

**FEATURES OF PRONUNCIATION CONTRIBUTING TO  
INTELLIGIBILITY AND COMPREHENSIBILITY IN L1  
FINNISH SPEAKERS' L2 ENGLISH**

Jarno Laitinen  
Master's Thesis  
English  
Department of Language and  
Communication Studies  
University of Jyväskylä  
May 2022

# JYVÄSKYLÄN YLIOPISTO

|   |  |
|---|--|
| <b>Tiedekunta</b><br>Humanistis-yhteiskuntatieteellinen tiedekunta  | <b>Laitos</b><br>Kieli- ja viestintätieteiden laitos |
| <b>Tekijä</b><br>Jarno Laitinen   |  |
| <b>Työn nimi</b><br>The Features of Pronunciation Contributing to Intelligibility and Comprehensibility in L1 Finnish Speakers' L2 English  |  |
| <b>Oppiaine</b><br>Englanti   | <b>Työn laji</b><br>Maisterintutkielma               |
| <b>Aika</b><br>Toukokuu 2022  | <b>Sivumäärä</b><br>59 + 4 liitettä                  |
| <b>Tiivistelmä</b> <p>Maisterintutkielmani keskittyy puheen ymmärrettävyyteen vaikuttavien piirteiden tutkimiseen äidinkielenään suomea puhuvien englannin kielessä. Englannin globaali asema alati kasvavana yleiskielenä, lingua francana, on siirtänyt puheen opettamisen painopisteen syntyperäisenkaltaisesta osaamisesta keskinäiseen ymmärrykseen, koska suurin osa englannin nykyisistä käyttäjistä ei puhu sitä äidinkielenään. Tutkielmani tarkoituksena on selvittää, mitkä puheen piirteet ovat oleellisia ymmärrettävyyden kannalta. Tämän lisäksi tutkimuskohteinani ovat eroavatko englantia äidinkielenään puhuvien ja englantia toisena kielenään puhuvien ymmärrettävyydsarviot toisistaan sekä auttaako saman äidinkielen jakaminen ymmärtämään vieraskielistä puhetta paremmin (ns. shared language benefit).</p> <p>Tutkielmani aineisto koostuu englantia äidinkielenään puhuvien (<math>n = 36</math>) ja suomea äidinkielenään puhuvien (<math>n = 31</math>) vastauksista, jotka keräsin internetpohjaisella arviointikyselyllä joulukuussa 2021 ja tammikuussa 2022. Kyseeseen osallistujat kuuntelivat suomalaisten puhuman englannin puhenäytteitä ja arvioivat näitä käyttäen liikusäädinasteikoita sekä avoimia kommenttikenttiä. Aineiston analyysissä hyödynsin mixed methods -lähestymistä, yhdistäen tilastollista analyysiä t-testein sekä teoriaohjaavaa sisällönanalyysiä.</p> <p>Tutkimustulokseni osoittivat, että englantia äidinkielenään puhuvat ymmärsivät suomalaisten puhumaa englantia paremmin jokaisella mitatulla osa-alueella. Näin ollen shared language benefit ei löytynyt tässä tutkimuksessa, eli saman äidinkielen jakaminen puhujan ja kuuntelijan välillä ei vaikuttanut puheen ymmärrettävyyteen positiivisesti. Tämän lisäksi arvioijaryhmien välillä oli tilastollisesti merkitseviä eroja koskien puheen kirjaimellista ymmärrettävyyttä, ymmärrettävyyden vaivan määrää, puheen segmenttejä sekä ydinsanomien ymmärtämistä. Kaikilla näillä osa-alueilla ensikieliset englannin puhujat ymmärsivät suomalaisten puhumaa englantia paremmin. Kaikista positiivisimmin puheen ymmärrettävyyteen vaikutti konteksti molemmilla kuuntelijaryhmillä. Ymmärrettävyyttä taas puolestaan häiritsivät eniten yksittäiset sanat, puherytmi sekä segmentit.</p> <p>Tutkimus antaa viitteitä ymmärrettävyyteen vaikuttavista piirteistä, jotka voivat olla hyödyllisiä niin kielten tutkimuksen kentälle kuin kielten opettajille.</p> |  |
| <b>Asiasanat</b> – Ääntäminen, prosodia, ymmärrettävyys, fonetiikka, mixed methods, shared language benefit   |  |
| <b>Säilytyspaikka</b> – Jyväskylän yliopiston julkaisuarkisto JYX   |  |
| <b>Muita tietoja</b> –  |  |

## ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to the following people and organizations, without whom this work would have not been possible in its present form.

Elina Tergujeff, for your humbling expertise and precise guidance.

Henna Heinonen, for your outstanding encouragement and spark.

Jukka Hintikka, for your never-ending council and help with statistical methods.

Viivi Kuusiahho, for your all-around support.

Anne Pitkänen-Huhta, for your excellent comments and seminar guidance.

Finally, I wish to thank the U.S Embassy of Finland for admitting this paper with the American Studies Grant, which had a profound effect on finishing it as best as I could. Thank you all.

## FIGURES

|   |    |
|---|----|
| Figure 1: Illustration of linguistic disciplines.....                   | 3  |
| Figure 2: The International Phonetic Alphabet chart.....                | 4  |
| Figure 3: Vowels in International Phonetic Alphabet .....               | 5  |
| Figure 4: Finnish vowel chart .....                                     | 6  |
| Figure 5: Monophthongs in Standard American English .....               | 6  |
| Figure 6: Consonants in Finnish phonology .....                         | 8  |
| Figure 7: Consonants in English phonology .....                         | 8  |
| Figure 8: The content word counting in the intelligibility section..... | 32 |
| Figure 9: The initial coding frame.....                                 | 34 |
| Figure 10: The final coding frame .....                                 | 34 |

## TABLES

|  |    |
|--|----|
| Table 1: Exemplification of parametrical and non-parametrical tests..... | 31 |
| Table 2: The results concerning the intelligibility section.....         | 37 |
| Table 3: Results of the comprehensibility section .....                  | 39 |
| Table 4: The times feature was mentioned in the comments.....            | 42 |

## Table of Contents

|   |  |    |
|---|--|----|
| 1 | INTRODUCTION.....                                | 1  |
| 2 | ON LINGUISTICS AND PHONETICS .....               | 3  |
|   | 2.1 Vowels .....                                 | 5  |
|   | 2.2 Consonants.....                              | 8  |
|   | 2.3 Stress .....                                 | 10 |
|   | 2.4 Intonation.....                              | 11 |
|   | 2.5 Rhythm .....                                 | 11 |
| 3 | ON INTELLIGIBILITY AND COMPREHENSIBILITY.....    | 13 |
|   | 3.1 Key terminology.....                         | 13 |
|   | 3.2 Studies on intelligibility .....             | 15 |
|   | 3.3 Studies on comprehensibility .....           | 16 |
|   | 3.4 Shared language benefit.....                 | 19 |
|   | 3.5 Pronunciation learning theories.....         | 20 |
|   | 3.6 Pedagogical considerations .....             | 24 |
| 4 | THE PRESENT STUDY.....                           | 26 |
|   | 4.1 Aim and research questions .....             | 26 |
|   | 4.2 The listener groups and speech samples ..... | 27 |
|   | 4.3 The rating task.....                         | 28 |
|   | 4.4 Mixed methods.....                           | 30 |
|   | 4.5 Statistical testing and the t-test.....      | 31 |
|   | 4.6 Content analysis .....                       | 32 |
|   | 4.7 Research strategy and epistemology.....      | 35 |
|   | 4.8 Ethical considerations.....                  | 36 |
| 5 | RESULTS.....                                     | 37 |
|   | 5.1 Intelligibility.....                         | 37 |
|   | 5.2 Comprehensibility.....                       | 38 |
| 6 | DISCUSSION.....                                  | 44 |
| 7 | CONCLUSION .....                                 | 50 |
|   | REFERENCES.....                                  | 52 |
|   | APPENDICES.....                                  | 60 |
|   | Appendix 1. The rating task form.....            | 60 |

|   |    |
|---|----|
| Appendix 2. The text used in audio samples .....                | 70 |
| Appendix 3. The phrases used in intelligibility section.....    | 71 |
| Appendix 4. The passages used in comprehensibility section..... | 72 |

# 1 INTRODUCTION

The role of the English language has increasingly developed from a language spoken by native speakers to a global lingua franca. In fact, of the over 1.3 billion worldwide speakers of English (Szmigiera 2021), nearly 1 billion speak it as a second or as a foreign language (Eberhard, Simons, and Fennig 2021). Hence, the idol of native-speaker-like pronunciation is no longer considered the highest priority (Jenkins 1998; Deterding 2010; Deterding and Lewis 2019). Consequently, the traditional pronunciation models of Southern Standard British English (formerly Received Pronunciation) and GA (General American) are not pursued in the Finnish education system or held in prestige *ad finem*. Moreover, the traditional models alone do not suffice in terms of intelligible communication, as the majority of English speakers are not natives (Sajavaara and Dufva 2001). Nevertheless, the profound need to convey information clearly and effectively, and to be understood via spoken English, remains the same.

The most important aspect of communication, regardless of the context, is widely agreed to be intelligibility (Levis 2018; Munro and Derwing 1995b). This principal premise of intelligibility is also noted in the guidelines of the Common European Framework of Reference for Languages (Council of Europe 2020), abbreviated as CEFR. The CEFR is a unified manual all around Europe used for assessing linguistical competence, and consequently in Finland as well. Surely, especially for English teachers both existing and becoming, such as myself, the information about the salient aspects on what contributes to intelligibility is highly valuable. Not only are they relevant in terms of guiding students towards understandable communication, but may be seen also useful in improving mutual understanding regarding oral communication outside classrooms.

