questionnaire responses

The audio samples from twenty advanced learners of English were rated by 31 native English speakers with various backgrounds. These rating scores were then analyzed statistically in order to determine the extreme cases, here the learners with the biggest change in their pre- and posttest samples. Additional analyses were required to define the reliability of the ratings as well as the relationship between the rated aspects. Native speaker scores were excluded from the statistical analysis and examined separately.

Table 3 The range, mean, and standard deviation of the scores by the four rated aspects

	N	Range	Mean	Std. Deviation
Rhythm mean	42	5.52	5.33	1.56
Accentedness mean	42	6.26	4.47	1.60
Fluency mean	42	5.52	5.58	1.54
Comprehensibility mean	42	5.90	6.57	1.26
Valid N (listwise)	42			

Table 3 above summarizes the mean values of each rated aspect: the comprehensibility of the learners received the highest mean score (6.57), fluency the second highest (5.58), then rhythm (5.33), and accentedness the lowest (4.47). The variation in the scores was wide: The range of given scores was the largest in accentedness (6.26) and the smallest in rhythm and fluency (5.52). Standard deviation was the largest in accentedness as well (1.60).

When pre- and posttest rhythm and fluency are viewed in bar charts (Figures 5 and 6), it is clear that some learners demonstrated changes after the pronunciation course. However, half of these changes are negative in rhythm (N = 10) and more than half in fluency (N = 13). In addition, accentedness was improved by half of the participants and decreased by nine, while one retained

the same mean score, and in comprehensibility seven improved and 13 decreased their scores. Some of the changes are seemingly marginal and thus cannot be used as an evidence for or against the influence of the course.

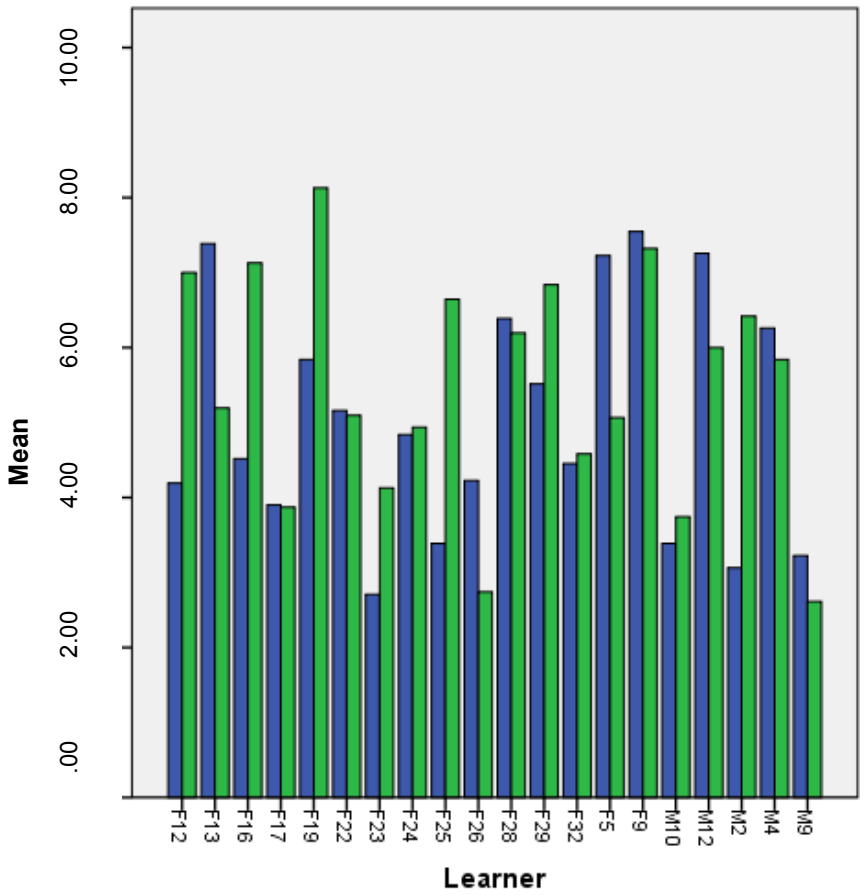

Figure 5 Pretest (blue) and posttest rhythm (green) means by each learner

For the following extreme case sampling, overall oral proficiency index was calculated as an average of all four mean values, and the learners were ranked accordingly. Mean values were used instead of medians to make the differences between learners more distinct. The average of the mean scores for the learner group was 5.40. Eight out of twenty students improved their overall oral proficiency scores, eleven decreased them, and one learner had the same score in both pre- and posttests. The biggest change, based on the mean scores, occurred in speech rhythm. The next largest change was in fluency, and then in comprehensibility, but the latter was a negative change.

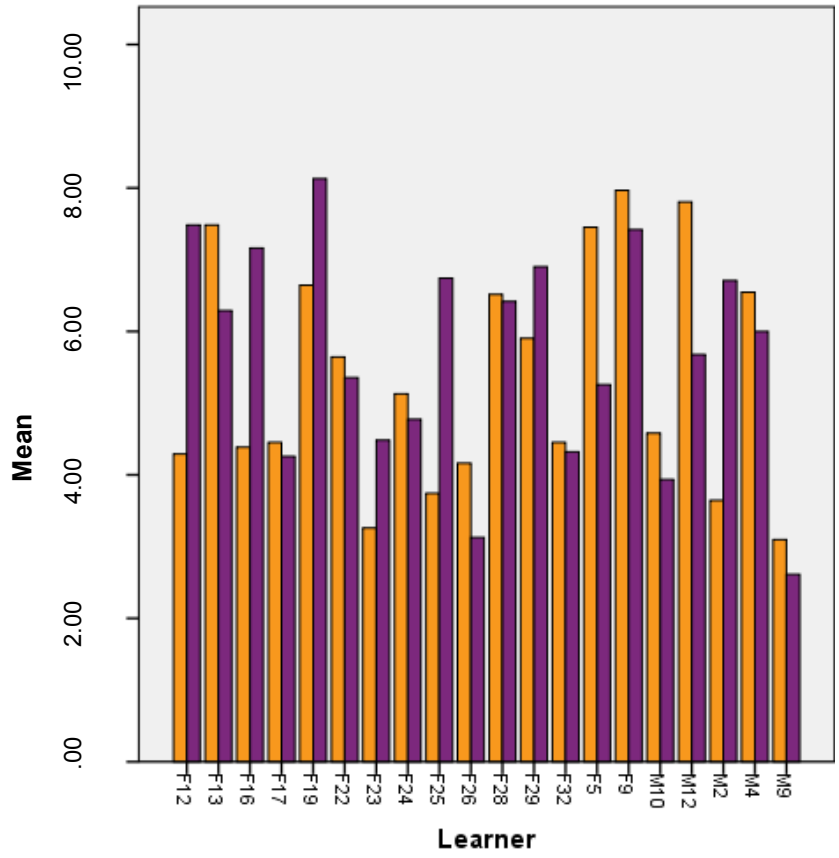

Figure 6 Pretest (orange) and posttest (purple) fluency means by each learner

In average, the students improved their rhythm, fluency and accentedness and decreased their comprehensibility. This makes the overall oral proficiency score difference slightly positive (see Table 4). However, when tested statistically with the Paired samples test (Table 5), none of the changes are significant ($p > 0.05$). If examined by the overall oral proficiency score rankings, three out of the top ten learners in the pretest improved their ranking, while four out of the ten lowest scoring learners improved it. The lower posttest comprehensibility score could indicate that there was something in the posttest text itself which reduced the comprehensibility, such as vocabulary, sentence structure, or context. On the other hand, it could also be the result of the difference between the scores of the weakest and the strongest learners.

Table 4 Mean scores and their changes for the learner group (N = 20)

	Rhythm	Accentedness	Fluency	Comprehensibility	Overall oral proficiency score
Pretest	5.02	4.26	5.36	6.65	5.32
Posttest	5.47	4.38	5.65	6.43	5.48
Change	**+0.45**	+0.12	**+0.30**	-0.22	+0.16

Table 5 Differences between pre- and posttest aspects according to the ratings

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	pretest rhythm - posttest rhythm	-.450	1.711	.383	-1.251	.351	-1.176	19	.254
Pair 2	pretest accentedness - posttest accentedness	-.123	.909	.203	-.548	.303	-.603	19	.554
Pair 3	pretest fluency - posttest fluency	-.295	1.664	.372	-1.074	.484	-.793	19	.437
Pair 4	pretest comprehensibility - posttest comprehensibility	.218	1.198	.268	-.343	.778	.813	19	.426

As for the relationship between rhythm and fluency, the mean rhythm and fluency scores given to all samples (N = 42) are close to each other, but fluency scores are a little higher. The mean rhythm score was 5.33 with the standard deviation of 1.56 while the mean fluency score was 5.58 (sd 1.54). This could indicate that the raters were able to evaluate the two aspects as two separate aspects. Correlation between rhythm and fluency scores was tested with Pearson

Correlation test (Table 6). The correlation is very strong (r = 0.98) and statistically significant (p = 0.00).

Table 6 Pearson Correlation between rhythm and fluency mean scores for each sample

		Rhythm mean	Fluency mean
Rhythm mean	Pearson Correlation	1	.98**
	Sig. (2-tailed)		.00
	N	42	42

**. Correlation is significant at the 0.01 level (2-tailed).

Overall, all four rated aspects correlated positively and significantly with each other (see Appendix 5). In both pretest and posttest scores, rhythm has the strongest correlation with fluency (0.98) but it correlates considerably with comprehensibility (pretest 0.88, posttest 0.87) and accentedness (both 0.85) as well. Both accentedness (pretest 0.86, posttest 0.87) and comprehensibility (both 0.91) correlate the most with fluency. For most of the learners (N = 15), comprehensibility rating increased or decreased along the three other variables, but in the five other cases fluency alone seemed to affect the comprehensibility score.

The influence of different listener variables was tested on Excel. If different subgroups give similar ratings, the external validity increases. When divided into three age groups, the mean score by the youngest (20-35) was 5.38, 5.45 by the middle group (36-55), and 5.48 by the oldest (56-88). Hence, the older the raters, the higher scores they gave, although the differences are marginal. As for the impact of gender, mean score from males was 5.51, and 5.59 from females. Again, the difference is scarce and may be caused by other variables. American English listeners gave a mean score of 5.40, and listeners with non-rhotic accents 5.37. The mean score by more educated respondents (bachelor's degree or higher, N = 22), 5.46, was slightly higher than by respondents with less education (5.39).

Bi-/multilingual background, foreign language studies, nor teaching experience seemed to have an effect on the overall mean scores. Respondents who have lived in a non-English speaking country (N = 14) gave an overall mean score of 5.40, whereas the mean score was 5.34 from respondents (N = 17) who have not lived in a non-English speaking country. Familiarity with Finnish accent made an interesting distinction: the mean score was 6.48 from those who have never

heard Finnish accent in English (N = 4), and 4.92 from those who hear English produced by Finns very often (N = 6). Out of those who rated the native speaker's accentedness '9' (N=11), 10 spoke American accent and 1 British, so it can be concluded that the speakers' accentedness scores were influenced by the evaluators' own accent. When comparing the backgrounds of the raters who gave the lowest and the highest mean scores, the only notable differences are that out of the four evaluators who gave the highest ratings, three had lived in a non-English speaking country and two had been engaged in language teaching. The seven highest mean ratings were from AmE speakers but otherwise no patterns could be detected.

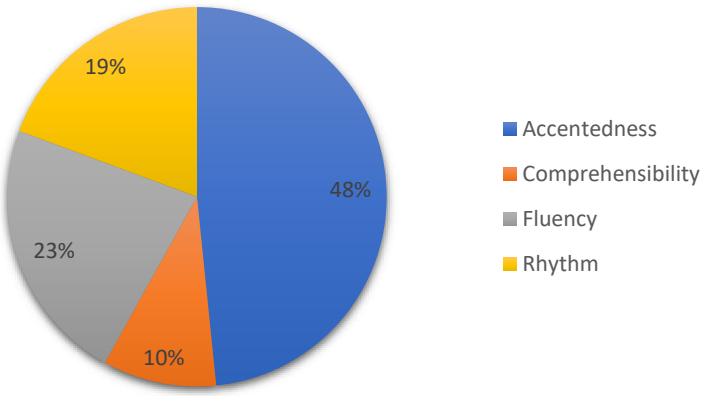

Figure 7 The most prominent feature according to the raters

At the end of the questionnaire, the participants were also asked to select the most prominent feature in the utterances (Figure 7). Almost half of them chose accentedness and a few elaborated that even a thick foreign accent did not make the utterance incomprehensible if the delivery was otherwise smooth and uninterrupted. On the other hand, one rater noted that if the accent was not familiar, the utterance was more difficult to understand. It was also pointed out that some of the speakers sounded accented but possibly because they spoke another variety of English. Most of the speakers were thought to be accented enough to distinguish them from native speakers. One rater noticed that when the speakers read aloud, they seem to focus rather on the pronouncing of the words than the flow of the utterance. Similarly, the speakers were thought to be mostly clear on a segmental level, but word stress, speed, and connectedness varied. Disconnected sentences were mentioned to hinder comprehension, and the differences between the speakers in pausing and speed were heard as noticeable. One rater found most of the speakers speaking too fast, and

therefore difficult to understand. Another noticed that rhythm affected comprehensibility more than accentedness. Rhythm was also considered a part of accentedness. The speakers were thought to be unsure of where to place the sentence stress. One rater who reported traveling a lot and having friends with different native languages highlighted comprehensibility in the samples.

To find out whether the evaluators were consistent in their ratings, Cronbach's alpha and the Guttman Split-Half Coefficient were tested (see Table 7). The number (0.72) indicates high correlation and strong relationship. Cronbach's alpha was also calculated for each aspect separately (Appendices 6 and 7): for rhythm and accentedness it was 0.95 and for fluency and comprehensibility 0.96. Therefore, it can be concluded that there was acceptable agreement amongst the raters in evaluating all four aspects as well as pre- and posttest items.

Table 7 Reliability Split Analysis on the rated items

Cronbach's Alpha	Part 1	Value	.96
		N of Items	4[a]
	Part 2	Value	.97
		N of Items	4[b]
	Total N of Items		8
Correlation Between Forms			.56
Spearman-Brown Coefficient	Equal Length		.72
	Unequal Length		.72
Guttman Split-Half Coefficient			.72

a. The items are: pretest rhythm, pretest accentedness, pretest fluency, pretest comprehensibility.

b. The items are: posttest rhythm, posttest accentedness, posttest fluency, posttest comprehensibility.

Based on the rater comments and the statistical analysis, it seems that native speakers, both with and without linguistic or teaching background are able to rate the speech of non-native speakers quite analytically and consistently. Yet, some comments reflect the global perspective on pronunciation: all aspects are connected and influence each other.

Finally, to select the extreme cases, the learners with the biggest differences in pre- and posttest scores were identified. According to their overall oral proficiency, learners F12 (+2.49), F25 (+2.45), M2 (+2.17), and F5 (-1.96) would qualify. When examining the rhythm scores, the most improved learners are M2 (+3.36), F25 (+3.26), F12 (+2.81) and F16 (+2.61) and based on

fluency scores, F12 (+3.19), M2 (3.06), F25 (+3) and F16 (+2.77). Since the present study focuses on speech rhythm and fluency, I decided to examine the samples of F12, F16, F25 and M2 more closely in the following acoustic analysis.