However, what exactly contributes to being understood is a complex matter. As stated by Zielinski (2008) our knowledge is not nearly complete on what the relevant features and cues that the listeners rely on when trying to understand the speaker's utterances are, and what features, in turn, lead astray. Furthermore, the features and aspects that contribute to intelligibility and comprehensibility in terms of Finnish English have not been studied as extensively as with other languages, or indeed, at all (see e.g. Jesney 2004). Hence, what contributes to intelligibility and comprehensibility are

highly relevant aspects for teachers and laymen alike, as communication through English is not contrived only into classrooms but is and continues to spread globally.

This master thesis is a study in the field of linguistics, namely acoustic phonetics. As such, it is set out to investigate the features of pronunciation in L1 Finnish speakers' L2 English, and how they relate to the concepts of intelligibility and comprehensibility. Within my study, I am interested in the *impressions of the hearer*, regarding the pronunciation features, namely the segmental and suprasegmental ones, and how they relate to the abovementioned concepts of intelligibility and comprehensibility. Segmentals are the smallest units of speech, such as vowels and consonants, and suprasegmental features consist of the larger units, such as word stress, sentence stress, intonation, and speech rhythm (Ogden 2009: 23). In turn, intelligibility corresponds to "the extent to which a speaker's message is actually understood by a listener" (Munro and Derwing 1995a: 76), and comprehensibility equaling to the "[listener's] effort involved in understanding" (Munro and Derwing 1995b). These concepts are the key focus of my study. They have been studied extensively before, together as well as separately, yet studies concerning Finnish spoken English have been scarce, expressing a clear research gap, which this study aims to contribute to.

Furthermore, the previous studies have shown inconclusive results concerning the so-called shared language benefit (section 2.4). The shared language benefit is a disputable phenomenon of gaining aid to mutual understanding due to sharing the same first language (Bent and Bradlow 2003; Foote and Trofimovich 2018). In other words, whether a native Finnish speaker understands another native Finnish speaking English somehow better due to them sharing Finnish as their mother tongue. This aspect has not been studied, at least extensively, in the Finnish English context. Thus, this study also aims to reduce this research gap.

The thesis is divided into seven chapters. After this introduction, in the second chapter, I contrast the differences between English and Finnish sound systems in detail. In chapter 3, I discuss and review the previous research done in the field and define the key concepts used in this study. I address, in addition, theoretical perspectives on how pronunciation is learned and how languages may affect each other. As a future teacher of English, I also consider some pedagogical perspectives. In chapter 4, I explain how this study was conducted and how the data was collected and analyzed. In chapter 5, I present the results obtained and in chapter 6, I discuss these results in relation to my theoretical framework. Finally, I conclude the thesis in chapter 7 by summarizing the study and adding notes on future research.

The research goals are ambitious and challenging for a Master's level study. As such, my motivation for transcendence comes from the aim to contribute to the scientific field with practical relevance from this study.



## 2 ON LINGUISTICS AND PHONETICS

In this chapter, I review the relevant linguistic features and aspects for my study. First, I position my study within the field of linguistics, and secondly, explain the key terminology on the concepts and marking systems used in the field. Thirdly, I present the features salient for this study, namely the vowels and consonants, and features of stress, intonation, and rhythm.

Linguistics is the overarching term for the study of language, encompassing such sub-disciplines as pragmatics, studying meaning in context, phonology, interested in sound systems, and phonetics, studying speech sounds (Ogden 2009). Phonetics can be further divided into several subdisciplines, wherein this study's position is founded in *auditory phonetics*, concentrating on how the human auditory system is involved in perceiving sounds (Gick, Wilson, and Derrick 2013).

Figure 1 depicts the position of phonetics within the innermost square. It is to be noted that the figure is not comprehensive, leaving aside such major branches as sociolinguistics, semiotics, and discourse analysis. Nevertheless, it serves as an illustration of the various disciplines and their connections.

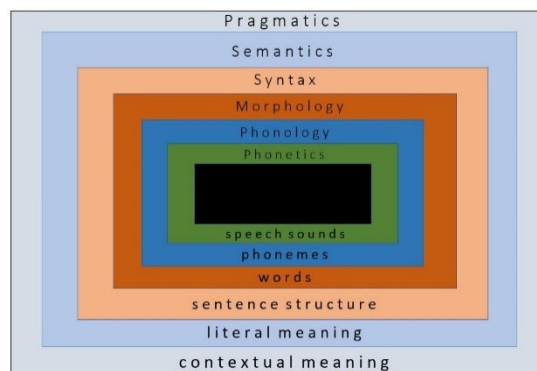


Figure 1: Illustration of linguistic disciplines and the position of phonetics within. Based on Thomas and Cook (2005: 110).

In linguistic terminology, the smallest unit of speech carrying a distinguishing significance is called the *phoneme*, written in between slashes, such as /æ/ or /e/. In Finnish, these examples put in word context “älä” (don’t) and “elä” (live) form a *minimal pair* as they change the meaning of the word (Savolainen 2001). In English, an illustration of a minimal pair could be “call” and “tall” wherein the phonemes /k/ and /t/ distinguish the word meaning (McMahon 2016: 14).

The *phone* is in close relation to the *phoneme*. Realized in actual speech, the vowels and consonants are called *phones*. *Phones* are the letters in speech, arranging to words consisting of multiple of these speech units, arranging to utterances and so on. They are the single speech units realized in speech; not abstractions but concrete manifestations of the phonemes (Savolainen 2001). In phonemic writing, and thus in this study, phonemes are written in between slash brackets //, and the phones within square brackets [ ], to differentiate them.

The international phonetic alphabet (IPA), in Figure 2, is an alphabet for phonetic transcription, used to display speech sounds in written form (International Phonetic Association 1999). In IPA, all the sounds have their corresponding place of origin in human anatomy: which parts of the mouth are active when a sound is uttered. For example, a voiceless glottal fricative means that the vocal cords are not vibrating when pronounced (hence it is called voiceless), the sound originates in the glottis of the vocal cords (hence it is a glottal) and the air flows from the mouth causing friction (hence it is called a fricative) (Savolainen 2001; McArthur, Lam-McArthur, and Fontaine 2018).

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2020)

CONSONANTS (PULMONIC) © 2020 IPA

|                     | Bilabial | Labiodental | Dental | Alveolar | Postalveolar | Retroflex | Palatal | Velar | Uvular | Pharyngeal | Glottal |
|---------------------|----------|-------------|--------|----------|--------------|-----------|---------|-------|--------|------------|---------|
| Plosive             | p b      |             |        | t d      |              | ʈ ɖ       | c ɟ     | k ɡ   | q ɢ    |            | ʔ       |
| Nasal               | m        | ɱ           |        | n        |              | ɳ         | ɲ       | ŋ     | ɴ      |            |         |
| Trill               | ʙ        |             |        | r        |              |           |         |       | ʀ      |            |         |
| Tap or Flap         |          | ⱱ           |        | ɾ        |              | ɽ         |         |       |        |            |         |
| Fricative           | ɸ β      | f v         | θ ð    | s z      | ʃ ʒ          | ʂ ʐ       | ç ʝ     | x ɣ   | χ ʁ    | ħ ʕ        | h ɦ     |
| Lateral fricative   |          |             |        | ɬ ɮ      |              |           |         |       |        |            |         |
| Approximant         |          | ʋ           |        | ɹ        |              | ɻ         | j       | ɰ     |        |            |         |
| Lateral approximant |          |             |        | l        |              | ɭ         | ʎ       | ʟ     |        |            |         |

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 2: The International Phonetic Alphabet chart. CC-BY-SA 3.0. © 2018 International Phonetic Association.

Furthermore, for instance, “car” and “key” are written in IPA as /ka:(r)/ and /ki:/ in Standard Southern British English. This phonemic transcription allows us to study and compare the speech sounds between two languages in depth. No two languages share the same phonological structure but their differences can be contrasted with the phonemic transcription.

I use IPA in my study, as it allows the presentation of an uttered sound more accurately than it would if it was written in standard Latin-script-based alphabet. IPA depicts 59 consonants and 28 vowels, each having multiple concurrent possibilities of further specification (diacritics) whereas the standard alphabets of both Finnish and English have 29 and 26 letters, respectively, with no further specification. When demonstrating the words, I write the original word in italics and place it within quotation marks followed by the IPA transcription in square brackets and translation in round brackets. For example, “*example*” [ɪɡzæmpl̩] (esimerkki). I keep the diacritics to the minimum, mostly using only the essential /:/, which is used to mark the lengthening of a phoneme.

It is to be noted that while the terms of phonemes and phones apply to all languages in general, they cannot be applied similarly to one another, as no two languages share the same phonology (McMahon 2016). Furthermore, a thorough explanation of these concepts is outside the scope of my thesis.

Next, I present the relevant features in my study, the segmentals divided into vowels and consonants, and the suprasegmentals, divided into stress, intonation, and rhythm.

## 2.1 Vowels

Wiik (1998: 35) defines the vowels to be speech sounds, wherein airflow is unobstructed and thus can continuously depart from the center of the mouth. This definition is convenient, as it does not only define vowels universally but provides a definition for consonants as well. More closely, Wiik (ibid.) defines consonants to be sound structures that do not meet the criteria for vowels. As said, these definitions are universals, and they fit both the Finnish and English sound structures.

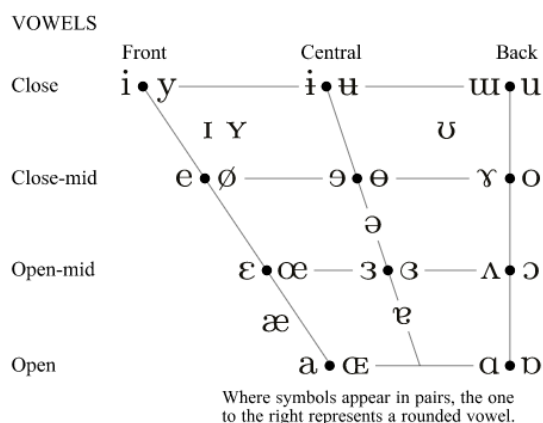


Figure 3: Vowels in International Phonetic Alphabet. CC-BY-SA 3.0. © 2018 International Phonetic Association.