4.2 Acoustic analysis of the extreme cases

Based on the quantitative analysis of the questionnaire responses, four learners with the biggest changes in rhythm and fluency scores were selected. As discussed above, on the group level the differences in scores were not significant but I shall now take a closer look at the learners whose rhythm and fluency scores increased the most: F12, F16, F25, and M2. A Paired samples t-test was conducted to find out whether their pre- and posttest scores were significantly different. As can be seen in Table 9 on the next page, F16 accentedness score difference is the only nonsignificant one.

The speech samples of these four learners were analyzed acoustically with Praat software in order to find the changes in fluency and speech rhythm. The fluency measures used were articulation rate, the number, duration, and location of pauses, as well as the number of repairs. The articulation rate for each speaker was calculated by dividing the number of syllables by the total phonation time. Micropauses (0.2 seconds or less), marked as (.), were not added to the total pause duration, and not taken into account when calculating articulation rate. The pauses were categorized as either boundary pauses or mid-unit pauses. In the transcribed utterances, dysfluencies are marked inside curly brackets ({}) and a period marks falling intonation. The changes in fluency measures are summarized in Table 8 below.

Table 8 Changes in fluency measures, pretest and posttest compared

Subject	Articulation rate	Number of unfilled pauses	Number of filled pauses	Mean duration of pauses	Number of repairs
F12	+1.69	-1	0	-0.05	-1
F16	+1.1	-4	0	-0.47	-1
F25	+2.48	-3	0	+0.30	0
M2	+1.11	-4	0	-0.21	-2

Table 9 Paired Differences of the pretest (A) and posttest (B) samples

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Rhythm	F12A - F12B	**-2.81**	1.87	.34	-3.49	-2.12	-8.36	30	**.00**
Accentedness	F12A - F12B	-2.19	2.60	.47	-3.15	-1.24	-4.70	30	**.00**
Fluency	F12A - F12B	**-3.19**	1.85	.33	-3.87	-2.51	-9.60	30	**.00**
Comprehensibility	F12A - F12B	-1.77	1.78	.32	-2.43	-1.12	-5.54	30	**.00**
Rhythm	F16A - F16B	**-2.61**	2.57	.46	-3.55	-1.67	-5.67	30	**.00**
Accentedness	F16A - F16B	-.61	2.03	.36	-1.36	.13	-1.68	30	.10
Fluency	F16A - F16B	**-2.77**	2.43	.44	-3.67	-1.88	-6.35	30	**.00**
Comprehensibility	F16A - F16B	-1.10	2.32	.42	-1.95	-.25	-2.64	30	**.01**
Rhythm	F25A - F25B	**-3.26**	2.34	.42	-4.12	-2.40	-7.76	30	**.00**
Accentedness	F25A - F25B	-1.65	2.18	.39	-2.45	-.84	-4.19	30	**.00**
Fluency	F25A - F25B	**-3.00**	2.19	.39	-3.80	-2.20	-7.62	30	**.00**
Comprehensibility	F25A - F25B	-1.87	2.67	.48	-2.85	-.89	-3.91	30	**.00**
Rhythm	M2A - M2B	**-3.36**	2.33	.42	-4.21	-2.50	-8.01	30	**.00**
Accentedness	M2A - M2B	-.65	1.70	.31	-1.27	-.02	-2.11	30	**.04**
Fluency	M2A - M2B	**-3.07**	2.57	.46	-4.01	-2.12	-6.64	30	**.00**
Comprehensibility	M2A - M2B	-1.61	1.98	.36	-2.34	-.89	-4.54	30	**.00**

After fluency, speech rhythm was considered. The acoustic analysis focused on the overall pitch range, amplitude, and the alignment of their peaks, vowel reduction and duration of stressed and unstressed syllables. Linking and elision phenomena were also considered. Prominent syllables were identified in Praat spectrograms according to Kang (2010, 305), having longer duration, greater amplitude, and higher pitch than unstressed ones. *Anticipated* and *delayed F0 peaks* are parts of the rise or fall of a stressed syllable, respectively (Paananen-Porkka 2007). The changes in rhythm measures can be seen in Table 10. The samples from the native speaker were also analyzed for comparison. This particular native speaker is a very fast speaker who omits sounds and links them together frequently, which is also reflected in the relatively lower comprehensibility scores. His speech does not show much variation in F0, except for a slight rise on every stressed syllable. The amplitude peaks, on the other hand, are consistent and regular. Using Praat, a spectrogram for each utterance was drawn (Figures 10 to 17). In the spectrograms the blue line follows pitch (Hz) and the yellow one intensity (dB). Next, the changes in pretest and posttest utterances will be analyzed by each speaker, starting from F12.

Table 10 Changes in rhythm measures, pretest and posttest compared

Subject	F0 range	F0 and amplitude peaks	Vowel quality	Duration	Linking, elision
F12	slightly less variation	better matching	more vowel reduction	better use of durational cues	more linking sounds
F16	less variation	both match fairly well, more anticipated F0 peaks	no vowel reduction	less use of durational cues	more linking sounds
F25	more variation	both have some alignment and anticipated F0 peaks	more vowel reduction	better use of durational cues	more linking sounds
M2	less variation	slightly more alignment	slightly more vowel reduction	better use of durational cues	no linking

Auditorily, the pretest utterance by F12 includes hesitation, which is manifested as pauses, repair, and slower speech rate. This is reflected in the speech flow. As for the fluency changes in pre- and posttest samples, F12 grew her articulation rate, decreased the number and duration of pauses, and used less repairs. The pauses in both utterances are at clause borders and do not disrupt the flow (see examples 1 and 2 below).

(1) however (.) not long after (.) a wolf (.) that had just {eks} escaped from the zoo (0.41) was looking for a change from its usual diet of chicken and duck. (total duration 9.17)

(2) all change is threatening. (0.36) but failures of imagination and bad decisions can also be fatal. (total duration 5.72)

(F12 pre- and posttest utterances)

Regarding rhythm measures, the pretest pitch and amplitude peaks are not congruent. In 'that had' the middle /t/ is omitted, otherwise there are no elisions or linking. In 'escaped' the first syllable receives greater intensity but the second one higher pitch. There are a lot of anticipated F0 peaks, as well as some delayed ones. The posttest sample shows a bigger range in F0 and amplitude, and their peaks match better. The words in the phrase "failures of imagination and" are all linked together, with reduced vowels in 'of' and 'and'. The last syllable of 'decisions' receives the highest pitch, whereas amplitude is almost constant throughout the word. In 'imagination' the F0 peaks fall on the first and the last syllable while the first one receives the greatest intensity. In general, F12 does not make a large distinction between the pitch and intensity of stressed and unstressed syllables (see Figures 8 and 9 on the next page).

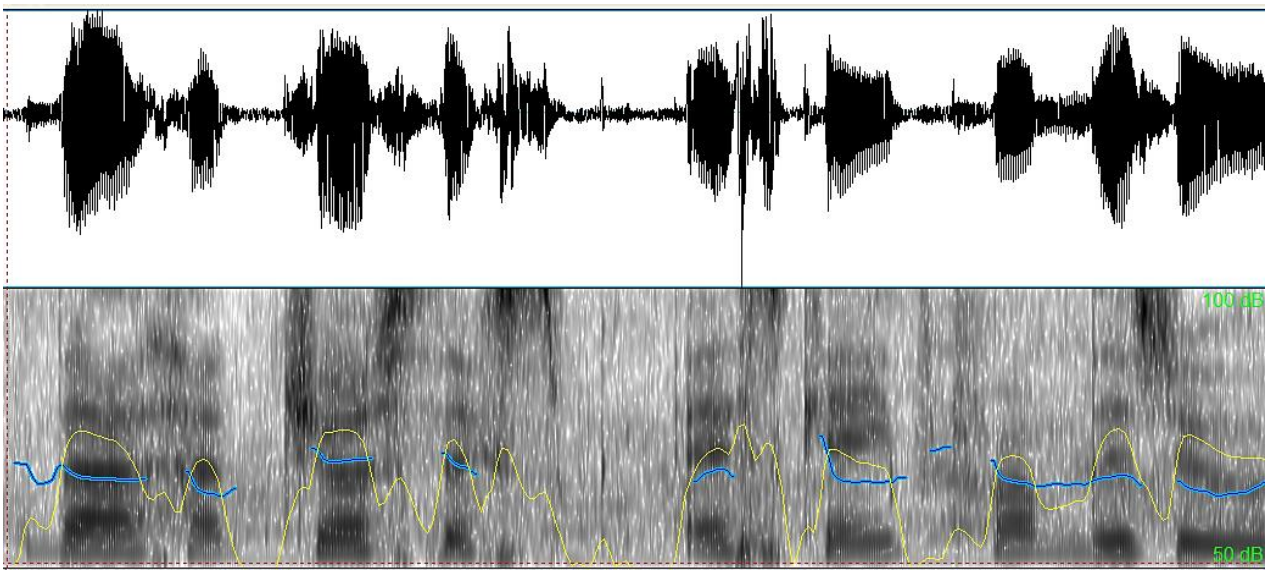

Figure 8 F12 saying "a wolf (.) that had just {eks} escaped from the zoo".

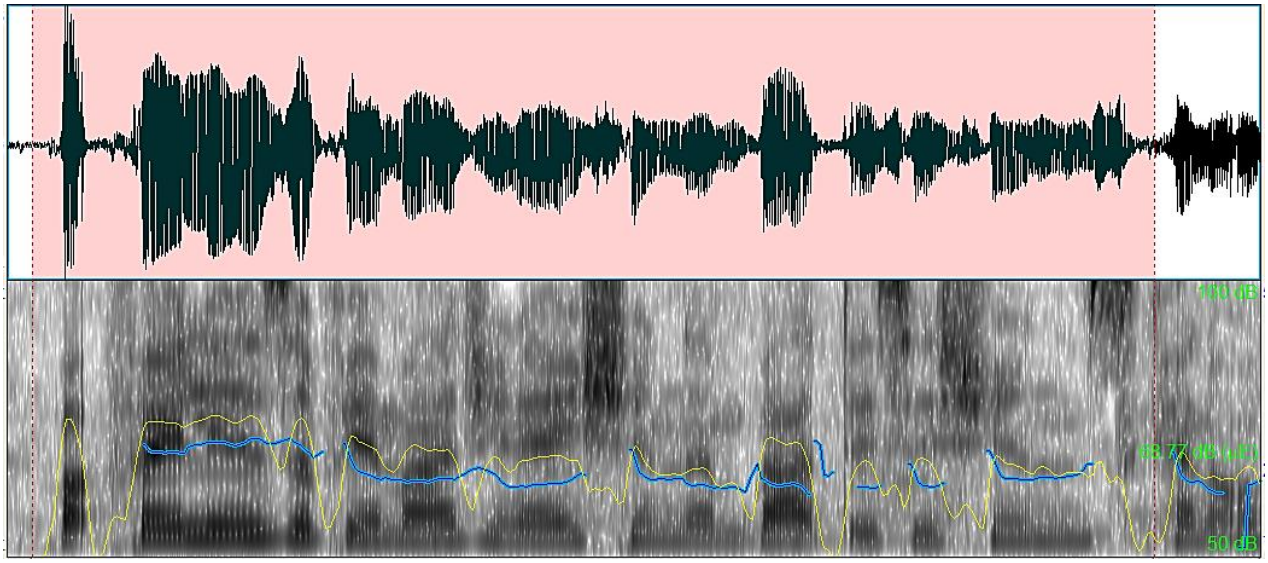

Figure 9 F12 saying "but failures of imagination and bad decisions" (painted area).

Learner F16 pauses frequently in the pretest sample, and for a rather long time. Albeit relatively long, the first three pauses can be considered functional while the last one is a mid-clause pause. Her articulation rate speeds up and pauses are dropped off completely in the posttest utterance (examples 3 and 4).

(3) however (0.43) not long after (0.48) a wolf that had just {ekseip} escaped from the zoo (0.33) was looking for a change (0.64) from its usual diet of chicken and duck. (total duration 10.09)

(4) the great transition is a tale of how it turned out right. (total duration 3.19)

(F16 pre- and posttest utterances)

Regarding the rhythm of F16, in the pretest utterance the second syllable of 'escaped' has greater amplitude and higher pitch, and the peaks match. Some F0 peaks are anticipated but otherwise they match the amplitude peaks. The words 'tale of' are linked and the second syllable of 'turned out' is omitted. In 'transition' the second syllable has greater intensity but the last one has the highest pitch. The F0 peaks are mostly anticipated. In general, F16 makes a difference between the amplitude and F0 of stressed and unstressed syllables in both samples (see Figures 10 and 11).

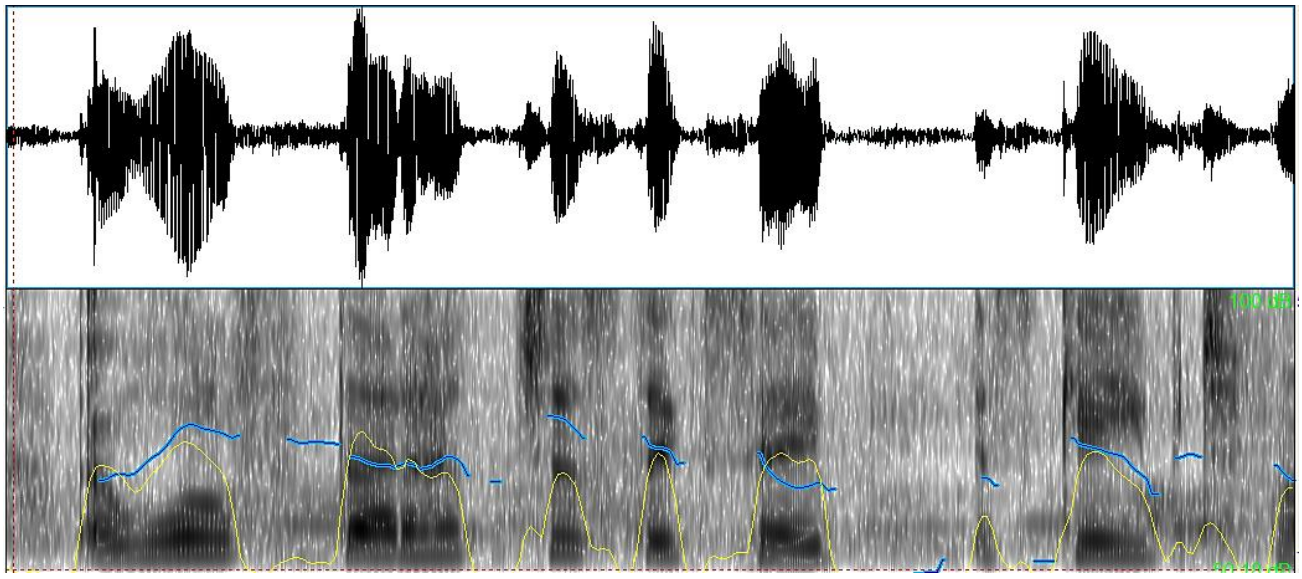

Figure 10 F16A saying "a wolf that had just {ekseip} escaped from"

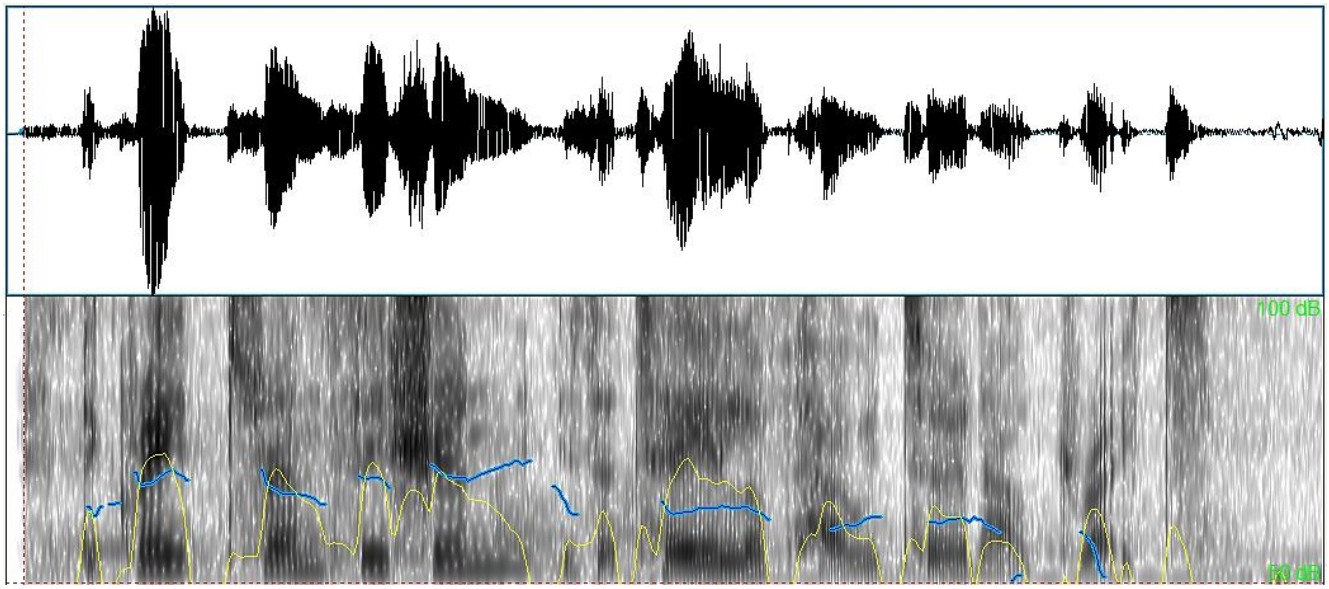

Figure 11 F16B saying "the great transition is a tale of how it turned out right"

Auditory analysis of the pretest speech of F25 reveals hesitation, missing weak forms, and no vowel reduction. The lack of weak forms and vowel reduction could be accounted for Finnish transfer. There is a great decrease in the number of pauses and their total duration. In the pretest there are two mid-clause pauses, but none in the posttest. F25 improved her articulation rate considerably (see examples 5 and 6 below).

(5) however (.) not long after (0.48) a wolf that had (0.30) just escaped from the zoo (0.46) was looking for a change from its usual diet of (0.34) chicken and duck. (total duration 10.86)

(6) all change is threatening. (0.69) but failures of imagination and bad decisions can also be fatal. (total duration 5.72)

(F25 pre- and posttest utterances)

The pretest utterance of F25 shows some alignment of amplitude and pitch peaks which are mostly anticipated (Figures 12 and 13 below). In 'escaped' the amplitude of both syllables is the same, while the F0 is slightly higher on the second syllable. There are some anticipated pitch peaks as well as some matching with amplitude peaks. The posttest of F25 has more variation in F0, with

some pitch and amplitude peaks coinciding although most pitch peaks are anticipated. F25 posttest speech demonstrates linking between all words in “of imagination and”. As a general observation, she does not make a distinction between the pitch and intensity of unstressed and stressed syllables.

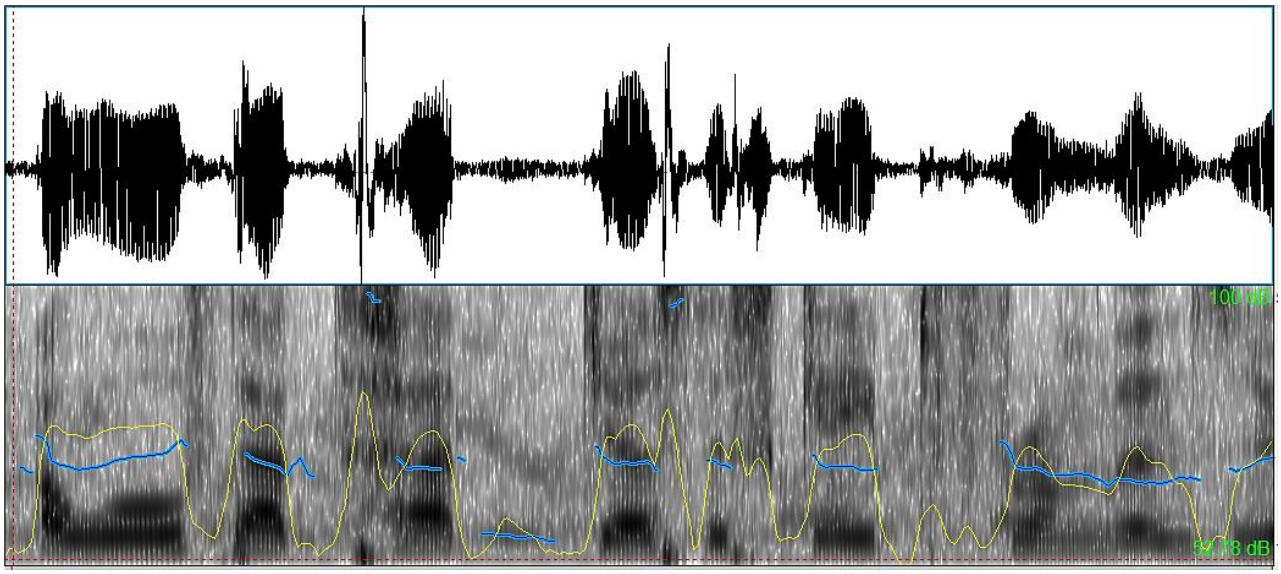

Figure 12 F25 saying “a wolf that had (0.30) just escaped from the zoo”

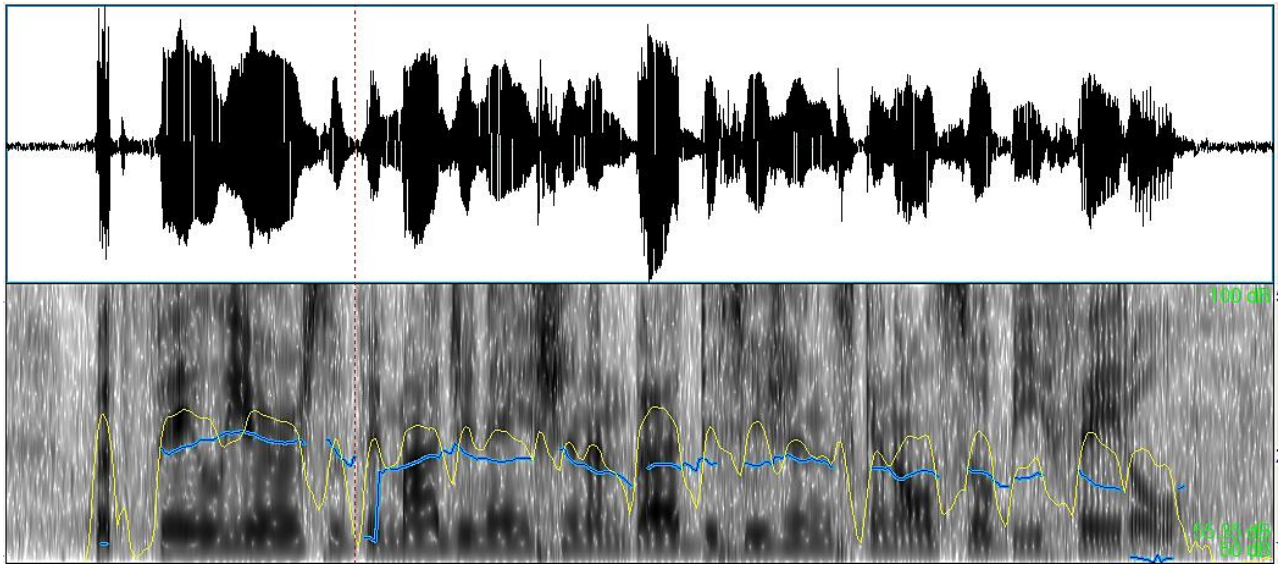

Figure 13 F25 saying “but failures of imagination and bad decisions can also be fatal”.

Auditorily, in his pretest utterance M2 hesitates a lot, uses self-correction, pauses within word groups, repeats a word, and does not reduce unstressed vowels. Measured by the fluency features, in the posttest samples his articulation rate increases, the number and mean duration of pauses decreases, and repairs are reduced completely. The location of the pauses changes as well: There are two relatively long mid-clause pauses in the pretest, but only one functional pause in the posttest (examples 7 and 8).

(7) however (0.71) not long after (.) a wolf that {has} (0.58) that had just escaped from the zoo (0.50) was looking {fo:r} (0.68) for a change from its usual diet of chicken and duck. (total duration 10.7)

(8) all change is threatening. (0.41) but failures of imagination and bad decisions can also be fatal. (total duration 5.85)

(M2 pre- and posttest utterances)

As to the rhythm features of M2 pretest utterance (Figures 14 and 15 on the next page), there is a little alignment in F0 and amplitude but no linking sounds. In 'escaped' the second syllable receives slightly greater amplitude and higher pitch. The F0 variation is scarce. In the posttest he shows again some alignment of the peaks and no linking. In 'imagination' the second syllable has the greatest intensity and the last the highest pitch, but the vowel in the prominent syllable (the fourth) is the longest in duration. The pitch barely peaks. In 'decisions' the second and the third syllable have equal amplitude and F0, and the vowel length is almost the same. The posttest utterance shows some distinction between F0 and amplitude peaks of stressed and unstressed syllable. As a general observation, the native speaker samples and the male learner samples differ greatly from the female samples in terms of pitch variation.

To summarize the results of the acoustic analysis, faster articulation rate, smaller number of unfilled pauses and their location at clause boundaries, as well as absence of repairs seemed to result in higher fluency scores for these four learners. Similar to previous studies, the four Finnish learners did not use filled pauses in either pre- or posttest speech. As for the influence of different acoustic features of rhythm, it seems that the learners utilize these to different extents, and they each have their own combination. All of the speakers have, however, somewhat matching amplitude and F0 peaks in either both pre- and posttest speech or just the latter. Thus, this could

be one influencing factor. It is also to be noted that the learners with the biggest difference in their rhythm scores, M2 and F25, both showed more vowel reduction and made a larger distinction between the duration of stressed and unstressed syllables in their posttest speech. Linking was increased by F12, F16 and F25 but not M2. Based on the present findings, speech rhythm and fluency go hand in hand, and for that reason it is difficult to tell the extent to which different rhythmic features affect the perceived fluency and, on the other hand, how much speech rate, pausing, and repairs influence the perceived rhythm.

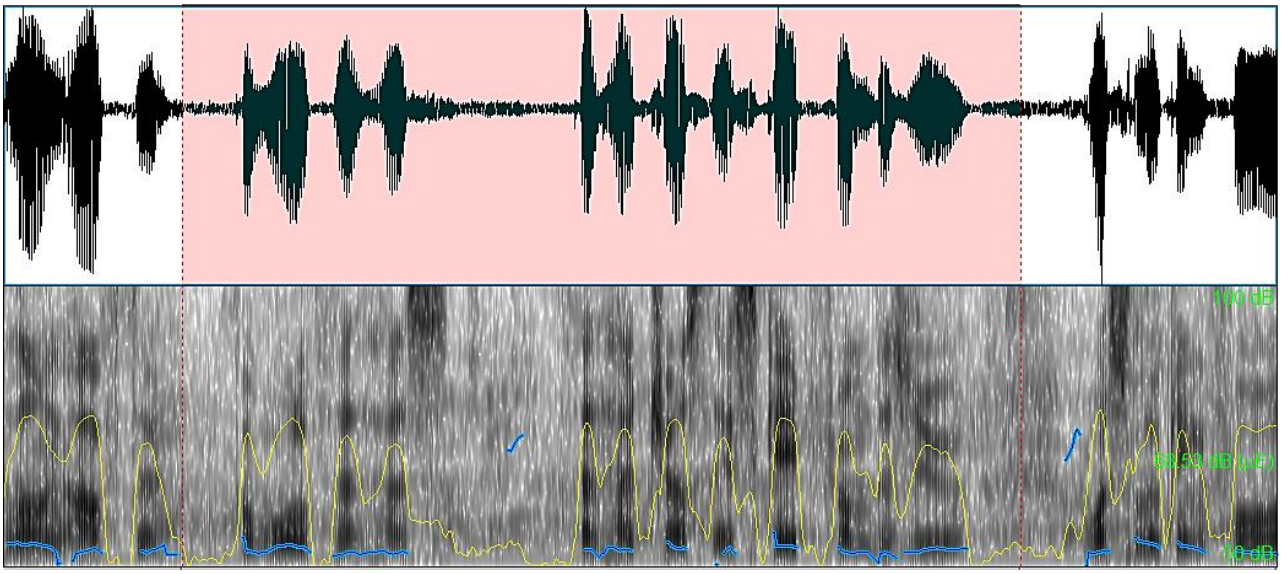

Figure 14 M2 saying "a wolf that {has} (0.58) that had just escaped from the zoo" (the painted area).

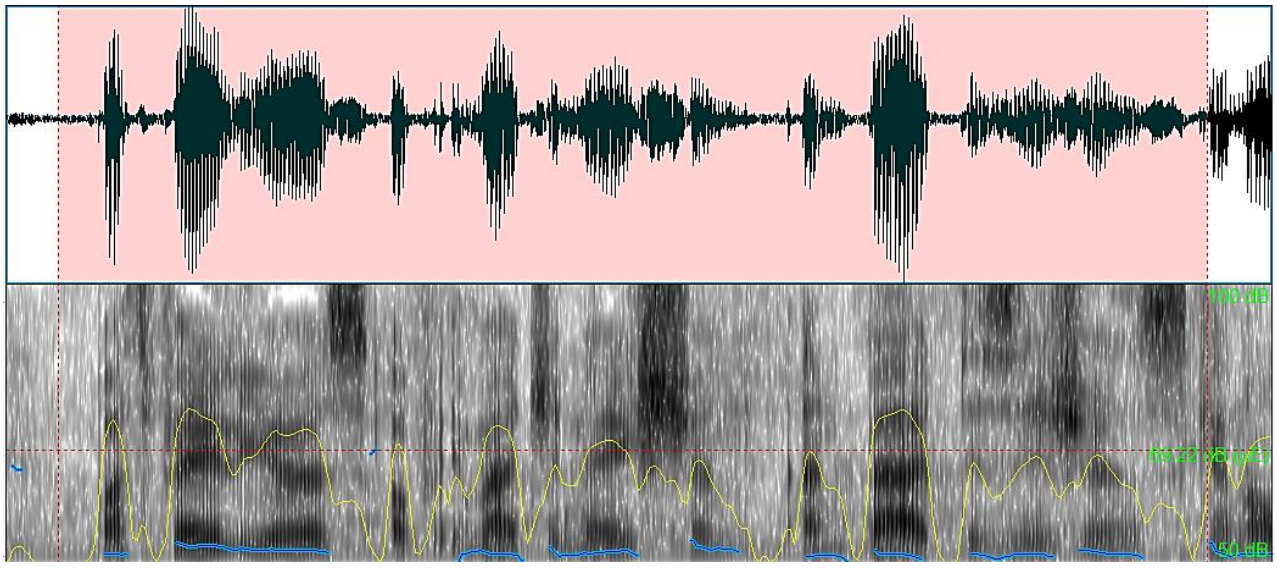

Figure 15 M2 saying "but failures of imagination and bad decisions can also be fatal" (painted area).