The quadrilateral diagram, Figure 3, represents the vowels as depicted in IPA and their place of production within the mouth. From this, we can see that for instance the phoneme /i/ can be called a *close front unrounded vowel*, found in English and Finnish words such as “free” [fri:] (vapaa) and “hiiva [hi.va]” (yeast). The name *close front unrounded vowel* signifies that when uttering a word like “see” [si:] (nähdä) the tongue is *close* to the roof of the mouth, the *front* of the tongue is being raised, and the lips are *unrounded* (Roach 2009). Similarly, a phoneme /o/ is called a *close-midback rounded vowel*, found in the center right in Figure 3, appearing in words such as “yawn” [jo:n] (haukotus) in English and “ohjus” [ohjus] (missile) in Finnish. Hence, its name signifies that the tongue’s position within the mouth is *close-mid* to the roof of the mouth, the *back* of the tongue is being raised, and the lips are *rounded*.

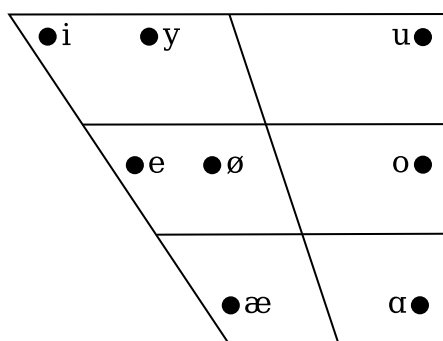


Figure 4: Finnish vowel chart. Based on Suomi, Toivanen, and Ylitalo (2008: 21). CC-BY-SA 4.0. Wikimedia Commons.

The Finnish vowels are /i e y ø æ a o u /, depicted in Figure 4, and realized in words such as *tikin, tekin, tykin, tökin, täkin, takin, token, tukin* (Suomi, Toivanen, and Ylitalo 2008: 20). Depiction of English vowels, however, is much more complicated as the variant and furthermore accent and individuality of English changes their realizations. Nevertheless, the vowels, as they appear in Standard American English are illustrated in Figure 5 and they are /i ɪ u ʊ e ε ə o æ ʌ ɔ /.

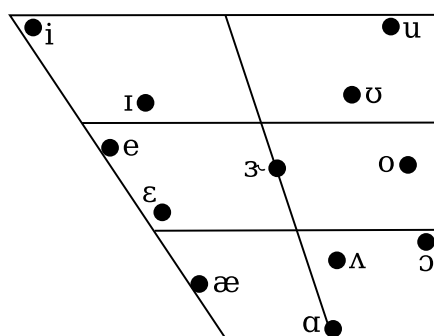


Figure 5: Monophthongs in Standard American English. Based on Wells (1982: 486). CC BY-SA 4.0. Wikimedia Commons.

In both Finnish and English, the vowels can be realized as monophthongs, meaning that there is no change in their quality within a syllable; on how they are realized (Ladefoged and Johnson 2010: 308). This can be seen with the phoneme /i/ in the English word “this” [ðis] (tämä), or in a Finnish word “tili” [tili] (account). In both examples, the /i/ remains unchanged.

In addition to monophthongs, the vowels can be realized as diphthongs, meaning that the quality of the vowel changes from one vowel to another (Ladefoged and Johnson 2010: 92). In English, there are eight diphthongs, such as /aɪ/ in the word “eye” wherein the first vowel “a” glides towards the accompanying /ɪ/ (Roach 2009: 17). In Finnish, the diphthongs are combinations of the eight vowels and as the vowels can also appear as double, there are 18 of them, such as “tuoli” [tuoli] (chair) (Suomi et al. 2008: 49).

As opposed to English, Finnish is a quantity language, meaning that the quantity of the vowels (and consonants) have a distinctive meaning (Suomi et al. 2008: 39). As such, Suomi et al. (ibid.) continue that all the vowels can appear as double in Finnish words, and the tensing of a vowel does not change the quality of the vowel itself but functions as a phoneme, as it changes the word meaning. This can be seen, for instance in the words “tuli” [tuli] (fire) in contrast to the tensed “tuuli” [tu:li] (wind) or in “lama” [lama] (recession) in contrast to the tensed “laama” [la:ma] (llama).

Finally, I address the concept of weak and strong forms regarding vowels. In connected speech, which means that we do not speak in individual, separated words, the sounds influence each other (Ladefoged and Johnson 2010). The authors continue that the *strong form* means how the sound appears when emphasized or uttered without accompanying words. They present an example of the indefinite article “a” (eräs) and “as” (kuten), which have the sounds [ei] and [æ], respectively, when uttered separately. The authors write that this changes when they are accompanied by other words. They explain that in the *weak form* (with accompanying words, when the syllable is not emphasized) the “a” article appears as schwa [ə], as in a word combination of “a cup” [ə kʌp] (kuppi) and the [æ] sound changes similarly to a schwa [ə], when accompanied by another word; “as good as” [əz gʊd əz] (yhtä hyvä kuin).

As such, at least Zielinski (2008) has found that the strong and weak syllables appear to be important when native English speakers distinguish sounds. Moreover, Cutler and Butterfield (1992) write that English speakers segmentate the heard words at least partially by relying on strong and weak syllables, deeming strong syllables as initiations or lexical words, whereas the weak syllables are perceived as non-initial or grammatical words. As with many other languages, assimilation is also a common phenomenon in Finnish (see Suomi. et al. 2008: 43) but since I am concentrating on samples spoken in English, I do not discuss assimilation in Finnish here.

## 2.2 Consonants

As stated previously, consonants can be defined as speech sounds that do not match the criteria for vowels (Wiik 1998: 35). Like vowels, consonants can be described by their place of production in the mouth and grouped accordingly to, for example, bilabials (contact between the lips), labiodentals (contact between the lower lip and upper teeth), or alveolars (contact between the tongue and alveolar ridge) (Ogden 2009).

Suomi et al. (2008: 25) point out that the number of Finnish consonants varies from 11 to 17 depending on the speaker's regional dialect and these are /p t k s h l r m n j v ɲ d f b g ʃ/. Figure 6 depicts them all, encompassing the highest number, 17 consonant sounds, in Finnish phonology.

|                 |                     | Way of art. → |             |                      |                 |                  |                |         |       |         |
|-----------------|---------------------|---------------|-------------|----------------------|-----------------|------------------|----------------|---------|-------|---------|
|                 |                     | Bilabial      | Labiodental | Laminal dentalveolar | Apical alveolar | Laminal alveolar | Palatoalveolar | Palatal | Velar | Glottal |
| Place of art. ↓ | Plosive             | p (b)         |             | t                    |                 |                  |                |         | k (g) |         |
|                 | Semiplosive         |               |             |                      | d               |                  |                |         |       |         |
|                 | Fricative           |               | f           |                      |                 | s                | (ʃ)            |         |       |         |
|                 | Glottal continuant  |               |             |                      |                 |                  |                |         |       | h       |
|                 | Nasal               | m             |             |                      | n               |                  |                |         | ŋ     |         |
|                 | Trill               |               |             |                      | r               |                  |                |         |       |         |
|                 | Lateral approximant |               |             |                      | l               |                  |                |         |       |         |
|                 | Central approximant |               | v           |                      |                 |                  | j              |         |       |         |

Figure 6: Consonants in Finnish phonology. The ones within brackets appear only in loan words. Based on Suomi et al. (2008).

|                 |             | Way of art. → |             |        |          |               |         |       |         |
|-----------------|-------------|---------------|-------------|--------|----------|---------------|---------|-------|---------|
|                 |             | Labial        | Labiodental | Dental | Alveolar | Post alveolar | Palatal | Velar | Glottal |
| Place of art. ↓ | Plosive     | p b           |             |        | t d      |               |         | k g   |         |
|                 | Nasal       | m             |             |        | n        |               |         | ŋ     |         |
|                 | Affricative |               |             |        |          |               | tʃ dʒ   |       |         |
|                 | Fricative   | (w)           | f v         | θ ð    | s z      | ʃ ʒ           |         | (x)   | h       |
|                 | Approximant | w             |             |        | l r      |               | j       |       |         |

Figure 7: Consonants in English phonology. The ones in brackets appear only in loan words. Based on McMahon (2016).

In contrast, as presented in Figure 7, English has 24 consonant sounds and, in some varieties, even 26 (McMahon 2016: 53). This adds 7-9 additional sounds for the Finnish speaker to learn.

Some of the consonant phonemes that appear in both languages are realized differently. For example, the /p t and k/ are realized as *unaspirated* voiceless plosives in Finnish (Suomi et al. 2008: 26), as in “*piina*” [pi:na] (torment), “*Tiina*” [ti:na] (Tiina) or “*Kiina*” [ki:na] (China). In turn, they are *aspirated* voiceless plosives in English, the aspiration meaning the “air passing through the glottis and then the vocal tract” (Ogden 2009: 102), realized in words such as “*pie*” [paɪ] (piiras), “*tie*” (ta) (solmio) or “*kye*” [kaɪ] (antiikkinen sana lehmälle). The term *voiced* refers to the vibration of the vocal cords whereas a *voiceless* sound is one without the vibration; contrasted to the above-mentioned “*pie*” which is voiceless, the /b/ in “*buy*” [baɪ] (ostaa) is in turn voiced (Ladegofed and Johnson 2010: 4). As such, the phonetical differences between voiceless and voiced sounds are relevant distinctive features in all languages (McMahon 2016: 26).

In addition to differences in aspiration and voicing, the consonants that are not present in the Finnish phonological system may cause trouble for a Finnish-speaking English learner, for instance, the /dʒ/ and /tʃ/ sounds (Sajavaara and Dufva 2001: 250). Moreover, the /v/ and /w/ are considered allophones in Finnish phonology, whereas they are distinct phonemes in English (ibid.: 246). The difference, therefore, between the labiodental central approximant /v/ and the labial approximant of /w/ may be understandably difficult for a Finnish speaker to understand. How it affects the concepts of intelligibility or comprehensibility, however, is not known.

Moreover, some of the consonants, /f b g and ʃ/, appear only in Finnish loan words, even if they appear nowadays regularly in the Finnish lexicon (Suomi et al. 2008). The /b/, and /g/ are present in the words such as “*baari*” [ba:ri] (bar) or “*galleria*” [gal:eria] (gallery) (Suomi et al. 2008: 35), /f/ in “*faarao*” [fa:rao] (faraao) and /ʃ/ in “*shakki*” [ʃakki] (chess).

As with vowels, the quantity as a distinctive feature applies to Finnish consonants as well, as in “*tuki*” [tuki] (support), and “*tukki*” [tuk:i] (log) (Suomi et al. 2008: 3, examples mine). Interestingly, the only consonant sound lacking a similar grapheme presentation in Finnish (ibid.: 141) is the /ŋ/, called the voiced velar nasal, found in words such as “*kuningas*” [kuniŋ:as] (king). The sound is also found in the English language, interestingly in the same word “*king*” [kiŋ] (kuningas).

These single phonological “errors” may not create a significant problem regarding understanding, as the listener is usually able to decipher the meaning from the context (Sajavaara and Dufva 2001: 242). Nevertheless, and perhaps unsurprisingly, a continuous phonological error may contribute to intelligibility or comprehensibility (Jenkins 1995: 9). The treatment of minimal pairs as homophonous does not contrive,

of course, only to Finnish. For example, in General American, the words “*Mary*” [mæri] (Mary), “*Merry*” [meri] (iloinen), and “*Marry*” [mæri] (avioitua) are often monophones (McMahon 2016: 67). Again, their relevance to understanding may not create a significant problem as the context usually provides clues to the meaning. However, constant dealignment from the norm may have an effect. At a minimum, the awareness of the different phonological structures between the languages may abridge the possible gap in understanding.

## 2.3 Stress

McMahon (2016: 118) defines stress as a syllable, which is phonetically more prominent in reference to others. They continue that this prominence can be made discernible by three “subsidiary phonetic factors” (ibid.) which are higher pitch, syllable duration, and intensity. In addition to these, Roach (2009: 73) has proposed a fourth factor involving the stress prominence; quality. By this, Roach refers to the quality of the vowels in relation to the stressed ones. Yet, they state that from these four factors, the pitch and duration produce the “strongest effect”, the intensity and quality have a lesser role.

Concerning languages, McMahon (2016: 119) states that they can be divided to two classes depending on their stress patterns: fixed-stress languages and free-stressed ones. The Finnish language has fixed stress (Suomi et al. 2008: 22). This means that the word stress is invariably on the first syllable and is not a distinctive feature (ibid.: 39). In turn, English is not entirely a fixed-stress language nor is it a free-stressed one; deriving from Germanic languages with fixed-stress, English has also been influenced heavily by romance languages such as French, ending in between both Germanic and Romance (McMahon 2016: 120). In addition, stress is a distinctive feature in English. Depending on the stress (illustrated in the next example with capitalization), English distinguishes the otherwise same words “*IMport*” (noun) from “*imPORT*” (verb) (Roach 2009: 36), whereas Finnish does not. Moreover, English tends to have prominence on the final content word of an utterance, 90% of the time (Levis 2018: 156). If this is not realized, it may influence the comprehensibility of the listener (Hahn 2004). Moreover, as shown in a study by Connell et al. (2018), English speakers differentiate words and recognize them at least partially by relying on stress.



## 2.4 Intonation

Intonation refers to the variation of the pitch to convey meaning (Levis 2018: 150). For example, with intonation, an English speaker can distinguish between a statement by using a falling tone in the word “yes” and a question, using a rising tone with the same word “yes?” (Roach 2009: 121). In turn, neutral Finnish intonation can be described as a “smoothly descending pitch contour” (Suomi et al. 2008: 114). As such, the questions or statements are not distinguished in Finnish by intonation (Sajavaara and Dufva 2001: 251).

Intonation conveys various meanings in speech, but it does not necessarily affect intelligibility, as defined in this study but instead can affect comprehensibility and the functions of the message (Levis 2018: 171, 178, 182). The functions, in turn, may relate to the finality of the message, when spoken with falling intonation, or by fall-rise pattern, signaling uncertainty (Roach 2009: 142). It is to be noted that Roach refers to English intonation descriptions, which do not necessarily apply to Finnish intonation features directly.

Furthermore, in the interaction between non-native speakers of English Jenkins (1998) suggests that intonation is not a relevant feature regarding intelligibility. This can be seen in the word “*please*”, usually marking politeness but serving other functions such as pleading (Roach 2009: 152) if intonated differently. This area of linguistics, however, falls under the field of pragmatics and is not discussed thoroughly here.

The intonation functions presented at the beginning of this section appear both in Finnish as well as in English, at least partially serving the same functions in communication and with a possible effect on comprehensibility. As discussed, even if intonation has high importance in conveying meaning (Levis 2018), its relevance to intelligibility and comprehensibility needs further studies.

## 2.5 Rhythm

As discussed before in section 2.3, the English language has features from both fixed-stress and free-stressed languages. However, in English, the syllables that are stressed have a tendency to appear regularly, regardless of unstressed syllables (Roach 2009: 147). In these strong syllables, the vowels are in their strong form, whereas in the unstressed syllables the vowels are weak, appearing as either schwa or other central vowels (Cutler and Butterfield 1992: 218).

The rhythm may affect comprehensibility and intelligibility, as deviations of the expected rhythm pattern may increase the effort to understand the speaker and at least indirectly affect the intelligibility (Levis 2018). Furthermore, the concept of rhythm also includes pauses. In English, pauses are a key differentiation method for the listener to distinguish between several tonic groups, which are units that carry meaning (Romero-Trillo and Newell 2012). This can be illustrated with an example of the Oxford comma. For example, if in the phrase *“To my parents, Ayn Rand, and God”* the commas are understood as pauses in speech, the sentence can be interpreted differently depending on if the pauses are uttered as commas. Even if this has no influence on the lexical level of understanding, it still can have an effect on the pragmatic meaning of the sentence, and thus affect comprehensibility indirectly.

In summary, this chapter has reviewed key aspects and features of Finnish and English sound systems. Both languages have their distinctions and unique ways of realizing sounds, in addition to how they are distinguished. In the next chapter, I move on to review the studies and aspects considering intelligibility and comprehensibility.

### 3 ON INTELLIGIBILITY AND COMPREHENSIBILITY

In this chapter, I define the concepts and introduce the terminology used in the study. First, I explain the key terminology relevant to my study. Secondly, I review some of the studies that have researched intelligibility, comprehensibility, and accentedness. Thirdly, I explain the phenomena involved in learning pronunciation and present an aspect of L1 (first language) influence on perceiving L2 (second or foreign language). Finally, I briefly discuss their pedagogical implications.

#### 3.1 Key terminology

Regardless of the language, the features of pronunciation can be roughly divided into two interrelated parts, segmentals and suprasegmentals. As explained by Ogden (2009: 23), the segmental features are the segments of otherwise continuous speech, namely the vowels and consonants. They continue that in turn, the suprasegmental parts, or the prosodic features, relate to the larger speech units, such as word stress, intonation, and rhythm.

In this study, in one part I am studying what pronunciation features relate to intelligibility. Thus, when referring to intelligibility, I am referring to “the extent to which a speaker’s message is actually *understood* by a listener” (Munro and Derwing 1995a: 76, emphasis mine). Referring to the abovementioned *understood*, I refer to the lexical<sup>1</sup> level of understanding, which in this study corresponds to the ability to transcribe a heard utterance, word by word, orthographically. This can be viewed as the “index of speaker intelligibility” (Munro, Derwing, and Morton 2006: 111). Index, in

---

<sup>1</sup> By lexical level of understanding I mean that one can understand the words in a sentence “*The hollowed bird read three pigs*” separately and thus be able to transcribe them accordingly, even though one cannot understand the (pragmatic) meaning of the sentence.

this sense, means the lexical level of understanding. Munro and Derwing (1995a) are the researchers who distinguished the concepts of intelligibility and comprehensibility and were the ones who defined them. Since 1995, the definitions of these concepts have been mainly followed in the field and as such, I use them as well. In addition to this, I chose to use their definition as it is clear, precise, and can be operationalized for measurement with content word counting. The content words are words that carry semantic meaning, such as nouns, adjectives, and verbs, whereas function words are conjunctions, articles, or “particles”, which appear in high numbers without semantic meaning (Haspelmath 2001: 16539). Hence, only content words were considered in the intelligibility section of my study. However, as the definition of intelligibility is quite narrow in the sense of overall understanding, I complement this by examining another phenomenon: comprehensibility.

By *comprehensibility*, I mean the “[listener’s] effort involved in understanding” (Munro and Derwing 1995b, emphasis mine). In other words, comprehensibility means how difficult or easy it is to understand a heard phrase. In this study, I focus on the comprehension of the *listener* as I examine what features the listener pays attention to and considers salient for comprehensibility. However, the importance of the speaker’s features must also be taken into account and indeed, the listener and the speaker are inextricable. The features of the speaker affect the listener and *vice versa*. There are studies focusing more on the speaker (e.g. Jenkins 1998) and some that are impartial (e.g. Zielinski 2008), taking both into consideration.