5 Discussion

In this section, the results of the present study are discussed and interpreted from a larger perspective and connected to SLA and pronunciation instruction. Moreover, the reliability and validity of the study are evaluated, and suggestions for further research are provided. Let us now return to the research questions presented in the introduction, beginning from the first.

5.1 Interpretation of the results

The first research question asked whether pronunciation instruction influences the speech rhythm and fluency of Finnish advanced learners of English according to native-speaker evaluations. The results of the quantitative analysis showed that half of the learners (N = 10) improved their speech rhythm and the other half lowered their rhythm scores. The fluency scores were lowered by thirteen learners and improved by seven. In addition, accentedness was improved by half of the participants and decreased by nine, while one learner retained the same mean score. In comprehensibility seven students improved and thirteen lowered their scores. Eight out of twenty students improved their overall oral proficiency scores, eleven lowered their scores, and one learner had the same score in both pre- and posttest.

It was hypothesized that the students with the highest scores in the pretest would show no notable changes in the posttest, or that these changes would be negative. Based on the results, it seems that the hypothesis was partially correct: seven out of the top ten students in the pretest lowered their ranking, but also six out of the ten lowest ranking students demonstrated a negative change. Therefore, it cannot be concluded that the scores for pretest samples would predict the posttest scores. A common statistical phenomenon that can be seen in the results is regression to the mean: the students with the lowest scores seem to improve, and the students with the highest scores demonstrate negative development. As hypothesized, the changes were not significant on the group level, but the four selected learners demonstrated significant changes in almost all aspects, especially rhythm and fluency. It is possible that non-native intonation and accentedness distracted some evaluators and caused them to give lower ratings in rhythm, but based on the rater comments, disfluent speech may have had a bigger impact.

The rating scores from the native-speaker evaluators indicate that the impact of the pronunciation course was positive for some and negative for other students. Although all learners speak the same L1, the development of L2 prosodic features does not necessarily progress

consistently, as mentioned earlier (Li & Post 2014), and may depend on individual differences. However, like Witton-Davies (2015, 317) points out, it should be borne in mind that the changes in the speech rhythm and fluency of the learners are also affected by contextual factors, here particularly the testing situation and the read texts. The testing situation may have caused anxiety in the learners, they may have focused more on segmentals than on suprasegmentals, and they may have tried to read the text faster increasing the risk of repairs. Furthermore, the difference between the pretest and posttest texts is notable. The posttest text is authentic and more complex in terms of vocabulary and sentence structure. It is also possible that the shorter length of posttest samples could have led to fewer errors and pauses and therefore higher scores. If these external and temporary factors are to be blamed for the negative changes instead of the instruction, perhaps the English students in the present study would show more improvement in the long run.

As mentioned before, some characteristics of the evaluators probably influenced the scores they gave. An interesting detail was that the evaluators who were familiar with Finnish accent gave lower mean scores. This could indicate that these listeners recognized the accent immediately and thus rated the speakers more accented. Major (2007, 549-550) has termed this kind of a concept *displaced foreign accent detection*. It happens when the listener focuses on detecting the speaker's L1 features in their speech instead of how it deviates from the L2 norms. The mean scores from female listeners as well as AmE speakers were slightly higher, which in turn could be connected to the learner characteristics: there were more female than male learners and two more learners with American accent compared to the BrE speakers. Further interpretation of the influence of different listener variables would require a detailed analysis by each rated aspect which was not conducted in the present experiment.

The second research question dealt with the acoustic correlates of the perceived rhythm and fluency. It was hypothesized that the biggest changes are due to the students learning how to use weak forms (i.e. to reduce the vowel quality), to alternate stressed and unstressed syllables, and to speak at faster rate. These students were expected to be the ones who received the lowest scores for their pretest productions, which was partly correct: F12, M2, and F25 ranked amongst the lower half of the learners in the pretest. Looking at the results from the acoustic analysis, it seems that faster articulation rate, fewer unfilled pauses and repairs, matching pitch and intensity peaks, vowel reduction, linking sounds, and better use of durational cues affected the rhythm and fluency scores positively. There was no single acoustic feature that could be said to be responsible

for the improved rhythm, but rather different kinds of combinations of the analyzed features. The finding is similar to Cumming's (2010, 209): perceived rhythm seems to be created by multiple integrated acoustic cues. Based on the results of the present study, it cannot be determined how much the rhythm measures affected to the perceived fluency or how much the fluency features affected the perceived rhythm. Similar to Paananen-Porkka (2007), the possibility that pausing influences rhythm even more than sentence stress should be considered. Speech rate, too, is a likely factor in rhythm as previously suggested (e.g., Arvaniti 2012).

On the other hand, the native listeners may subconsciously compensate for the predicted periodicity (e.g., Ratchke and Smith 2015). As Couper-Kuhlen (1993, 14) already argued, it is possible that the perceived regularity or periodicity does not correspond directly to any measurement of the acoustic signal. This would contribute to the theory that speech rhythm is primarily a perceived phenomenon. As observed in the acoustic analysis, even when all the syllables in a multisyllabic word are given equal pitch, intensity, and duration, somehow the syllable which is expected to be stressed sounds prominent. The role of transfer can only be considered based on the previous research on Finnish learners of English as I did not have recordings of the learners reading aloud in Finnish and therefore could not compare their L1 and L2 rhythm or fluency. The missing weak forms and vowel reduction were assumed to be caused by Finnish transfer, similar to Hackman (1978), Lehtonen, Sajavaara, and May (1977), Lehtonen (1979), Morris-Wilson (1981), Pihko (1994), and Paananen-Porkka (2007).

The third and final research question asked how the rated features correlate with each other. As hypothesized and in line with the results by Dilley, Wallace, and Heffner (2012) and Tominaga (2011), speech rhythm and fluency scores had a strong correlation ($r = 0.98$). Similar to previous studies, fluency and rhythm correlated with comprehensibility more than accentedness (e.g., Derwing et al. 2004). Accentedness, too, correlated the most with fluency. Consistent with the results of Hahn (2004), Major (2007), Kang (2010), Trofimovich and Isaacs (2012), Crowther et al. (2015), and Saito, Trofimovich, and Isaacs (2017), it seems that the accentedness scores were influenced by word stress and rhythm. Speech rate probably also had an impact on both accentedness and comprehensibility ratings. Comprehensibility scores seemed to be affected by word stress and fluency. Since all learners read a pre-determined text, the previously discovered role of lexis and grammar in comprehensibility rating could be eliminated. These results imply the essential role of fluency in L2 speech and justify the increasing focus on it in language classrooms.

In addition, as rhythm and fluency correlate strongly, rhythm should not be neglected in pronunciation instruction.

5.1 Educational implications

One of the main motivations for this type of study is to bring up issues connected to the learning of L2 pronunciation, prosodic features in this case. The results of the present study are, therefore, interpreted from the perspective of pronunciation instruction: What kind of impact did the pronunciation course have in the rhythm and fluency rated by native speakers and, on the other hand, measured from the acoustic signal? Which features should be emphasized in instruction if the goal is to sound native or to be more intelligible? Derwing and Munro (2015) note that rhythm research findings cannot be directly interpreted as pedagogical recommendations but based on previous studies it does seem that it is well-grounded to teach L2 rhythm. According to my results, speech rhythm and fluency are closely connected, and hence both of them should be a fundamental part of the L2 syllabus.

I support the communicative goal of pronunciation instruction at schools: as long as the learner can be understood by native and non-native English speakers, it does not matter if they sound foreign. However, learners who wish to become English teachers or other language professionals should pay more attention to their pronunciation as they will be the models for their future students. A nonnative English teacher who is able to produce even the most difficult phonemes, natural rhythm, and intonation can be a huge source of inspiration to their students. Unfortunately, pronunciation courses like *Spoken English* offered by the University of Turku are not very common in higher education. I agree with Paananen-Porkka (2007) and Tergujeff (2014) on their recommendation for providing more phonetic training for English teachers. On the other hand, learning to pronounce by one specific model of English (BrE or AmE) may not be the best option regarding comprehensibility.

When considering useful practices in pronunciation instruction, studies show that raising learners' awareness of (at least segmental) phonetic features can improve their L2 pronunciation. Kennedy and Trofimovich (2010, 171) noticed that students' language awareness measured by the number of qualitative journal comments during a pronunciation course correlated positively with the pronunciation ratings. Derwing and Munro (2015, 106) recommend classroom activities with a focus on prosody, such as shadowing, mirroring, dialogues, and self or peer video monitoring.

Particularly useful for L1 Finnish learners of English would be teachers' focus on reduced vowels and weak forms in the classroom in order to improve their students' speech rhythm (Low 2006, 121). The results of the present study indicate that native listeners attend to certain combinations of acoustic features so it may be worthwhile to direct learners' attention to the several different ways to signal prominence, not only duration, intensity, pitch, or vowel quality. High motivation, personal interest in the language, and a positive attitude towards it may result in better pronunciation without additional instruction. Some of these strategies are employed in *Spoken English* but it is unclear to what extent and how explicitly.

In general, it was quite surprising how low scores the learners received for their speech rhythm, accentedness, and fluency, even after the course. Thomson and Derwing's (2015) review of pronunciation efficacy studies reveals that varying results of the effect of instruction can be due to e.g., individual differences, goals and duration of the instruction, and assessment procedures. As mentioned earlier, the course focused mostly on segmentals, and prosody had a secondary role in the assessment. The duration and the focus of the course should be critically evaluated: is one semester enough to improve the rhythm and fluency of the students permanently? The total amount of prosodic instruction may not be sufficient or its nature explicit enough. Since it is likely that the testing situation affects the pronunciation in the final test of the pronunciation course, like discussed above, I suggest employing formative assessment during the course instead. Students could record their speech at home after each set of practice and thus reduce anxiety if there were more chances to show the possible gains. If a summative test was still required, it could be a spontaneous conversation with the teacher or in groups. It is true that this type of test would allow learners to avoid certain difficult sounds in their speech, but it is one of the language learning strategies and can be done in real-life situations, too. Regardless, during the course, the learners would have been made aware of which phonemes or prosodic features are challenging and require additional attention.

5.2 Evaluation of the present study

When discussing the results of any study, reliability and validity should be considered. I have attempted to take these cornerstones of scientific research into account in every decision made during the research process, from choosing references to methodology and analysis. Of course, often times it is a balancing act between what is ideal and what realistic. The methodology of the

present study was carefully designed based on previous studies, such as Munro and Derwing (1999), Derwing et al. (2004), and Kennedy and Trofimovich (2010) but also restricted by the data available. In my opinion, the quantitative and qualitative approaches complemented each other and were justified. The diversity and number of native-speaker evaluators were satisfactory considering the resources of the present study.

However, there are several issues that could compromise reliability and validity in the present study. Firstly, the background information of the learners was unavailable. Individual differences in the learner scores were large and could be explained by factors such as bilingual background or longer residence in an English-speaking country. On the other hand, my experiment measured the development of the learners, so the focus was on the differences between pre- and posttest productions rather than between individual learners. Secondly, the lack of control regarding the native-speaker evaluators and the rating situation decreases the value of the results. Ideally, the rating session would have been organized in person so that it could have been confirmed that the raters followed the instructions. In addition, more native speakers as a control group would have resulted in better generalizability. Since the evaluators were able to choose to participate themselves, the results can be claimed to represent a certain type of people. As mentioned before, finding a sufficient number of native-speaker evaluators locally would have been difficult and required more time.

Thirdly, as the pre- and posttest texts were different, their comparison is problematic. The differences in the length of the samples (3 to 12 seconds) may be reflected in the rating scores as well: the shorter the production the easier it is to keep fluent. Yet, I would like to point out that he selection of the utterances included many variables and was far from straightforward. Moreover, Derwing and Munro (1997) used samples between 4.5 and 10.5 seconds and found that the duration did not affect the results. Finally, both texts were read aloud, which means that the results are not directly applicable to spontaneous speech, but as mentioned earlier, measuring read aloud speech is very common in rhythm and fluency studies, and thus a comparison to previous findings is possible. Based on these observations, I will provide suggestions for further research on this subject.

5.3 Suggestions for further research

The research on L2 fluency has been substantial in the past few decades but in my opinion, the nature of speech rhythm, the acquisition of L2 rhythm, and the importance of correct rhythm in L2 deserves more attention. First, a more comprehensive acoustic analysis encompassing the whole learner group would be required. The analysis would definitely benefit from a skilled use of Praat scripts and filters. Although the sample size of the present study was sufficient, a larger group would improve the generalizability. As for the data elicitation, I suggest using some other online questionnaire form for a better implementation of the audio samples or ideally, arranging the rating session under the guidance of the researcher. To better define the acoustic features influencing perceived rhythm and fluency, I propose collecting ratings from two different groups, one for rhythm and one for fluency. These raters would also verbalize what in the utterance made them think it was more or less rhythmic or fluent. To increase scientific validity, the research design could include a control group, which was not possible in the present study due to the use of pre-recorded data.

Classroom observation would be beneficial to see how prosodic features are taught in practice. Two or more English learner groups on different levels could be examined, possibly even recording the same text in the first and third year of studies. A longitudinal study on the same learners could show whether the course has long-term effects, similar to Derwing, Thomson, and Munro (2006). Of course, a longer period of time would also generate more explaining factors to consider (e.g., exposure to English outside studies). Some of these suggestions for further research are clearly quite demanding in terms of time, technology, and personnel but certainly feasible with proper resources. Next, the thesis will conclude with a short overview of the main points of the present study.

6 Conclusion

The present study aimed to reveal the influence of pronunciation instruction on the speech rhythm and fluency of advanced Finnish learners of English and the acoustic features which native English speakers attend to. I also wanted to see whether the ratings of the different aspects correlate with each other. I used a mixed methods approach combining listener ratings with an acoustic analysis on four extreme cases. The results indicate that on average, the students improved their speech rhythm, accentedness, and fluency but comprehensibility was reduced. The overall oral proficiency score increased slightly, but the changes between pre- and posttests were not significant on a group level. However, when examining the four most improved learners individually in an extreme case analysis, the differences in rhythm and fluency were significant ($p = 0.00$).

As for the acoustic measures of fluency, it was found that faster articulation rate, smaller number of unfilled pauses and their location at clause boundaries, as well as absence of repairs resulted in higher fluency scores for the four learners. There was no single rhythm measure that could be said to affect the ratings, but all four learners had matching amplitude and F0 peaks in either their posttest or both samples, and most of them used more durational cues, linking sounds, and vowel reduction in their posttest utterances. Each learner demonstrated a various set of these features to different extends. Regarding the relationships between the four aspects, all four of them correlated strongly with each other, particularly rhythm and fluency. This suggests that at least the tested aspects of pronunciation are tightly connected and affect one another. Based on the quantitative analysis as well as the evaluator comments, fluency has more impact on comprehensibility than accentedness. Most of the findings are in line with the previous research on rhythm, fluency, accentedness, and comprehensibility of L2 speech.