In addition to the concepts of intelligibility and comprehensibility, a third factor, accent, defined as “the degree to which the pronunciation of an utterance sounds different from an expected production pattern” (Munro et al. 2006: 112), is also often considered in the field. Indeed, many studies incorporate comprehensibility and accentedness, or all three with intelligibility within the same study (e.g. Munro and Derwing 1995; Kang 2010). It has been found that these three concepts are interrelated yet they are not identical (Munro and Derwing 1997). Hence, they must be defined and studied as separate concepts.

However, the concepts of intelligibility and comprehensibility often intertwine and are not mutually exclusive. On the contrary, they are frequently inextricable. Among the wide variety of features that have been studied are such as familiarity with the speaker’s accent (Beinhoff 2014; Wilkerson 2013), fluency (Derwing and Rossiter 2003), primary stress (Hahn 2004; Low 2006), and lexical stress and vowel quality (Field 2005) and they have been all concluded to influence intelligibility and comprehensibility.

## 3.2 Studies on intelligibility

The intelligibility of speech has been studied for long and mostly by using a transcription task in which the study participants write down the heard utterances using the standard orthography (Munro et al. 2006; Zielinski 2008; Bent and Bradlow 2003; Burda et al. 2003). Other measures employed include true/false statement verification (Munro and Derwing 1995b), reaction time measuring (Hahn 2004), and acoustical measurements such as the effect of background noise and reverberation (Bradley 1986). Many of the studies have also investigated the concept of comprehensibility and/or accentedness in the same study with intelligibility (e.g. Munro and Derwing 1995a, 1997, 2009; Munro et al. 2006).

Curiously, regarding the speaker's accent, it has been found that the accent can be evaluated as "harsh" or "heavily accented" by the listener, but it does not necessarily correlate with intelligibility (Munro and Derwing 1997; Munro et al. 2006: 112). Hence, in this study, I left the concept of accentedness aside, and I concentrated on intelligibility and comprehensibility.

Some studies of intelligibility have narrowed their focus to a single word, such as in Field's (2005) research. As their data, Field used 12 disyllabic words and varied their stress patterns to see if they correlated with intelligibility. In essence, Field shifted the stress from its usual location to an unusual one to see if it had an effect on the intelligibility. Field found that if the stress was shifted rightward (e.g. to the last syllable on the word *vacATION*, normally stressed on the second syllable *vaCAtion*), it affected intelligibility more so than when the shift was to an unusually stressed syllable, leftward (e.g. to the first syllable on *VAcation*). Hence, the shift to a normally unstressed syllable, especially rightward, caused a decrease in intelligibility.

Field (2005) also reported that the effect of the stress shift affected similarly both the native listeners as well as the non-native listeners. As such, one of my research questions is set to investigate this particular aspect; do the native listeners and non-native listeners respond similarly or report similar features to be salient in terms of intelligibility. Moreover, aiming to follow Field's method, I use a similar stress-shift rightward dislocating the normal stress pattern in one study sample, to see if it has an effect on intelligibility.

The shifting of the stress and its effects on intelligibility has been also studied by Hahn (2004) but instead of the word level, they concentrated on the sentence level. Hahn found indications that the stress placement affects the recollection of the content favorably, as in emphasizing the content word of a sentence *this PAINTING is a fine piece of art* as opposed to emphasizing the auxiliaries *this painting is A fine piece of art*. Thus, emphasizing the content word made the phrases easier to remember, and favored the evaluation of the speaker positively. This significance of suprasegmentals

has also been found in several other studies (e.g. Isaacs and Trofimovich 2012; Saito and Saito 2016) discussed further in section 3.3.

In their study, Zielinski (2008) studied what features L1 speakers of English paid attention to when listening to L2 speech. Even though they utilized only three subjects, Zielinski found that all of them were attuned to the stress patterns of strong and weak forms. Essentially, strong and weak forms refer to the phenomenon that certain words are pronounced differently depending on several factors, such as their position within a sentence or their function (Roach 2009: 89). To exemplify, Roach utilizes the word “*that*” which in its strong form would be /ðæt/ and in its weak form /ðət/. In context, this could be seen in sentences such as “*That* is what she likes”, in its strong form, and “One hopes *that* she will”, in its weak form.

Zielinski also found that deviations in standard stressing (in line with the previously mentioned studies by Field 2005, and Hahn 2004) had the most effect on understanding, especially regarding segmentals on strong syllables. This, in turn, points out that the importance of segmental features needs to be taken into account in studies such as mine. As such, Saito et al. (2016) point out the importance of segmentals and word stress regarding intelligibility, yet they nevertheless advocate a more holistic approach to the matter, taking into account such concepts as fluency, grammar, and discourse, as they are all related to intelligibility and comprehensibility in complex ways.

In summary, the studies on intelligibility have been multifaceted and with different alignments and foci. In previous studies, the most notable effects in terms of intelligibility have been the shifting of the stress, both in word and sentence levels. As such, stress is one of the pronunciation features that I concentrate on in this study. In addition, the suprasegmentals have been found to influence intelligibility more so than the segmentals. However, this is not to say that segmentals can be disregarded, as their relationship as integral and inextricable parts of suprasegmentals is complex, as shown in the previous paragraph. Hence, I examine both of these overarching concepts in this study. For a more thorough review on these concepts and previous studies on intelligibility, see e.g. Levis (2018).

Next, I review some relevant studies that have focused on the second concept salient for my study, comprehensibility.

### **3.3 Studies on comprehensibility**

As intelligibility, the concept of comprehensibility has also been in focus in many previous studies. The somewhat broader definition of comprehensibility supplements and expands the concept of intelligibility.

One of the more recent studies was done by Tergujeff (2021), as they studied the relation between comprehensibility and accentedness. They utilized two study groups, Finnish and Finland-Swedish teenagers both as the English-speaking groups, and native English teenagers from the United Kingdom as listeners and evaluators. The native listeners evaluated the speakers with a 9-point scale for comprehensibility and accentedness. Curiously, Tergujeff found that the L1 speakers of Finnish in the Common European Framework of Reference for Languages at the levels B1-B2<sup>2</sup> were evaluated as “less comprehensible and more accented” than their Finland-Swedish peers on the same level. Tergujeff discusses the potential affecting factors to be such as “segmental inaccuracy, unusual rhythm, dysfluent pausing and slow speech rate” referring to numerous previous studies. Finally, they consider if there were potential errors in the preliminary assessment but did not find sufficient evidence to rule this as the *raison d’être* for the results.

In addition to Tergujeff, other studies (e.g. Munro and Derwing 1995b; Kang 2010) have examined the relation between comprehensibility and accentedness. For instance, in their study, Munro and Derwing (1995b) found that speech that was only partially intelligible took a longer time to evaluate and process, and thus contributed to comprehensibility, yet they found no correlation between the accent and comprehensibility. Regarding the L1 background, Munro et al. (2006) found that despite the listener’s background, the listeners evaluated the speakers’ comprehensibility similarly. This points out that L1 and L2 speakers, at least to some extent, evaluate listeners using the same features. Curiously, Munro et al. (2006) also found out that familiarity with the speaker accent did not aid the comprehensibility nor did the evaluation of the accent affect comprehensibility.

Regarding segmental features, there is a relevant concept called *functional load* (abbreviated as FL from now on) to be considered. For simplicity’s sake, FL and whether it is considered *high* or *low* can be defined in this paper as “the number of *minimal pairs* that [the phonemic contrast] serves to distinguish” (Brown 1988: 600, emphasis mine). In turn, *minimal pairs* refers to a change of a single phoneme, effectively changing the meaning of a word (Roach 2009: 52). For instance, the words “*bad*” [bæd] (paha) and “*bed*” [bed] (säinky) the phonemes /æ/ and /e/ function as a minimal pair, as they change the meaning of the word. Similarly, the phonemes /æ/ and /e/ can be said to bear a *high functional load* as they are frequent in English words, function similarly in all English native accents and have a high number of minimal pairs (Brown 1988: 602). On the opposite end, functioning as *low functional load* bearers in English, are phonemes such as /u/ and /ʊ/, forming minimal pairs in words such as “*pull*” [pʊl] (vetää) and “*pool*” [pu:l] (allas). These can be considered to bear a *low*

---

<sup>2</sup> More on the basis of CEFR evaluations: <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>.

*functional load* as they are not as frequent and have far fewer minimal pairs (Brown 1988: 603).

Regarding functional load and comprehensibility, Munro and Derwing (2006) studied the predictions about high FL and low FL made by Brown (1988) and Catford (1987). Their preliminary results confirmed that errors in high functional load sounds may influence comprehensibility negatively to a greater extent than the low functional sounds. In addition, other researchers have also tested Brown's predictions. A two-phase study by Suzukida and Saito (2021), examining Japanese speakers' English samples, and how they were evaluated by native English speakers, showed in both sections that high FL errors affected listeners' comprehensibility.

Another recent study by Foote and Trofimovich (2018) studied comprehensibility in relation to shared language benefit, with French, Mandarin, and Hindi speakers and listeners. Interestingly, they found that the shared language benefit affecting comprehensibility was greatest with the lower proficiency level speakers of Mandarin, as they were understood better by the listeners from the same background. Overall, the authors speculate that some segmental features and features relating to fluency may be relevant for all L2 listeners in terms of comprehensibility. However, they cannot state exactly what these features are, suggesting grammar, discourse and vocabulary as the potential ones.

Another study by Saito, Trofimovich, and Isaacs (2017) used extemporaneous speech description of a picture task with 40 French speakers of English. They utilized 60 native English speakers from Canada as raters. In line with previous studies, such as Munro et al. (2006) and Saito et al. (2016), their results implicated that accentedness does not necessarily correlate with comprehensibility and that the native speakers evaluate the speech similarly, relying on such features as segmental errors and word stress, in addition to pronunciation in overall.

Ambitious, all-encompassing approaches have also been executed in an effort to gain an in-depth understanding of the underlying concepts, such as Isaacs and Trofimovich's (2012) two-phase study which had 19 quantitative measurements from segmental and suprasegmental features, in addition to studying comprehensibility as another part of their study. They used scalar evaluations of the samples' comprehensibility made by the Canadian raters and contrasted them to the multiple features of speech, utilizing statistical methods. Their results implied that word stress was the most distinguishing factor regardless of the language speaker's level of comprehension. Despite this being much narrower a study, I too aim for a larger spectrum of features in search of the relevant ones that contribute to comprehensibility.

Furthermore, one of the major findings of the previous studies has been that the suprasegmentals seem to contribute to intelligibility and comprehensibility more than the segmentals (e.g. Kang 2010; Kang, Rubin, and Pickering 2010; Munro 1995; Munro



and Derwing 1995a; Hahn 2004; Derwing and Rossiter 2003). For instance, a study by Saito and Saito (2016) examined the effects of language instruction focusing on suprasegmentals on comprehensibility with Japanese learners of English, evaluated by North American speakers of English. They found that the training in suprasegmentals, namely in word stress, intonation, and rhythm, in addition to showing differences between the L1 and L2, enhanced comprehensibility. Nevertheless, especially the research concerning FL (Munro and Derwing 2006; Zielinski 2008) points out the relevance of segmental features at least considering how the native English speakers perceive and evaluate intelligibility and comprehensibility.

In summary, despite the extensive research, the features in relation to comprehensibility remain anything but unambiguous, and no feature has been shown conclusively to be the most salient one above others. However, in line with intelligibility, the previous research suggests word stress to be among the most relevant ones regarding comprehensibility. Hence, I examine its relevance in this study as well. In addition to stress deviations, also imprecisions in segmental realization, unexpected speech rhythm and pauses may all influence comprehensibility, among the high or low functional load of the words. All these features are taken into consideration in this study in an effort to contribute to their salience.

### **3.4 Shared language benefit**

Shared language benefit (abbreviated as SLB from now on) refers to an aspect that there would be a benefit to the mutual understanding between the listener and the speaker while conversing in their non-native language while sharing the same native language background (Bent and Bradlow 2003; Foote and Trofimovich 2018). More comprehensively defined, this benefit means that the “performance by non-native listeners or on nonnative speech exceeds that by native listeners or on native speech” (Hayes-Harb et al. 2008: 665).

Curiously, the studies have shown disputable and inconclusive results. As such, the L1 background does not necessarily affect the intelligibility or comprehensibility of the spoken L2 favorably, even if the listener is from the same L1 background as the speaker (Munro et al. 2006).

Moreover, both the NS (native speaker) and NNS (non-native speaker) listeners have responded similarly to the influence of the researched factors (Field 2005). This suggests that the concepts of intelligibility and comprehensibility are, *super omnia*, qualities of the speaker more so than the qualities of the listener.

In their study, Bent and Bradlow (2003) found that indeed there was SLB within the four listener groups that they used. They hypothesized that there is a so-called

*matched* benefit, meaning that the listener and the talker share the same native language background, and that the SLB may be due to the sharing of the sound structure.

In turn, a study conducted by Lima (2016) focused on the relation between comprehensibility and L1 background. Their study also examined accent and the effects of oral training. In their study, four Chinese speakers were evaluated by listeners from 14 different backgrounds. The results pointed out that accentedness and comprehensibility are not necessarily related, as the accent did not affect comprehensibility. The authors tested the participants twice, before and after guidance on speaking. Curiously, on two occasions the Chinese listeners evaluated the Chinese speakers more favorably before the guidance, yet after it, when they would have been supposedly evaluated as more comprehensible, there was no significant difference to be found anymore.

Finally, a study by Riney, Takagi, and Inutsuka (2005), examined what features non-native speakers (NNS) and native speakers (NS) of English paid attention to when making judgments about pronunciation. While their study was not focusing on SLB *per se*, it has some valuable insights on this matter as well. Their study showed that NS and NNS focused on different features when listening to the speakers and evaluating their accents. The NS from America focused more on segmental parts whereas the NNS from Japan used suprasegmental cues. As the authors point out, Japanese is a language wherein the pitch has meaning separating functions and thus may be the reason why the NNS used suprasegmentals over segmentals; intonation and pitch are suprasegmental features.

In summary, SLB is an interesting aspect that has inconclusive results. Some studies have found that indeed the L1 background may provide an additional benefit in understanding NNS spoken English, but this is contested by other studies. Hence, more studies, such as mine, are needed to examine this aspect. As such, one of my study questions explicitly addresses this phenomenon; do the L1 speakers of Finnish experience a shared language benefit when listening to their peers' English.

### **3.5 Pronunciation learning theories**

It is known that aspects not found in L1 may be hard to perceive in L2. This relates to the concept of cross-linguistic influence, defined as the effect of how the knowledge of one language influences learning or use in another (James 2012). In terms of segmentals, the studies (e.g. Escudero and Boersma 2004; Aoyama 2003) have shown that L1 has a role in how the segmental features are perceived in L2. An example demonstrating possible difficulties for the English speakers would be the quantity as a distinctive feature in Finnish, realized in words such as "tuli" [tuli] (fire), and "tuuli"

[tu:li] (wind) (Jarvis and Pavlenko 2008: 63). Contrariwise, the same difficulty could apply to Finnish speakers, for example in the words “grove” /gɹɔʊv/ (lehto), and “groove” /gɹu:v/ (syväne), which could be difficult for Finnish learners to perceive or pronounce due to the influence of the L1.

One of the reasons for the perceiving difficulties may be the phoneme-grapheme correspondence. In Finnish, the phone-grapheme correspondence is *shallow* (meaning that they correspond highly), whereas it is *deep* (the correspondence is low) in English (Lintunen 2015). In Finnish, almost every sound has a corresponding grapheme, for example in the word “äiti” [æiti] (mother). In contrast, English has a more complex writing system, and thus the speech sounds are not necessarily realized in graphemes. For instance, the examples of the abovementioned “grove” and “groove” could be pronounced as /krove/ and /kro:ve/, when following the Finnish shallow grapheme correspondence. This is an aspect that may create difficulties for the Finnish learner of English. However, the graphemes and their shallowness or depth are not the key focus of my study, but I mention them to provide background and understanding, as these perspectives illustrate how languages are perceived.

Pronunciation is a skill that encompasses both cognitive and motoric skills (Lintunen 2015). In effect, this means both how the speech is processed in the mind (perception) and how it is produced physically in the mouth and throat areas (motoric skills). There have been many hypotheses and models on how pronunciation is learned, and I review some of the essential ones here. A key concept in learning pronunciation is the speech learning model (SML) developed by James Flege, among his colleagues. The model is built from and upon several hypotheses, and hence, I review them first.

One of the hypotheses preceding SML is the Critical Period Hypothesis (abbreviated as CPH), introduced by Lenneberg (1967) and explained in a study by Snow and Hoefnagel-Höhle (1987) to mean that there is a certain age after which learning a native-like pronunciation becomes difficult. The authors place this age around puberty and write that its possible reason may be due to the brain development and changes in its plasticity. However, in their study, Snow and Hoefnagel-Höhle did not find sufficient evidence to support the hypothesis. In addition, other studies over the years (e.g. Flege 1998; Flege, MacKay, and Meador 1999) have been unable to find evidence in favor of CPH, yet there have been some that have (e.g. Patkowski 1990). What has been indisputable is that age relates to language learning. How exactly, has remained inconclusive.

Another hypothesis preceding the SML is the Contrastive Analysis Hypothesis (CA or CAH), first introduced by Robert Lado in 1957. In its strong form, the CAH presents that by contrasting two languages, and their grammatical and phonological differences, one can predict and describe what kinds of difficulties will arise in

learning the L2 (Lado 1957). In essence, this means that the greater the difference between the L1 and L2 phonology or grammar, the harder it can be to produce. Lado (ibid.: 61) exemplifies this with a comparison of English and Spanish sound systems. They explain that in English, a question such as “*Are you a student?*” is marked with a rising intonation towards the end of the phrase, as opposed to falling intonation signifying a statement. They explain that in addition to changes in intonation, there are grammatical changes, as the word order changes from a question to a statement “*You are a student.*” Lado continues that is in reference to Spanish, where there is no difference in the word order, as the word order in “*Usted es un estudiante?*” remains the same both as a question and as a statement, only the intonation changes similarly as in English. Finally, Lado concludes that from this example one can assume that the word order in English creates problems for the Spanish speakers, and the intonation (without the grammatical changes) creates problems for the English speakers.

However, the strong form of the CAH as a predictor of rising problems has been disputed from early on. For instance, a study by Whitman and Jackson (1972) examining English syntax with 2500 Japanese students found that the indications made by CAH are not sufficient as predictors of problems. Another issue with CAH has been the impossibility of universal theories dealing with semantic, syntactic, and phonological issues, regarding all languages (Wardhaugh 1970). However, Wardhaugh continues that in its weak form, as an explainer of the possible errors due to one language effect on another, instead of a model for predictions, it remains plausible.