Several interpretations of the results were discussed in the present thesis. There were clearly individual differences between the learners, but contextual and external factors were considered having even more impact on the results. The most influential of these would be the differences in the pre- and posttest text genre and vocabulary as well as the posttest testing situation. The posttest was the final oral exam of the course and thus added a lot of pressure on the students. Based on the results, it was suggested that the assessment methods as well as the amount and nature of prosody instruction of the pronunciation course be critically evaluated to better cater to the needs of the students.

Nevertheless, the results should be interpreted with caution since some methodological details may have affected the scores: for instance, the length of the samples which were cut for the questionnaire varied from 3 to 12 seconds. Hence, it is recommended to ensure the similarity of the samples in further research. In the future, research on speech rhythm will need to focus more on combining quantitative and qualitative approaches, testing and comparing both L1 and L2 rhythm through native- and nonnative-listener ratings, and deepening our understanding of rhythm as a perceived as well as an acoustic phenomenon. Adding fluency ratings and measures into rhythm research would help define their relationship further. No matter how compelling the claim that rhythm only exists in the mind of the beholder (e.g., Molczanow and Wiese 2014), there is not enough evidence to abandon the acoustics just yet.

List of references

Adams, Corinne. 1979. *English Speech Rhythm and the Foreign Learner.* The Hague: De Gruyter Mouton.

Arvaniti, Amalia. 2012. "Rhythm Classes and Speech Perception". In *Understanding Prosody: The Role of Context, Function and Communication*, edited by Oliver Niebuhr, 75-92. De Gruyter. Accessed January 18 2019. https://www-degruyter-com.ezproxy.utu.fi/viewbooktoc/product/186201.

Arvaniti, Amalia. 2009. "Rhythm, Timing and the Timing of Rhythm". *Phonetica* 66, no. 1-2: 46–63. Accessed January 19 2019. https://doi.org/10.1159/000208930.

Barry, William J. 2008. "Rhythm as an L2 Problem: How Prosodic Is It?" In *Non-Native Prosody: Phonetic Description and Teaching Practice,* edited by Jürgen Trouvain and Ulrike Gut, 97-120. Berlin: De Gruyter Mouton. Accessed February 4 2019. https://doi.org/10.1515/9783110198751.1.97.

Celce-Murcia, Marianne, Donna M. Brinton, and Janet M. Goodwin. 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages.* Cambridge: Cambridge University Press.

Council of Europe. 2018. Qualitative Aspects of Spoken Language Use - Table 3 (CEFR 3.3): Common Reference Levels. Accessed March 3 2019. https://www.coe.int/en/web/common-european-framework-reference-languages/table-3-cefr-3.3-common-reference-levels-qualitative-aspects-of-spoken-language-use.

Couper-Kuhlen, Elizabeth. 1993. *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Pragmatics & Beyond New Series 25. Amsterdam: John Benjamins Publishing Company.

Crowther, Dustin, Pavel Trofimovich, Talia Isaacs, and Kazuya Saito. 2015. "Does a Speaking Task Affect Second Language Comprehensibility?" *The Modern Language Journal* 99, no. 1: 80-95. Accessed December 16 2018. https://doi.org/10.1111/modl.12185.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. 8th ed. Oxon: Routledge.

Cumming, Ruth E. 2010. *Speech Rhythm: The Language-Specific Integration of Pitch and Duration*. PhD dissertation, University of Cambridge. Accessed January 15 2019. https://doi.org/10.17863/CAM.16499.

Cutler, Anne. [1984] 2011. "Stress and Accent in Language Production and Understanding". In *Intonation, Accent and Rhythm: Studies in Discourse Phonology*, edited by Dafydd Gibbon and Helmut Richter, 77-90. De Gruyter. Accessed December 16 2018. https://doi.org/10.1515/9783110863239.

Dauer, Rebecca M. 1983. "Stress-Timing and Syllable-Timing Reanalyzed". *Journal of Phonetics* 11, no. 1: 51-62. Accessed January 20 2019. https://search-proquest-com.ezproxy.utu.fi/docview/58127558?accountid=14774.

De Jong, Nivja H. 2016. "Fluency in Second Language Assessment". In *Handbook of Second Language Assessment*, edited by Dina Tsagari and Jayanti Banerjee, 203-218. De Gruyter Mouton. Accessed February 9 2019. http://langsnap.soton.ac.uk/linked_files/LANGSNAP_dejong.pdf.

Derwing, Tracey M. and Murray J. Munro. 1997. "Accent, Intelligibility, and Comprehensibility: Evidence from Four L1s". *Studies in Second Language Acquisition* 19, no. 1: 1-16. Accessed November 20 2018. https://doi.org/10.1017/S0272263197001010.

———. 2013. "The Development of L2 Oral Language Skills in Two L1 Groups: A 7-Year Study". *Language Learning* 63, no. 2: 163-85. Accessed November 20 2018. https://onlinelibrary.wiley.com/doi/epdf/10.1111/lang.12000.

———. 2015. *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing Company.

Derwing, Tracey M., Murray J. Munro, and Grace Wiebe. 1998. "Evidence in Favor of a Broad Framework for Pronunciation Instruction". *Language Learning* 48, no. 3: 393–410. Accessed November 18 2018. https://onlinelibrary.wiley.com/doi/epdf/10.1111/0023-8333.00047.

Derwing, Tracey M., Murray J. Munro, Ron I. Thomson, and Marian J. Rossiter. 2009. "The Relationship Between L1 And L2 Fluency Development". *Studies in Second Language Acquisition* 31, no. 4: 533-57. Accessed November 18 2018. https://doi.org/10.1017/S0272263109990015.

Derwing, Tracey M., Marian J. Rossiter, Murray J. Munro, and Ron I. Thomson. 2004. "Second Language Fluency: Judgments on Different Tasks". *Language Learning* 54, no. 4: 655-79. Accessed November 20 2018. https://onlinelibrary-wiley-com.ezproxy.utu.fi/doi/epdf/10.1111/j.1467-9922.2004.00282.x.

Derwing, Tracey M., Ron I. Thomson, and Murray J. Munro. 2006. "English Pronunciation and Fluency Development in Mandarin and Slavic Speakers". *System* 34, no. 2: 183-193. Accessed November 20 2018. https://doi.org/10.1016/j.system.2006.01.005.

Deterding, David. 2012. "Issues in the Acoustic Measurement of Rhythm". In *Pragmatics and Prosody in English Language Teaching*, edited by Jesús Romero-Trillo, 9-24. Dordrecht: Springer. Accessed November 2 2018. https://doi.org/10.1007/978-94-007-3883-6.

Dilley, Laura C., Jessica Wallace, and Christopher C. Heffner. 2012. "Perceptual Isochrony and Fluency in Speech by Normal Talkers under Varying Task Demands". In *Understanding Prosody: The Role of Context, Function and Communication*, edited by Oliver Niebuhr, 237-258. Accessed December 16 2018. https://www.degruyter.com/downloadpdf/books/9783110301465/9783110301465.237/9783110301465.237.pdf.

Dörnyei, Zoltan. 2007. *Research Methods in Applied Linguistics: Quantitative, qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Eckman, Fred R. 1977. "Markedness and the Contrastive Analysis Hypothesis". *Language Learning* 27, no. 2: 315-30. Accessed January 18 2019. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x.

Eerola, Osmo, Janne Savela, Juha-Pertti Laaksonen, and Olli Aaltonen. 2012. "The Effect of Duration on Vowel Categorization and Perceptual Prototypes in a Quantity Language". *Journal of Phonetics* 40, no. 2: 315-28. Accessed February 2 2019. https://doi.org/10.1016/j.wocn.2011.12.003.

Espinosa, Gonzalo Eduardo. 2016. "On the Uncertain Naure of Speech Rhythm and Its Relevance in L2 Acquisition". *Argentinian Journal of Applied Linguistics* 4, no. 1: 5-22. Accessed January 17 2019. http://www.faapi.org.ar/ajal/issues/401/EspinosaAJALVol4(1).pdf.

Flege, James. 1981. “The Phonological Basis of Foreign Accent: A Hypothesis”. *TESOL Quarterly*, 15, no. 4: 443-55. Accessed January 20 2019. https://www.jstor.org/stable/3586485.

———. 1992. “Speech Learning in a Second Language”. In *Phonological Development: Models, Research, and Implications*, edited by Charles Ferguson, Lise Menn, and Carol Stoel-Gammon, 565-604. Timonium, MD: York Press. Accessed January 20 2019. http://www.jimflege.com/files/Flege_in_Ferguson_1992.pdf.

Grabe, Esther and Ee Ling Low. 2002. “Durational Variability in Speech and the Rhythm Class Hypothesis”. In *Laboratory Phonology 7*, edited by Carlos Gussenhoven and Natasha Warner, 515–46, [2008] 2013. De Gruyter Mouton. Accessed November 18 2018. https://www-degruyter-com.ezproxy.utu.fi/downloadpdf/books/9783110197105/9783110197105.2.515/9783110197105.2.515.pdf.

Hackman, Dorothea. 1978. “Rhythm in Finnish and English”. In *Nordic Prosody: Papers from a symposium*, edited by Eva Gårding, Gösta Bruce, and Robert Bannert, 263-269. Department of Linguistics, Lund University, Malmö. Accessed November 20 2018. https://journals.lub.lu.se/LWPL/article/view/17045/15423.

Hahn, Laura D. 2004. “Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals”. *TESOL Quarterly* 38, no. 2: 201-23. Accessed December 20 2018. https://doi.org/10.2307/3588378.

Hirose, Keikichi and Hiromichi Kawanami. 2002. “Temporal Rate Change of Dialogue Speech in Prosodic Units as Compared to Read Speech”. *Speech Communication* 36, no. 1-2: 97-111. Accessed February 7 2019. https://doi.org/10.1016/S0167-6393(01)00028-0.

Hirvonen, Pekka. 1973. *Measurement of Achievement in English at the End of Upper Secondary School: 4. Speaking Test*. Turku: AFinLA Publications 8.

Howell, Peter and Karimai Kadi-Hanifi. 1991. “Comparison of Prosodic Properties Between Read and Spontaneous Speech Material”. *Speech Communication* 10, no. 2: 163–9. Accessed February 7 2019. https://doi.org/10.1016/0167-6393(91)90039-V.

Isaacs, Talia and Ron I. Thomson. 2013. “Rater Experience, Rating Scale Length, and Judgments of L2 Pronunciation: Revisiting Research Conventions”. *Language Assessment Quarterly*

10, no. 2: 135-59. Accessed November 20 2018. https://doi.org/10.1080/15434303.2013.769545.

Jenkins, Jennifer. 2005. *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Kainada, Evia and Angelos Lengeris. 2015. "Native Language Influences on the Production of Second-Language Prosody". *Journal of the International Phonetic Association* 45, no. 3: 269-287. Accessed February 20 2019. https://doi.org/10.1017/S0025100315000158.

Kang, Okim. 2010. "Relative Salience of Suprasegmental Features on Judgments of L2 Comprehensibility and Accentedness". *System* 38, no. 2 :301-15. Accessed December 5 2018. https://doi.org/10.1016/j.system.2010.01.005.

Kennedy, Sara and Pavel Trofimovich. 2010. "Language Awareness and Second Language Pronunciation: A Classroom Study". *Language Awareness* 19, no. 3: 171-185. Accessed January 5 2019. https://doi.org/10.1080/09658416.2010.486439.

Kim, Taesung. 2008. "Accentedness, Comprehensibility, Intelligibility, and Interpretability of NNESTs". *The CATESOL Journal* 20, no. 1: 7-26. Accessed February 20 2019. http://www.catesoljournal.org/wp-content/uploads/2014/06/CJ20_kim.pdf.

Kohler, K. J. 2008. "The Perception of Prominence Patterns". *Phonetica* 65: 257-69. Accessed January 5 2019. https://doi.org/10.1159/000192795.

———. 2009. "Rhythm in Speech and Language: A New Research Paradigm". *Phonetica* 66: 29-45. Accessed January 5 2019. https://doi.org/10.1159/000208929.

Kormos, Judit and Mariann Denes. 2004. "Exploring Measures and Perceptions of Fluency in the Speech of Second Language Learners". *System* 32, no. 2: 145-64. Accessed January 7 2019. https://www-sciencedirect-com.ezproxy.utu.fi/science/article/pii/S0346251X0400017X?via%3Dihub.

Lehtonen, Jaakko. 1979. "Speech Rate and Pauses in the English of Finns, Swedish-speaking Finns, and Swedes". In *The Kuopio-Vaasa-Stockholm Tests. Perception and Production of English: Papers on Interlanguage*, edited by Rolf Palmberg, Håkan Ringbom, and Jaakko Lehtonen, 35-52. Turku: Åbo Akademi.

Lehtonen, Jaakko, Kari Sajavaara, Anthony May. 1977. *Spoken English: The Perception and Production of English on a Finnish-English Contrastive Basis.* Jyväskylä: Gummerus.

Levis, John M. 2005. “Changing Contexts and Shifting Paradigms in Pronunciation Teaching”. *TESOL Quarterly* 39, no. 3: 369-77. Accessed January 5 2019. http://www.personal.psu.edu/kej1/APLNG_493/old_site/levin.pdf.

———. 2016. “Research into practice: How Research Appears in Pronunciation Teaching Materials”. *Language Teaching* 49, no. 3: 423–37. Accessed January 6 2019. https://doi.org/10.1017/S0261444816000045.

Levis, John M., Sinem Sonsaat, Stephanie Link, and Taylor Anne Barriuso. 2016. “Native and Nonnative Teachers of L2 Pronunciation: Effects on Learner Performance”. *TESOL Quarterly* 50, no. 4: 894-931. Accessed January 5 2019. https://onlinelibrary.wiley.com/doi/epdf/10.1002/tesq.272.

Li, Aike and Brechtje Post. 2014. “L2 Acquisition of Prosodic Properties of Speech Rhythm: Evidence from L1 Mandarin and German Learners of English”. *Studies in Second Language Acquisition*, 36, no. 2: 223-55. Accessed January 5 2019. https://doi.org/10.1017/S0272263113000752.

Lintunen, Pekka. 2014. ”Ääntämisen oppiminen ja opettaminen” [Learning and Teaching Pronunciation]. In *Kuinka Kieltä Opitaan* [How Languages Are Learnt], edited by Päivi Pietilä and Pekka Lintunen, 165-187. Helsinki: Gaudeamus.

Low, Ee Ling. 2006. “A Review of Recent Research on Speech Rhythm: Some Insights for Language Acquisition, Language Disorders and Language Teaching”. In *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice,* edited by Rebecca Hughes, 99-125. New York: Palgrave Macmillan. Accessed December 6 2018. https://link-springer-com.ezproxy.utu.fi/content/pdf/10.1057%2F9780230584587.pdf.

Major, Roy. 2007. “Identifying a Foreign Accent in an Unfamiliar Language”. *Studies in Second Language Acquisition* 29, no. 4: 539-56. Accessed December 8 2018. https://doi-org.ezproxy.utu.fi/10.1017/S0272263107070428.