Finally, before the SML, I present the Markedness Differential Hypothesis (MDH), coined by Eric Eckman in 1977. As explained by Benson (1986), it builds on CAH, but adds a typological variable; a phonologically more complex sound is called *marked* and a less complex one is called *unmarked*, and this *markedness* predicts how difficult a feature is for an L2 learner. For instance, universally speaking, vowels (e.g. a, e, i, o) can be seen as the *least marked ones* in their word-final positions (as in words such as *academia*, *age*, *bikini*, *go*) when compared to other sounds, particularly to voiced obstruents in the same word-final positions (e.g. d, g, v, z) which are considered as the *most marked ones* (as in *complicated*, *waterboarding*, *Kalashnikov*, *tchervonetz*) (Eckman 1985: 294, examples mine). Eckman continues that in addition, the existence of a marked feature in a language entails the existence of an unmarked feature in the same language, but not *vice versa*. Hence, following the MDH, one could predict that the previously mentioned d, g, v, and z sounds are much harder to learn (in the word-final position at least) than the vowels. The research has found some evidence supporting the MDH regarding its predictions on the abovementioned obstruents being more difficult than vowels (e.g. Carlisle 1991) and considering the word order (e.g. Jin 2008).

These hypotheses, among others, have each influenced the speech language model (SML), which was coined by James Flege in 1995. The model presents several hypotheses on how pronunciation is learned. In one of them, Flege (1995) states that the sounds in the L2 are related in the listener's perception to a sound that is closest to it in L1. For example, an English speaker could associate the Finnish /r/, which is realized as a trilled alveolar (see section 2.2), with the English /r/, which is realized in the General American variant as an alveolar approximant. Contrariwise, the sounds that could be predicted to cause trouble for a Finnish learner of English might be, for example, the /b/ and /g, namely those that are not present in the native words of the Finnish language (Kuronen 2019: 60).

In turn, Flege (1995) continues with another hypothesis, stating that the greater the difference between the sounds in L1 and L2, the greater the chance for the listener to establish a new sound *category* for it within their mind. By category, Flege means that the speech sounds are contrasted to the L1 sounds of the listener, and consequently can be divided into three categories: new, identical, and similar sounds. They explain that new ones are the sounds not found on the listener's L1, identical ones are those found and realized similarly in L1 and L2, and the similar sounds are the ones similar but not identical to the L1 sounds. Flege continues that this hypothesis presupposes that the listener has indeed perceived a difference within those sounds. They write that chance for this to happen increases as the difference between the sounds increases. Finally, Flege argues that the reason for the L2 learner's inability to produce the L1 sounds has more to do with their perceptive skills than with their motoric, physiological, abilities.

To put the previous paragraph in context, an example could be the differences in distinguishing the phonemes /v/ and /w/ in Finnish and English (see section 2 on phonemes). In English they are treated as different phonemes and in Finnish as allophones, meaning that they are expressions of the same phoneme. This, in terms of SML, could mean that the difference between the /v/ and /w/ are hard for the Finnish learner of English to perceive, due to the influence of the L1 (see section 2.2 for further discussion on the matter).

In 2021, the SLM was revised, and its current form is entitled as Speech Language Model Revised (SML-r). The key changes are that the revised model does not have early or late learners in focus as the older model had, it does not concentrate on skilled L2 learners, and the hypothesis about the limitations of L2 segmental productions has been replaced by a hypothesis that the production and perception coexist and evolve together (Flege and Bohn 2021).

In reference to the sound categories of new, identical, and similar, Lintunen (2015) writes that the hardest to learn can be the similar ones, and not the new ones, as could be intuitively thought. Lintunen base their argument on the *Perceptual Assimilation*

*Model* (PAM), a model on how the L2 sounds are perceived in relation to L1, by Catherine Best. In short, Best (1995) argues that there is assimilation in the speech sounds that are not identical but similar if there is not a category for the perceived new sound. Hence, as explained by Lintunen (2015), a Finnish person may assimilate the /z/ to /s/ sound, as there is no distinction between them in Finnish phonology. They continue that this may not create a problem in terms of intelligibility or comprehensibility if the sound is not confused by the listener to a different sound. To exemplify, this would mean that the pronunciation of the word “zeal” [zi:l] (*innokkuus*) as “seal” [si:l] (*tunnus*) may be understood through context as it is intended but separately could be mistaken as another word.

Several, partially overlapping hypotheses and models have been developed throughout the years on how pronunciation is learned. The aspect needs further research on determining how, but it can be said that L1 has a key role in perceiving and categorizing the sounds in L2. Considering the speech samples that I use in this study (see section 4.2) I predict that the sounds that are not present in the Finnish phonological system may cause peculiarities in the speakers’ pronunciation.

### **3.6 Pedagogical considerations**

The research findings presented in sections 3.2 and 3.3 imply that emphasizing the prosodic features over the segmental ones could yield better results in pronunciation if indeed the profound meaning of all communication is to become understood. Despite this, the study materials on pronunciation tend to focus almost exclusively on the production of segmentals (Levis 2018: 15). The discrepancy raises thoughts about the hierarchy of teaching pronunciation marking segmentals as a priority. Fortunately, in the Finnish National Core Curriculum (NCC) for Basic Education (2014: 220) the features of word stress, sentence stress, rhythm, and intonation are mentioned as consequential in relation to pronunciation taught in the elementary school (grades 3–6). In contrast, in the NCC (*ibid.*: 349), there is a mention of “conveyance of the message” as one of the most important objects in junior high school (grades 7–9). In addition, “good pronunciation” is one of the targets, as is the “appliance of numerous ground rules for the pronunciation” (*ibid.*: 351–352). Nevertheless, how these objects would be attained and what are the features of pronunciation relevant to reaching the ambiguous goals are not explicit, except in the passage regarding elementary school, which notes the relevance of prosodic features.

Another guideline for teaching in Finland is the Common European Framework of Reference for Languages, CEFR. Regarding phonology, the CEFR notes the

importance of intelligibility<sup>3</sup> in oral communication and prosodic features as relevant to “convey meaning in an *increasingly precise* manner” (Council of Europe 2020: 133, emphasis mine). Segmentals are mentioned first among the features contributing to intelligibility. The importance of suprasegmentals is also highlighted in numerous occasions within CEFR (ibid.: 64, 130, 133, 164). Indeed, the research has found their salience yet the importance of segmentals in terms of FL and intelligibility cannot be left aside. Thus, the importance of one over another is difficult to assume, as the concepts are intertwined and in connected speech inextricable.

Based on the findings concerning FL, by Munro and Derwing (2006) and Suzukida and Saito (2021) discussed in section 3.3, one could argue that it is more useful to concentrate on the high FL pairs in terms of being understood, than on the low FL. However, the abovementioned studies have focused on how the words are interpreted by native English speakers and do not necessarily apply to English as a lingua franca, or English as L2 speakers. Further studies on non-native English speakers, such as mine, are needed.

The SML by Flege (1995) and their advocacy of *perception* on learning pronunciation entails that pedagogically the language learners should be taught to distinguish different sounds as a priority, rather than to concentrate on motoric realizations. However, using repetitive exercises (drills), may also influence perception. As the matters are complex, it can be said that perception and motoric skills influence one another. Finally, L1 effects on learning and perceiving L2 have seen disputable results in previous studies. L1 certainly affects the L2, but in complex ways that need to be studied more.

In summary, features relating to the overall concepts of intelligibility and comprehensibility have been studied extensively but not exhaustively. The concepts remain multifaceted and may be approached in a myriad of views and ways. However, for future language teachers, in terms of being understood, these phenomena are of utmost importance both from didactic and pedagogical perspectives.

---

<sup>3</sup> Intelligibility in CEFR (2020: 133) is defined as “how much effort is required from the interlocutor to decode the speaker’s message”, thus being equivalent to the concept “comprehensibility” as defined in my study.

## **4 THE PRESENT STUDY**

In this chapter, I explain my research methodology. First, I introduce the study aim, the research questions and explain the listener groups and data collection. Next, I present the rating task, mixed methods approach, and the analysis methods. Then, I explain the research strategy and epistemology, and finally, I conclude the chapter with ethical considerations.

### **4.1 Aim and research questions**

My study aim is to seek out which factors of pronunciation affect intelligibility and comprehensibility. I hypothesized that there would be a difference between the listener groups, as English and Finnish differ in their phonology (see chapter 2) and as L1 affects the perceiving of L2 (see section 3.5). My research questions are:

- 1) What segmental and suprasegmental features affect intelligibility and comprehensibility in L1 Finnish speakers' L2 English?
- 2) Are there differences in the abovementioned features between native listeners of English, and non-native listeners who share the speakers' L1, and if so, what are they?
- 3) Do L1 speakers of Finnish experience a shared language benefit when listening to their peers' English?

I developed the questions in an effort to gain an in-depth view on the matter, and to examine whether natives and non-natives have differences in perceiving pronunciation. Moreover, the shared language benefit has received mixed results in previous studies and I wanted to contribute to that aspect, whether in support or dispute.



## 4.2 The listener groups and speech samples

In this study, I had two listener groups, one consisting of native Finnish speakers ( $n = 31$ ), and the second consisting of native English speakers ( $n = 36$ ). No limitations about age or gender were applied, and I considered only the nativeness (either a Finnish native who can speak English, or a native English speaker) as a crucial factor. To answer the research questions, a rating task was arranged and developed utilizing the Question Pro online platform (see section 4.3). The participants were obtained by distributing the rating task through university mailing lists and on social media networks, such as Facebook groups.

I used readily available speech samples of Finnish spoken English, derived from the International Dialects of English Archive (IDEA<sup>4</sup>). The IDEA is an archive of English language samples with more than 130 countries included. I chose to use the ready-made samples, as the speech samples were high quality and the text (see Appendix 2.) utilized was made by researchers, aiming to include an as wide a variety of phonological features as possible for any researcher to consider. I used three speakers' samples, incorporating six short phrases (samples 1–6) in the intelligibility section, and six longer ones (samples 7–12) in the comprehensibility section. The speakers' level of English was evaluated by me, in an addition to two Finnish English teachers and highly accomplished researchers, to correspond to levels B1, B2, and C1. All three of us used the scale presented by the Council of Europe (2020: 135) concerning aspects of overall phonological control, sound articulation, and prosodic features. We deemed these three aspects sufficient for the evaluation criteria for the purposes of this study.