———. 2010. “First Language Attrition in Foreign Accent Perception”. *International Journal of Bilingualism* 14, no. 2: 163–83. Accessed December 9 2018. http://journals.sagepub.com/doi/pdf/10.1177/1367006910363063.

Major, Roy C., Susan F. Fitzmaurice, Ferenc Bunta, and Chandrika Balasubramanian. 2002. “The Effects of Nonnative Accents on Listening Comprehension: Implications for ESL

Assessment". *TESOL Quarterly* 36, no. 2: 173-90. Accessed December 8 2018. https://www.jstor.org/stable/3588329.

Molczanow, Janina and Richard Wiese. 2014. "Rhythm Is in the Mind of the Beholder. Remarks on the Nature of Linguistic Rhythm". *Linguistica Copernicana* 11: 169-82. Accessed February 20 2019. http://dx.doi.org/10.12775/LinCop.2014.044.

Morris-Wilson, Ian. 1981. *An Introduction to English Segmental Phonetics for Finns*. Helsinki: Gaudeamus.

Moyer, Alene. 2013. *Foreign Accent: The Phenomenon of Non-Native Speech*. Cambridge: Cambridge University Press.

Munro, Murray J. and Tracey M. Derwing. 1994. "Evaluations of Foreign Accent in Extemporaneous and Read material". *Language Testing* 11, no. 3: 253-66. Accessed November 21 2018. https://doi.org/10.1177/026553229401100302.

———. 1999. "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners". *Language Learning* 49, no. 1: 285-310. Accessed November 20 2018. https://search-proquest-com.ezproxy.utu.fi/docview/85692911?pq-origsite=primo.

Nieminen, Tommi and Michael O'Dell. 2009. Suomen puherytmi typologisessa katsannossa. 36. Kielitieteen päivät Jyväskylässä. Accessed March 20 2019. http://papers.legisign.org/presentations/nieminen-odell2009.pdf.

Paananen-Porkka, Minna M. 2007. "Speech Rhythm in an Interlanguage Perspective: Finnish Adolescents Speaking English". PhD dissertation, University of Helsinki.

Peltola, Maija, Pekka Lintunen and Henna Tamminen. 2014. "Advanced English Learners Benefit from Explicit Pronunciation Teaching: An Experiment with Vowel Duration and Quality". *AFinLA-E* 6: 86-98. Accessed January 16 2019. https://journal.fi/afinla/article/view/46282.

Peltonen, Pauliina and Pekka Lintunen. 2016. "Integrating Quantitative and Qualitative Approaches in L2 Fluency Analysis: A Study of Finnish-Speaking and Swedish-Speaking Learners of English at Two School Levels". *European Journal of Applied Linguistics* 4, no. 2: 209-38. Accessed January 15 2019. https://search-proquest-com.ezproxy.utu.fi/docview/1818037555?pq-origsite=primo.

Pihko, Marja-Kaisa. 1994. *Cross-linguistic Intelligibility of Native and Non-Native L2 Speech Varieties*. Research Reports from the Language Centre for Finnish Universities 1. Jyväskylä: University of Jyväskylä.

———. 1997. "'His English Sounded Strange': The Intelligibility of Native and Non-Native English Pronunciation to Finnish Learners of English". PhD dissertation, University of Jyväskylä.

Purcell, Edward T. and Richard W. Suter. 1980. "Predictors of Pronunciation Accuracy: A Reexamination". *Language Learning* 30, no. 2: 271-87. Accessed January 15 2019. https://doi.org/10.1111/j.1467-1770.1980.tb00319.x.

Ramus, F., M. Nespor and J. Mehler. 1999. "Correlates of Linguistic Rhythm in the Speech Signal". *Cognition*, 73: 265–92. Accessed December 20 2018. https://doi.org/10.1016/S0010-0277(99)00058-X.

Rathcke, Tamara V. and Rachel H. Smith. 2015. "Speech Timing and Linguistic Rhythm: On the Acoustic Bases of Rhythm Typologies". *The Journal of the Acoustical Society of America* 137, no. 5: 2834. Accessed December 15 2018. https://doi.org/10.1121/1.4919322.

Riazantseva, Anastasia. 2001. "Second Language Proficiency and Pausing: A Study of Russian Speakers of English". *Studies in Second Language Acquisition* 23, no. 4 :497-526. Accessed February 15 2019. https://www-jstor-org.ezproxy.utu.fi/stable/44486959.

Rossiter, Marian J. 2009. "Perceptions of L2 Fluency by Native and Non-native Speakers of English". *The Canadian Modern Language Review* 65, no. 3: 395 –412. Accessed December 8 2018. https://doi.org/10.3138/cmlr.65.3.395.

Saito, Kazuya, Pavel Trofimovich, and Talia Isaacs. 2017. "Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness: A Validation and Generalization Study". *Applied Linguistics* 38, no. 4: 439–62. Accessed November 20 2018. https://doi.org/10.1093/applin/amv047.

Segalowitz, Norman. 2007. "Access Fluidity, Attention Control, and the Acquisition of Fluency in a Second Language". *TESOL Quarterly* 41, no. 1: 181-6. Accessed February 15 2019. https://doi-org.ezproxy.utu.fi/10.1002/j.1545-7249.2007.tb00047.x.

———. 2010. *Cognitive Bases of Second Language Fluency*. New York: Routledge.

Segalowitz, Norman and Barbara Freed. 2004. "Context, Contact, and Cognition in Oral Fluency Acquisition: Learning Spanish in at Home and Study Abroad Contexts". *Studies in Second Language Acquisition* 26, no. 2: 173-99. Accessed February 15 2019. https://doi-org.ezproxy.utu.fi/10.1017/S0272263104262027.

Tavakoli, Parvaneh, Colin Campbell, and Joan McCormack. 2016. "Development of Speech Fluency Over a Short Period of Time: Effects of Pedagogic Intervention". *TESOL Quarterly* 50, no. 2: 447-71. Accessed December 8 2018. https://doi-org.ezproxy.utu.fi/10.1002/tesq.244.

Tavakoli, Parvaneh and Peter Skehan. 2005. "Strategic Planning, Task Structure, and Performance Testing". In *Planning and Task Performance in a Second Language*, edited by Rod Ellis, 239–273. John Benjamins Publishing Company. Accessed February 13 2019. https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=622981&ppg=247.

Tergujeff, Elina. 2014. "Three Claims in English Pronunciation Instruction in Finland". *Puhe ja Kieli* 34, no. 1: 43–47. Accessed January 6 2019. https://journal.fi/pk/article/view/45206/11474.

Thomson, Ron I. and Tracey M. Derwing. 2015. The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics* 36, no. 3: 326-344. Accessed January 16 2019. https://doi.org/10.1093/applin/amu076.

Toivola, Minnaleena., Mietta Lennes, and Eija Aho. 2009. "Speech Rate and Pauses in Non-Native Finnish Speech". Proceedings of the 10th Interspeech 2009 Conference, Brighton, 1707-10. Accessed January 20 2019. https://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_1707.pdf.

Tominaga, Yuko. 2011. "An Analysis of English Pronunciation of Japanese Learners: From the Viewpoint of EIL". *Journal of Pan-Pacific Association of Applied Linguistics* 15, no. 2: 45-57. Accessed December 20 2018. https://eric.ed.gov/?id=EJ97991.

Trofimovich, Pavel and Wendy Baker. 2006. "Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech". *Studies in Second Language Acquisition* 28, no. 1: 1-30. Accessed November 15 2018. https://doi.org/10.1017/S0272263106060013.

Trofimovich, Pavel and Talia Isaacs. 2012. “Disentangling Accent from Comprehensibility”. *Bilingualism: Language and Cognition* 15, no. 4: 905-16. Accessed December 5 2018. https://doi.org/10.1017/S1366728912000168.

Ullakonoja, Riikka. 2011. “Da. Eto vopros!: Prosodic Development of Finnish Students’ Read-aloud Russian during Study in Russia”. PhD dissertation, University of Jyväskylä.

Wennerstrom, Ann K. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis* Oxford: Oxford University Press. Accessed November 15 2018. https://login.ezproxy.utu.fi/login?url=http://site.ebrary.com/lib/uniturku/Doc?id=10084857.

Witton-Davies, Giles. 2015. “Measuring Oral Fluency: How It Can Be Done, and the Implications for Teaching and Testing”. The Proceedings of 2015 International Conference and Workshop on TEFL & Applied Linguistics: 311-21. Department of Applied English, Ming Chuan University. Accessed February 14 2019. http://www.forex.ntu.edu.tw/en/download.php?filename=4463_064ede15.pdf&dir=writing&title=witton-davies+15a+MCU.

Ylinen, Sari, Maria Uther, Antti Latvala, and Sara Vepsäläinen. 2010. “Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English”. *Journal of Cognitive Neuroscience* 22, no. 6: 1319-32. Accessed January 15 2019. https://doi.org/10.1162/jocn.2009.21272.

Appendices

Appendix 1 Pretest text [selected phrases in italics]

Read aloud the following text and the individual words below. Remember to say your name at first.

There was once a poor shepherd boy who used to watch his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting "Wolf, Wolf." *As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety*, and two of his cousins even stayed with him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. *However, not long after*, *a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck*. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before. Unfortunately, as all the villagers were convinced that he was trying to fool them a third time, they told him, "Go away and don't bother us again." And so the wolf had a feast.

Appendix 2 Posttest text [selected phrases in italics]

Western governments are desperate to restore the global economy along "business as usual" lines. But, argues Andrew Sims, that is a short-sighted approach; a radical redevelopment would bring much bigger environmental, social and economic benefits.

If someone offered you a plan that would get rich countries on to a radical path of deep, immediate carbon cuts to tackle climate change and also solved a vast range of social problems, would you take it?

A team of scientists and economists at the New Economics Foundation has come up with a plan called The Great Transition. This provides a blueprint - or rather, a greenprint - for how the UK can make a step-change in delivering quality of life for all, whilst living within our collective environmental means. One thing is sure: in spite of being well-intended, the recent Climate Change Committee call for more electric cars and nuclear power was a disastrously inadequate distraction from the scale and speed of the measures actually needed.

At one extreme, allowing runaway climate change to occur is infinitely expensive and therefore unthinkable. But even over the next few decades, simply by factoring-in reasonable, even highly conservative, estimates of how much we can save by tackling social and environmental problems with proven solutions, the results are astonishing.

All change is threatening. But failures of imagination and bad decisions can also be fatal. When Greenland was occupied by Icelandic and Scandinavian settlers in the early Middle Ages, *they soon made themselves at home with customs and methods of food cultivation.* When the great chill bit deep in the 15th Century, instead of adapting by learning from the climate-adjusted indigenous people, they clung to what they knew, and died out. Far from that grim scenario, today the great re-skilling of society to manage this transition could even break the zombie walk of consumer society and bring us alive again as individuals and communities. *The Great Transition is a tale of how it turned out right.*

Appendix 3 Posttest alternative text [selected phrases in italics]

Military commanders know they have to make quick decisions: lives can depend on their conclusions. But what is decision-making like for politicians, far removed from the battlefield, when it comes to military matters?
It is now several months since President Obama received a report from General Stanley McChrystal, the commander of Nato forces, outlining a revised military strategy. The president has still to decide if he should support General's call for many more troops.
Some say he is dithering. Others say he is carefully weighing up advice given to him from many quarters.

Balancing the advice with the need for a decision is an experience familiar to John Hutton and Malcolm Rifkind, both former British defence secretaries.
"If you have the right information and advisers who can give some background, your job as a minister is not to become an expert on the subject but to understand the options and to decide which to choose. "*If I had the opportunity to do that properly, then I could live with the decisions"*, Rifkind added.

Hutton told Broadcasting House that he was struck by the professionalism when it came to giving ministers "clear and concise advice." He admitted he was not afraid to challenge that advice where necessary, even if that carried dangers. "There has to be a clear recognition of responsibilities. Politicians who become armchair generals run a very serious risk of getting things horribly wrong,"

Historically, politicians have taken a varied approach to decisions, with varied levels of success. For Anthony Howard Lady Thatcher was decisive, some would say to the point of impulsive: "The decision to send the taskforce to the Falklands Islands in 1982 was made on the very night that she was told it was possible.

Mr Howard identified similarities between Tony Blair and Lady Thatcher in terms of their decision-making styles.

"He was also impulsive. He believed in 'sofa government', not big cabinet meetings. He certainly goes into the category of those who liked making snap decisions."

Tony Blair's immediate predecessor had a very different way of choosing what to do, leading to accusations of dithering.
Anthony Howard believed President Obama's reputation had similarly been "badly damaged" by his failure to make a decision over future strategy.

Mr Blakeman had a warning for a president receiving advice from many quarters over Afghanistan: "Indecision can be as perilous as making the wrong decision."
Malcolm Rifkind, in contrast, was in less of a hurry:"I would rather President Obama takes eight weeks and gets it right than takes eight days and gets it wrong."

Appendix 4 Questionnaire introduction

M.A. Thesis Questionnaire of English Speech Rhythm and Fluency

This questionnaire is targeted at native speakers of English. The purpose of this questionnaire is to collect data for my master's thesis for the Department of English at the University of Turku (Finland). I attempt to examine the correlation between the auditory ratings and the acoustic measures of speech rhythm and fluency. Although, the main focus is on rhythm and fluency, ratings for accentedness and comprehensibility are included as well.

This questionnaire consists of background information questions and 42 short recordings in four sets. You should be able to complete the questionnaire in about 30 minutes. Before you start, please make sure that you are in a quiet place, you are wearing headphones, and that you are not doing anything else at the same time. It is recommended that you take short breaks between the sets. You are allowed to listen to each recording as many times as you wish.

Keep in mind that there are no right or wrong answers; I am interested in your genuine reactions and opinions. The questionnaire is completely anonymous. Please note that this questionnaire has not been formatted for mobile devices. You will hear recordings of different English phrases read aloud by speakers 1 to 10. Click on each link to open the recording, listen to it, and rate it by choosing the number that best represents your perception.

Appendix 5 Correlations between pretest and posttest aspects

		pretest rhythm	pretest accentedness	pretest fluency	pretest comprehensibility	posttest rhythm	posttest accentedness	posttest fluency	posttest comprehensibility
pretest rhythm	Pearson Correlation	1	.851**	.978**	.878**	.397	.513*	.405	.346
	Sig. (2-tailed)		**.000**	**.000**	**.000**	.083	.021	.076	.135
	N	20	20	20	20	20	20	20	20
pretest accentedness	Pearson Correlation		1	.859**	.814**	.616**	.816**	.631**	.572**
	Sig. (2-tailed)			**.000**	**.000**	**.004**	**.000**	**.003**	**.008**
	N		20	20	20	20	20	20	20
pretest fluency	Pearson Correlation			1	.908**	.422	.543*	.428	.387
	Sig. (2-tailed)				**.000**	.064	.013	.060	.092
	N			20	20	20	20	20	20
pretest comprehensibility	Pearson Correlation				1	.502*	.637**	.528*	.584**
	Sig. (2-tailed)					.024	**.003**	.017	**.007**
	N				20	20	20	20	20
posttest rhythm	Pearson Correlation					1	.853**	.981**	.870**
	Sig. (2-tailed)						**.000**	**.000**	**.000**
	N					20	20	20	20
posttest accentedness	Pearson Correlation						1	.868**	.853**
	Sig. (2-tailed)							**.000**	**.000**
	N						20	20	20
posttest fluency	Pearson Correlation							1	.910**
	Sig. (2-tailed)								**.000**
	N							20	20
posttest comprehensibility	Pearson Correlation								1
	Sig. (2-tailed)								
	N								20

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix 6 Reliability statistics

Rhythm

Cronbach's Alpha	N of Items
.953	42

Accentedness

Cronbach's Alpha	N of Items
.950	42

Fluency

Cronbach's Alpha	N of Items
.962	42

Comprehensibility

Cronbach's Alpha	N of Items
.956	42

Appendix 7 Cronbach's alpha if item deleted for each aspect

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
rhythm1	218.52	2230.99	.525	.952
rhythm2	217.03	2217.03	.597	.951
rhythm3	216.32	2276.63	.259	.953
rhythm4	219.48	2223.26	.580	.951
rhythm5	216.29	2270.61	.405	.952
rhythm6	221.06	2235.13	.551	.951
rhythm7	215.55	2323.52	.030	.953
rhythm8	219.45	2192.32	.681	.951
rhythm9	219.16	2170.61	.717	.950
rhythm10	218.74	2178.80	.622	.951
rhythm11	217.84	2249.27	.433	.952
rhythm12	218.16	2250.14	.479	.952
rhythm13	218.58	2214.85	.594	.951
rhythm14	220.61	2236.05	.497	.952
rhythm15	217.48	2206.99	.545	.952
rhythm16	219.77	2185.85	.673	.951
rhythm17	220.97	2198.77	.679	.951
rhythm18	218.61	2171.25	.699	.950
rhythm19	220.29	2228.48	.493	.952
rhythm20	217.32	2215.76	.410	.953
rhythm21	216.68	2230.49	.513	.952
rhythm22	216.45	2242.86	.528	.952
rhythm23	216.13	2244.38	.524	.952
rhythm24	218.48	2206.13	.546	.952
rhythm25	217.84	2220.07	.479	.952
rhythm26	216.42	2236.19	.517	.952
rhythm27	216.55	2236.19	.557	.951
rhythm28	219.23	2203.91	.623	.951
rhythm29	220.45	2164.12	.711	.950
rhythm30	216.35	2197.44	.650	.951
rhythm31	220.94	2267.80	.340	.953
rhythm32	217.42	2239.45	.453	.952
rhythm33	219.55	2196.52	.769	.950

rhythm34	217.68	2168.43	.737	.950
rhythm35	219.81	2202.36	.717	.951
rhythm36	218.84	2236.07	.383	.953
rhythm37	219.94	2181.73	.657	.951
rhythm38	217.29	2218.81	.671	.951
rhythm39	217.26	2233.40	.529	.952
rhythm40	220.29	2208.08	.624	.951
rhythm41	216.84	2238.21	.588	.951
rhythm42	219.10	2189.96	.655	.951

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
accent1	184.32	1970.29	.307	.951
accent2	182.87	1939.25	.437	.950
accent3	180.32	2016.49	.062	.952
accent4	183.81	1930.23	.549	.949
accent5	181.13	1968.25	.369	.950
accent6	186.03	1987.43	.423	.950
accent7	179.77	2018.45	.072	.952
accent8	184.58	1929.39	.578	.949
accent9	183.23	1904.05	.652	.948
accent10	185.52	1963.39	.611	.949
accent11	183.42	1947.79	.556	.949
accent12	183.39	1924.05	.657	.949
accent13	184.16	1916.74	.658	.948
accent14	183.77	1916.78	.674	.948
accent15	183.48	1898.06	.699	.948
accent16	184.35	1931.30	.622	.949
accent17	185.03	1945.23	.758	.948
accent18	183.65	1926.64	.572	.949
accent19	185.00	1933.80	.670	.949
accent20	180.00	1995.67	.237	.951
accent21	181.61	1894.45	.689	.948
accent22	182.23	1929.58	.534	.949
accent23	180.94	1937.53	.458	.950
accent24	182.06	1893.33	.645	.949
accent25	180.84	1973.07	.295	.951

accent26	181.19	1952.50	.480	.950
accent27	182.61	1887.91	.685	.948
accent28	183.97	1928.97	.550	.949
accent29	185.10	1933.62	.550	.949
accent30	181.16	1954.47	.363	.951
accent31	184.58	1915.92	.812	.948
accent32	183.23	1915.85	.663	.948
accent33	184.32	1931.96	.681	.949
accent34	181.84	1904.74	.652	.948
accent35	184.84	1963.14	.488	.950
accent36	184.87	1928.18	.691	.948
accent37	184.68	1939.89	.647	.949
accent38	182.97	1926.77	.583	.949
accent39	183.13	1922.05	.651	.949
accent40	184.52	1942.73	.772	.948
accent41	182.58	1907.52	.626	.949
accent42	183.68	1921.69	.633	.949

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
fluency1	228.68	2626.89	.572	.961
fluency2	227.58	2619.52	.591	.961
fluency3	226.97	2683.90	.233	.963
fluency4	230.03	2596.83	.698	.961
fluency5	226.84	2656.61	.573	.961
fluency6	231.71	2666.21	.334	.962
fluency7	226.19	2732.63	.066	.963
fluency8	230.16	2558.61	.734	.960
fluency9	229.94	2540.20	.810	.960
fluency10	229.55	2619.12	.547	.961
fluency11	228.32	2656.36	.418	.962
fluency12	228.42	2599.39	.679	.961
fluency13	228.97	2588.43	.708	.961
fluency14	230.68	2612.36	.593	.961
fluency15	227.90	2598.36	.649	.961
fluency16	229.87	2549.85	.824	.960
fluency17	231.06	2603.13	.713	.961

fluency18	229.06	2573.40	.760	.960
fluency19	229.74	2600.40	.683	.961
fluency20	227.58	2612.39	.456	.962
fluency21	226.84	2642.14	.631	.961
fluency22	226.87	2653.98	.582	.961
fluency23	226.35	2686.70	.425	.962
fluency24	228.03	2592.97	.652	.961
fluency25	227.68	2602.83	.629	.961
fluency26	226.52	2648.93	.631	.961
fluency27	227.16	2626.01	.597	.961
fluency28	229.87	2601.12	.630	.961
fluency29	231.23	2544.58	.788	.960
fluency30	226.90	2646.22	.462	.962
fluency31	231.19	2623.03	.641	.961
fluency32	227.77	2669.98	.399	.962
fluency33	229.84	2574.01	.838	.960
fluency34	228.65	2539.70	.839	.960
fluency35	230.06	2594.13	.695	.961
fluency36	229.19	2662.10	.390	.962
fluency37	230.39	2610.11	.592	.961
fluency38	227.81	2627.63	.615	.961
fluency39	227.61	2653.05	.467	.962
fluency40	230.58	2568.19	.817	.960
fluency41	227.42	2647.12	.555	.961
fluency42	230.00	2595.20	.688	.961

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
comprehensibility1	268.71	2078.88	.341	.956
comprehensibility2	268.52	2089.13	.253	.956
comprehensibility3	269.03	2080.00	.214	.957
comprehensibility4	269.58	2043.05	.568	.955
comprehensibility5	268.13	2100.32	.260	.956
comprehensibility6	273.71	2125.88	.024	.958
comprehensibility7	268.13	2140.25	-.071	.958
comprehensibility8	269.65	1983.30	.742	.954
comprehensibility9	269.68	1972.23	.833	.953
comprehensibility10	271.48	2027.66	.480	.955

comprehensibility11	269.03	2079.57	.442	.955
comprehensibility12	269.06	2031.73	.695	.954
comprehensibility13	270.16	2023.67	.539	.955
comprehensibility14	270.19	1969.36	.752	.953
comprehensibility15	269.00	2032.53	.601	.955
comprehensibility16	269.74	1992.40	.812	.953
comprehensibility17	271.42	1972.92	.771	.953
comprehensibility18	270.06	1967.60	.760	.953
comprehensibility19	269.81	2003.50	.639	.954
comprehensibility20	268.68	2036.76	.481	.955
comprehensibility21	267.81	2088.96	.438	.955
comprehensibility22	268.00	2081.27	.574	.955
comprehensibility23	268.23	2095.18	.269	.956
comprehensibility24	268.74	2052.60	.567	.955
comprehensibility25	268.23	2046.45	.669	.954
comprehensibility26	267.90	2089.69	.493	.955
comprehensibility27	268.58	2060.56	.475	.955
comprehensibility28	269.97	1993.23	.749	.954
comprehensibility29	272.16	1971.34	.759	.953
comprehensibility30	268.61	2041.91	.583	.955
comprehensibility31	271.23	1964.85	.804	.953
comprehensibility32	268.29	2090.68	.368	.956
comprehensibility33	270.23	1986.11	.821	.953
comprehensibility34	269.23	2027.78	.652	.954
comprehensibility35	270.65	2008.77	.738	.954
comprehensibility36	270.52	2051.66	.422	.956
comprehensibility37	270.23	2002.98	.651	.954
comprehensibility38	268.65	2044.50	.654	.954
comprehensibility39	268.58	2065.32	.550	.955
comprehensibility40	270.39	1938.71	.902	.952
comprehensibility41	268.65	2028.17	.645	.954
comprehensibility42	270.03	1968.03	.879	.953

Finnish summary

Johdanto ja teoriatausta

Tämä pro gradu -tutkielma käsitteli englannin kielen puherytmiä ja sujuvuutta suomalaisten edistyneiden englanninopiskelijoiden puheessa. Tutkimuksen tavoitteena oli selvittää, muuttuvatko nämä ominaisuudet ääntämiskurssin tuloksena, kun kurssia ennen ja sen jälkeen äänitettyjä tuotoksia arvioivat englannin natiivipuhujat. Valitsin arviointien perusteella neljä opiskelijaa, joiden puheessa tapahtui suurin muutos ja vertailin akustisia ominaisuuksia puhesignaalissa löytääkseni ne, joihin natiivipuhujat kiinnittivät huomiota.

Puherytmiä on tutkittu paljon erityisesti englannin kielessä jo 1900-luvun alkupuolelta lähtien. Tämä aiempi tutkimus keskittyi kuitenkin merkittävissä määrin ajoitteisuusdikotomiaan. Tämän näkemyksen mukaan kielet ovat joko paino- tai tavuajoitteisia. Esimerkiksi suomen kieltä on pidetty yleisesti tavuajoitteisena, mutta väitteelle ei ole löytynyt perusteluja (Nieminen & O'Dell 2009). Yhden tietyn rytmityypin määrittäminen olisi hankalaa, sillä tavujen pituus muuttaa ajoitteisuutta. 1990-luvulla rytmin tutkimuksessa tapahtui paradigman muutos, ja keskiöön nostettiin havaittu rytmisyys. Kahden rytmityypin ajoitteisuusdikotomiasta ja mitattavissa olevasta isokroniasta eli painojen säännöllisestä esiintymisestä on jo pitkälti luovuttu rytmitutkimuksessa. Puherytmiä on yritetty analysoida akustisilla mittareilla: esimerkiksi Ramus, Nespor ja Mehler (1999) ovat mitanneet rytmiä eri kielissä vokaalisten intervallien suhteella sekä vokaali- että konsonantti-intervallien keston keskihajontaan. Grabe ja Low (2002) puolestaan ovat käyttäneet vokaalisia ja vokaalienvälisiä kestoja hyödyntävää summalauseketta (*Pairwise Variability Index)*. Monet tutkijat kritisoivat tällaisia mittareita, sillä ne mittaavat ainoastaan ajoitteisuutta sivuuttaen muut puherytmin ominaisuudet, eivätkä auta selittävän mallin rakentamisessa (esim. Arvaniti 2009; Kohler 2009; Arvaniti 2012; Deterding 2012). Muun muassa Arvaniti (2009) peräänkuuluttaa tutkimuksen keskittämistä havaittuun rytmiin, akustisten mittareiden kielikohtaista soveltamista sekä natiivipuhujien kokemusten huomioimista. Derwin ja Munro (2015) sekä Paananen-Porkka (2007) muistuttavat, että akustisia mittauksia voidaan käyttää kuulija-arviointien tukena, muttei korvaamaan niitä. Vaikka akustisessa signaalissa pystyttäisiin näyttämään muutoksia, kuulijan kokemus tekee niistä merkityksellistä.

Puherytmi koostuu useista osatekijöistä: lausepainosta, junktuurista, puhenopeudesta ja tauotuksesta (Paananen-Porkka 2007). Omassa tutkimuksessani käsittelin lausepainon

korrelaatteja eli äänenvoimakkuutta, sävelkorkeutta, pituutta ja vokaalin laatua ensisijaisesti rytmin ominaisuuksina, kun taas junktuuri, puhenopeus ja tauot liittyivät sujuvuuden analysointiin. Englannin kielessä sanapaino kuuluu olennaisena osana fonologiseen tietoon jokaisesta sanasta. Lausepaino määräytyy paitsi lauseeseen sisältyvien sanojen painojen myös semanttisten ja pragmaattisten tekijöiden kautta. Toisinaan luonnollisen rytmin saavuttamiseksi painoja on lisättävä yleensä painottomille tavuille tai sanoille (esim. artikkelit, prepositiot ja apuverbit). Lisäksi spontaani puhe poikkeaa usein selvästi luetusta juuri säännöllisen rytmin kohdalla. Tällaisten ilmiöiden vuoksi englannin puherytmiin on vaikea määritellä tiukkoja sääntöjä.

Rytmin lisäksi keskeinen aspekti tutkimuksessani oli sujuvuus. Vieraan kielen sujuvuuden tutkimus on ollut runsasta viime vuosikymmeninä. Arkikielessä sujuvuudesta puhutaan usein laajassa merkityksessä eli kielen kokonaisvaltaisena hallintana, kun taas vieraan kielen oppimisen tutkimuksessa sitä tarkastellaan rajatun merkityksen kautta, suullisen (joskus myös kirjoitetun) kielitaidon ominaisuutena. Segalowitz (2010) jakaa sujuvuuden kolmeen ulottuvuuteen: tuotoksen sujuvuuteen, havaittuun sujuvuuteen ja kognitiiviseen sujuvuuteen. Tuotoksen sujuvuus jaetaan edelleen kolmeen (Tavakoli & Skehan 2005): nopeuteen, tauottamiseen ja korjaussujuvuuteen. Rytmin ja sujuvuuden suhdetta ei ole tutkittu kattavasti, mutta muun muassa Dilley, Wallace ja Heffner (2012), Tominaga (2011) ja Paananen-Porkka (2007) ovat löytäneet tutkimuksissaan merkittävän yhteyden näiden aspektien välillä.

Koska tutkimukseni käsitteli puherytmiä ja sujuvuutta nimenomaan vieraan kielen oppimisen näkökulmasta, oli tärkeää ottaa huomioon erilaiset seikat prosodian oppimisessa. Trofimovich ja Baker (2006) ehdottavat, että prosodian oppimiseen voisi soveltaa samoja teorioita ja malleja kuin yksittäisten äänteiden omaksumiseen. Samalla tavalla kuin segmentaalien oppimisessa näkyy äidinkielen siirtovaikutus, myös vieraan kielen prosodian kehityksen on osoitettu olevan epätasaista ja riippuvaista äidinkielestä (Li & Post 2014). Kainada ja Lengeris (2015) puolestaan huomasivat, että vaikka oppijat käyttivät äidinkielen prosodisia piirteitä kohdekielessä, puhe poikkesi myös äidinkielen normeista. Erityisesti suomenkielisille ongelmia englannin rytmissä tuottavat virheellinen ja riittämätön perustaajuuden vaihtelu, riittämätön vokaalireduktio painottomilla tavuilla, virheelliset segmenttien kestot sekä glottalisaation virheellinen käyttö junktuurin merkkinä (Hackman 1978; Lehtonen, Sajavaara, and May 1977; Lehtonen 1979; Morris-Wilson 1981; Pihko 1994; Paananen-Porkka 2007). Vieraan kielen oppijoiden aksenttisuutta ja ymmärrettävyyttä on perinteisesti tutkittu natiivikuuntelijoiden

arvioinneilla. Monissa tutkimuksissa on osoitettu, että aksenttisuus ei välttämättä vähennä ymmärrettävyyttä (esim. Derwing ym. 2004). Vieraan kielen ääntämisen opettamisessa onkin 1960-luvun jälkeen siirrytty korostamaan ymmärrettävyyttä natiivinkaltaisuuden sijaan (Levis 2005). Valitettavasti prosodian opetusta pidetään usein hankalana tai jopa turhana segmentaaleihin verrattuna ja se saatetaan siksi jättää kokonaan väliin. Prosodiaan keskittyvä ääntämisenopetus on kuitenkin tutkitusti parantanut oppijoiden ymmärrettävyyttä verrattuna pelkkään segmentaaliseen opetukseen (Derwing, Munro, ja Wiebe 1998).

Tämän teoreettisen viitekehyksen pohjalta tutkimuksen keskiöön nousivat seuraavat kolme tutkimuskysymystä: Ensiksi, tapahtuuko edistyneiden englanninoppijoiden puherytmissä ja sujuvuudessa muutoksia ääntämiskurssin jälkeen natiivikuuntelijoiden arviointien perusteella? Hypoteesini oli, että yksilötasolla muutoksia tapahtuu, mutta ryhmätasolla ne eivät ole merkitseviä. Toiseksi, jos muutoksia löytyy, minkä akustisten tekijöiden kanssa ne korreloivat? Arvelin, että yksilöiden muutoksia aiheuttavat puhenopeuden kasvu, vokaalien redusointi ja selkeämpi vaihtelu painollisten ja painottomien tavujen välillä. Kolmanneksi, korreloivatko arvioidut ominaisuudet (puherytmi, sujuvuus, aksenttisuus ja ymmärrettävyys) keskenään? Aiempien tutkimusten perusteella odotin vahvaa korrelointia etenkin rytmin ja sujuvuuden välillä. Seuraavassa osiossa esittelen, millaisin menetelmin pyrin vastaamaan asettamiini tutkimuskysymyksiin sekä millaisia koehenkilöitä ja aineistoa tutkimus sisälsi.

Aineisto ja menetelmät

Tutkimukseni hyödynsi sekä määrällisiä että laadullisia menetelmiä, joista kerron tarkemmin tässä osiossa. Koehenkilöinä tutkimuksessa toimi kaksi eri ryhmää: 20 ensimmäisen vuoden englannin kielen pääaineopiskelijaa sekä 31 arvioijaa, jotka olivat kaikki natiivienglanninpuhujia. Kaikki koehenkilöt osallistuivat tutkimukseen anonyymisti. Opiskelijat valittiin niiden 45 opiskelijan joukosta, jotka olivat antaneet luvan käyttää äänityksiään tutkimustarkoituksiin. Määrä karsittiin kahteenkymmeneen kahdesta syystä: ensiksi, arvioinnin pituutta ajatellen ja toiseksi, huonon äänenlaadun vuoksi. Arvioijat valikoituivat lumipallo-otannan kautta; sähköinen kyselylomake lähetettiin ensin muutamalle natiivienglanninpuhujalle sähköpostitse, ja he jakoivat linkkiä eteenpäin. Tähän menetelmään päädyttiin, koska se arvioitiin tehokkaammaksi tavaksi löytää sellaisia englantia äidinkielenään puhuvia, jotka eivät puhu sujuvasti suomea, kuin etsiä heitä tarvittava määrä paikallisesti.

Aineistona tutkimuksessa käytettiin ennen ääntämiskurssia ja sen jälkeen äänitettyjä tuotoksia sekä natiiviarvioijille suunnatun kyselyn vastauksia. Äänitykset toteutettiin syyslukukaudella 2017 Turun yliopiston englannin oppiaineessa. Spoken English -ääntämiskurssi kuuluu englannin kielen pakollisiin perusopintoihin ja suoritetaan tavallisesti ensimmäisenä syksynä. Opiskelijat lukivat molemmilla kerroilla ääneen eri tekstit. Ennen kurssia lukiessaan opiskelijat eivät tienneet, mitä suorituksessa arvioidaan. Kurssin jälkeen tehty testi oli kurssin loppukoe ja opiskelijat tiesivät, mitä heidän odotettiin oppineen kurssin aikana. Tämän oletettiin aiheuttavan tarkempaa segmentaalien ääntämistä ja mahdollisesti myös hermostuneisuutta. Näytteiden joukkoon lisättiin kontrollin vuoksi myös yhden natiivipuhujan otokset samoista teksteistä. Sähköinen kyselylomake suunniteltiin varta vasten tätä tutkimusta varten, mutta sen pohjana käytettiin aiemmissa tutkimuksissa hyödynnettyjä menetelmiä. Arvioijat vastasivat ensin taustakysymyksiin ja kuuntelivat sitten 42 lyhyttä äänitystä, jotka arvioitiin yhdeksänportaisella Likert-asteikolla. Puherytmi, aksenttisuus, sujuvuus sekä ymmärrettävyys arvioitiin kyselyn ohjeiden mukaisesti.

Tutkimus eteni seuraavalla tavalla: ensin opiskelijoiden tuotokset kuunneltiin ja niiden pohjalta valittiin 20 opiskelijaa. Näiden opiskelijoiden sekä yhden natiivipuhujan tuotoksista leikattiin lyhyet näytteet tarkkaan valituista kohdista, ja ne liitettiin natiivipuhujille suunnattuun kyselylomakkeeseen. Kysely luotiin Google Forms -palvelun avulla ja sitä levitettiin sähköpostitse sekä sosiaalisessa mediassa. Kun kyselyyn oli kerätty tarpeeksi vastauksia, ne analysoitiin tilastollisesti SPSS- ja Microsoft Excel-ohjelmien avulla. Arvioijien avoimet kommentit otettiin huomioon tuloksia tulkittaessa. Jokaiselle opiskelijalle laskettiin pisteet ennen- ja jälkeen-tuotoksille, ja niiden perusteella valittiin neljä edistyneintä opiskelijaa akustiseen analyysiin. Akustinen analyysi tehtiin Praat-puheanalyysiohjelman avulla ja suurimmaksi osaksi manuaalisesti. Sujuvuuden mittareina käytettiin artikulaationopeutta, taukojen määrää, kestoa ja sijaintia sekä korjauksia. Puherytmiä varten analysoitiin perustaajuuden vaihteluväli, perustaajuuden ja amplitudin huippujen yhtenevyys, vokaalien laatu, keston käyttö sanapainon merkkinä sekä peräkkäisten äänteiden nivonta ja äänteiden poisjättö. Kaikki kahdeksan arvioitua äänitystä litteroitiin sujuvuustekijöiden havainnollistamiseksi. Rytmin osatekijöiden analyysia varten jokaisesta kahdeksasta tuotoksesta piirrettiin Praatin avulla spektrogrammi.

Tutkimustulokset ja johtopäätökset

Tilastollisen analyysin perusteella osalla opiskelijoista tapahtui positiivista ja osalla negatiivista kehitystä. Rytmissä puolet oppijoista (N = 10) paransi pisteitään ja puolet huononsi, kun taas sujuvuus huononi peräti kolmellatoista opiskelijoista. Aksenttisuudessa pisteitään paransi puolet puhujista ja ymmärrettävyydessä vain seitsemän. Keskimäärin puherytmi, aksenttisuus sekä sujuvuus paranivat hieman, kun taas ymmärrettävyys huononi. Kuten hypoteesissani oletin, tilastollisesti mitattuna nämä erot eivät olleet kuitenkaan merkittäviä. Vastausten reliabiliteettia mitattiin Cronbachin alphalla ja Split Half -menetelmällä, joiden mukaan arvioijat olivat riittävän yhdenmukaisia arvioissaan sekä ennen-jälkeen -vertailussa että jokaisen yksittäisen aspektin kohdalla. Arvioijien kommentit viittaavat siihen, että he osasivat arvioida neljää ominaisuutta erillisinä, vaikkakin osa huomasi näiden vaikuttavan toisiinsa. Määrällisen analyysin perusteella valittiin neljä opiskelijaa, joiden ennen- ja jälkeen-tuotoksissa oli suurimmat erot.

Kun näiden suurimmat erot saavuttaneiden opiskelijoiden otoksia tarkasteltiin erikseen akustisesti ja auditiivisesti, kävi ilmi, että kasvanut artikulaationopeus, hiljaisten taukojen keston vähentyminen ja niiden paikan muutos lausekkeiden rajoille sekä korjausten puuttuminen johtivat korkeampiin sujuvuuspisteisiin. Aiempien tutkimustulosten tapaan nämä neljä suomenkielistä englannin oppijaa eivät käyttäneet täytettyjä taukoja puheessaan. Myös rytmin kannalta puheessa näytti olevan useita osatekijöitä, joita opiskelijat yhdistelivät puheessaan eri suhteissa. Kaikilla oppijoilla sävelkorkeuden ja amplitudin huiput osuivat jokseenkin samoihin kohtiin joko molemmissa tai vain jälkeen-tuotoksessa. Huomionarvoista oli myös, että oppijat, joiden puherytmissä oli suurimmat erot (M2 ja F25), oppivat käyttämään vokaalireduktiota ja tekemään selkeämmän eron painollisten ja painottomien tavujen kestoissa. Hypoteesi osui siis osittain oikeaan. Kaikki neljä arvioitua puheen piirrettä korreloivat vahvasti keskenään, kuten hypoteesissani oletin, rytmi ja sujuvuus eniten ($r = 0.98$). Aikaisempien tutkimustulosten tapaan sujuvuus ja ymmärrettävyys korreloivat selkeämmin keskenään kuin aksenttisuus ja ymmärrettävyys. Kaikki korrelaatiosuhteet olivat tilastollisesti merkittäviä.

Tutkimuksen merkitystä pohdittiin ääntämisen opetuksen kannalta. Tuskin on tarkoituksenmukaista, että opiskelijoiden puheen ymmärrettävyys laskee kurssin myötä. Kurssin segmentaalinen painopiste ja prosodian vähäinen merkitys loppukokeen arvioinnissa saattoivat osaltaan vaikuttaa tuloksiin. Koska on todennäköistä, että jälkeen-äänityksen tilanne on vaikuttanut negatiivisesti tuotoksiin, ehdottaisin Spoken English -kurssille formatiivista arviointia:

jokaisen harjoituskerran jälkeen opiskelijat saisivat tehtävän, jonka voisivat äänittää itsekseen kotona ja lähettää opettajalle arvioitavaksi. Tämä vähentäisi summatiivisessa loppukokeessa koettua jännitystä sekä tuon kokeen painoarvoa kurssin kokonaisarvioinnissa.

Tutkimuksen reliabiliteettia ja validiteettia on syytä arvioida kriittisesti. Olen pyrkinyt ottamaan nämä huomioon tutkimuksen kaikissa vaiheissa aina lähteiden valinnasta metodologisiin kysymyksiin ja analyysiin. On kuitenkin selvää, että resurssien ollessa rajalliset, täydellinen reliabiliteetti ja validiteetti eivät ole realistisia. Tutkimuksen yleistettävyyttä voisi parantaa lisäämällä koehenkilöiden määrää. Laajemmat taustatiedot opiskelijoista puolestaan saattaisivat selittää yksilöllisiä eroja. Oppijoiden kehitystä olisi myös parempi tutkia yhdellä samalla tekstillä ja samanpituisilla otoksilla, jolloin tuotosten vertailu olisi luotettavampaa. Natiiviarviointia voisi parantaa osallistujien valvonnalla sekä ääninäytteille paremmin soveltuvalla kyselylomakkeella. Aiheen tutkimusta edistäisi pitkittäistutkimus, jossa testattaisiin samojen oppijoiden kehitystä esimerkiksi ensimmäisenä ja kolmantena opintovuonna. Kontrolliryhmä auttaisi varmistamaan, että erot johtuvat nimenomaan opetuksesta, ja kielen oppijoiden äänitykset myös äidinkielellään lisäisivät tietoa prosodian siirtovaikutuksesta. Puherytmin ja sujuvuuden välisen suhteen tarkempi määrittely olisi myös tarpeen – onko mahdollista erottaa vain rytmiin ja toisaalta vain sujuvuuteen vaikuttavat tekijät puheessa? Toisin kuin monet alan tutkijat, pitäisin akustisen analyysin vielä osana rytmitutkimusta, ennen kuin saadaan lisää todisteita rytmin luonteesta pelkästään havaittuna ilmiönä.

Lopuksi

Tämän pro gradu -tutkielman tavoitteena oli tutkia, muuttuvatko opiskelijoiden puherytmi ja sujuvuus ääntämiskurssin tuloksena, ja jos kyllä, niin mitkä akustiset ominaisuudet saavat aikaan nämä havaitut muutokset. Halusin selvittää myös, korreloivatko neljä arvioitua ominaisuutta keskenään. Tulosten perusteella muutoksia tapahtui kaikilla osa-alueilla, mutta ryhmätasolla ne eivät olleet merkitseviä. Kun neljää eniten kehittynyttä opiskelijaa tarkasteltiin erikseen, erot olivat merkitseviä erityisesti rytmissä ja sujuvuudessa. Opiskelijat saivat korkeammat sujuvuuspisteet, kun heidän artikulaationopeutensa kasvoi, hiljaisten taukojen määrä väheni ja niiden sijainti siirtyi lausekkeiden rajoille ja korjaukset jäivät pois. Akustisen analyysin perusteella rytmiin positiivisesti vaikuttavat tekijät näyttivät olevan pääasiassa sävelkorkeuden ja intensiteetin huippujen yhtenevyys, vokaalien redusointi sekä keston vaihtelu. Lisäksi kaikki neljä arvioitua

piirrettä korreloivat voimakkaasti keskenään, rytmi ja sujuvuus eniten. Lopuksi, tutkimustulosten perusteella sekä puherytmin että sujuvuuden opettaminen kannattaa, sillä ne vaikuttavat merkittävästi paitsi toisiinsa myös havaittuun aksenttisuuteen sekä ymmärrettävyyteen. Tuloksiin on kuitenkin suhtauduttava varovaisuudella, sillä useat tekijät, kuten kurssin opetuksen fokus, koehenkilöiden määrä, erilaiset tekstit testauskerroilla sekä äänitysten kestoerot, ovat saattaneet osaltaan vaikuttaa tutkimuksen lopputulokseen. Rytmin tutkimisessa tullaan tulevaisuudessa toivon mukaan keskittymään erityisesti sen luonteeseen sekä tuotettuna että havaittuna puheen ilmiönä ja selvittämään sen ja sujuvuuden suhdetta tarkemmin.