I chose these particular samples as they contained a plentitude of both segmental and suprasegmental features which were the core focus of my study. To verify that the samples were evaluated correctly, a control sample by a native speaker of General American variety of English was incorporated into both sections of intelligibility and comprehensibility. This was also done to increase the study's reliability and validity.

The transcription of the full sample text, entitled "Comma Gets a Cure", can be found in Appendix 2. and the individual sample transcriptions concerning the intelligibility section can be found in Appendix 3., and the transcriptions for the sample sentences used in the comprehensibility section can be found in Appendix 4.

Next, I present briefly the 12 samples and their content:

---

<sup>4</sup> For the open-access website, see <https://www.dialectsarchive.com/>.

- Sample 1 incorporated a phrase “Sarah Perry (was a) veterinary nurse”. I did not consider the words within brackets as content words and thus I left them unaccounted. The word “veterinary” usually has a word stress on the first syllable, but here the word stress was shifted to the second syllable. The results of this word stress shift are discussed in sections 5 and 6.
- Sample 2 was a native speaker control sample, stating “Even so, (on) her first morning, she felt stressed”, without any deviations in pronunciation.
- Sample 3 stated “Her efforts (were not) futile” with attention to segmental deviations.
- Sample 4 had a phrase “She ate (a) bowl (of) porridge”, with a focus again on segmental peculiarities.
- Sample 5 contained a phrase “When she got there, there (was a) woman (with a) goose”. This sample had a peculiarity in sentence stress, with both of the words “there” receiving an emphasis in terms of intensity, intonation, and duration, leading to “When she got THERE, THERE was a woman with a goose”.
- Sample 6 read “Either five (or) six times (the) cost (of) penicillin”, wherein the intonation and rhythm of the speaker were very even throughout the whole sentence.
- Samples 7 to 12 were longer passages used to measure comprehensibility from the same speakers as in samples 1–6.

### 4.3 The rating task

For data collection purposes, I opted to use an internet-based survey platform. However, as my set of questions and integrated audio samples was not a survey *per se*, I refer to it as a rating task from now on. The advantages of an internet questionnaire are its speed in gaining answers, ease of distribution to the target groups, and the data being ready for exporting to Excel, without the preliminary need for literation (Heikkilä 2014: 66). Consequently, these were the reasons for choosing to use the internet-based form.

I collected the data utilizing Question Pro<sup>5</sup>, as this survey platform allowed the integration of audio samples, which was the most crucial factor when deciding which one to use. Other reasons were its ease of use, clear user interface when designing the questions, and the ability to distribute the rating task online.

---

<sup>5</sup> The site can be accessed via [www.questionpro.com](http://www.questionpro.com).

The rating task consisted of three parts. First, there was the intelligibility section with six questions, of which one was a control sample by an English native speaker. All the questions incorporated a short audio sample, ranging from 2 to 4 seconds, or from 3 to 8 content words (cf. Beinhoff 2014), wherein the listener was asked to write down what they heard in the audio sample using standard orthography.

Secondly, the comprehensibility section consisted of six questions of which one was a control sample. In this part, the samples were longer, from 17 to 24 seconds in length (cf. Tergujeff 2021) and their content words were uncounted. Each sample in this section entailed three questions. First, the participant was asked to evaluate the sample's comprehensibility with a slider scale, ranging from 0 to 100 (see Appendix 1. for details). This was followed by seven additional questions wherein the participant was asked to evaluate how much a certain feature affected their comprehensibility using again a slider scale, ranging from 0 to 100. Zero, in this study, meant that a feature, such as speech sounds, was not perceived as difficult at all to understand. In turn, 100 meant that this feature was very difficult to understand. Not answering meant neither agreeing nor disagreeing.

Finally, with every sample in the comprehensibility section, an open-ended question was provided for additional comments. As described by Anders (2012: 70), the open-ended questions give the research participants a chance to broaden their numerical answers and point out issues that may have not been considered by the researcher.

The final section of the questionnaire was for background information, which included age, mother tongue, and other languages spoken. I put this section as the last one since Valli (2018) has stated that the research participant motivation may decline towards the end of the questionnaire, and putting the background section last may help. Furthermore, I designed the questionnaire according to the guidelines of Dörnyei (2007) and Valli (2018) considering the language, word choices, neutrality, visual form, and layout.

I piloted the rating task with layman respondents and I initially planned the comprehensibility section to use Likert-scale answers but decided to use the 0-100 slider scale instead. This was because the sliding scale has been used in many studies before (e.g. Flege, Munro, and MacKay 1995; Major 1987), it is user-friendly, and clear when seeking numerical answers. As such, Jesney (2004: 3) describes its advantages to be much finer details in answers, when compared to the Likert scale. However, Jesney also points out that these details are not necessarily noted by the participants. Nevertheless, the reasons for opting to use the slider scale were the ease of quantification and statistical comparison of the answers in the latter analysis, and the ease of use for the research participants.

During the pilot phase, I was able to consider and receive feedback on listener fatigue and modified the rating task accordingly. As such, Dörnyei (2007: 110) advocates keeping the questionnaire within a 30-minute limit. This limit is backed up by Alanen (2011: 152). On the other hand, slightly longer durations have been used, such as in a study by Tergujeff (2021), who used a 45-minute limit with a small break in the middle. Keeping these factors in mind, I aimed for a 15-minute completion time. I wanted to keep the questionnaire as short as possible, since it included audio samples which, I speculated, would require more attention than mere text-based questions. As such, the average time to answer the rating task was 14 minutes, which was within my intended 15-minute limit.

#### **4.4 Mixed methods**

Mixed methods is a study that utilizes and combines quantitative and qualitative methods (Dörnyei 2007: 163). This approach has been used in the field of linguistics previously (e.g. Isaacs 2013; Derwing and Munro 2009; Tergujeff 2021) and is utilized in an effort to gain a fuller understanding of the phenomena. Combining different methods is called *triangulation*, which is used to increase validity, understanding, developing and paving way for further studies, and the increase the breadth of the study (Hesse-Biber 2010). Triangulation in my study realizes through conjoining quantitative and qualitative methods in both data collection methods as in the subsequent analysis.

Moreover, I chose mixed methods in an effort to gain a deeper and more thorough understanding of the study phenomena than the previous studies have been able to reach. Thus, I wanted to incorporate the study participants' own experience into the data and effectively combine quantitative and qualitative methods both in data collection and analysis.

Furthermore, I chose to combine different methods in order to increase the validity and understanding of the studied phenomena. I used three different types of methods in the rating task, which were content word counting and analyzing the transcriptions in the intelligibility section, semantic differential slider scale in the comprehensibility section, and open comments in the comprehensibility section. In addition to these, the methods of analysis were mixed with t-testing the word count and slider scales, and using theory-driven content analysis with the open comments and in intelligibility section transcriptions.

## 4.5 Statistical testing and the t-test

I chose to analyze the quantifiable results with an unpaired, two-tailed t-test, whose position within the various statistical test options is illustrated in Table 1. The unpaired t-test is used to compare two individual samples, and their possible statistical difference, assuming equal variances and normal distribution.

Subsequently, this was the case in my study. The normal distribution means that the mean and standard deviation act similarly in both ends of the distribution spectrum (Singh 2007: 95). This important aspect was determined both visually and using the Shapiro-Wilks' *W* test, which is "one of the best tests for normality" (Singh 2007: 100). Furthermore, Singh describes that two-tailed test tests if the mean of the samples is equal. In other words, the presumption is that the listener groups differ from each other, but not how or which one (Hypoteesien testaus n.d.).

However, it is to be noted that the process did not take into account the randomization of the observational units. This means that the principles of probability sampling were not met and this aspect needs to be considered when referring to the samples (Valli and Perkkilä 2018). Hence, my data consists of *non-probability samples*, which means that their generalizability to the general population is not as valid as it would be when using *probability samples*, which in turn means that the observational units have a probability above zero to be selected for the research (Otos ja otantamenetelmät n.d.). Despite this, the Shapiro-Wilks *W* test, in addition to the visual observation of the data diagrams, showed that the data is indeed valid in terms of using parametric methods for the analysis, and hence I opted to use t-testing as planned.

Table 1: Exemplification of parametrical and non-parametrical tests and depending on the purpose and listener groups. Based on Singh (2007: 162).

| Purpose                                  | Parametric tests        | Non-parametric tests       |
|--|-------------------------|----------------------------|
| To compare single group to a value       | One-sample t-test       | Wilcoxon, Chi-square       |
| To compare two unpaired groups           | <b>Unpaired t-test</b>  | Mann-Whitney, Fisher's     |
| To compare two paired groups             | Paired t-test           | Wilcoxon, McNemar's        |
| To compare three or more unpaired groups | One-way ANOVA           | Kruskal-Wallis, Chi-square |
| To compare three or more paired groups   | Repeated-measures ANOVA | Friedman, Cochran Q        |

First, after gaining enough research participants, I exported the rating task data from Question Pro to the Excel -software. Next, I counted the content words in responses and compared them to the actual number in the audio samples (see Appendix 2.). The process is depicted in Figure 8.

I repeated the preceding process with all the answers, samples 1 to 6, and they were analyzed with a t-test, in aggregate, as reported in section 5. The samples 7 to 12 were about comprehensibility, using a slider scale ranging from 0 to 100. These numbers were also combined and analyzed similarly with the t-test.

Considering the predicted results, I hypothesized that there is a difference in the perception of the intelligibility and comprehensibility between the listener groups of L1 English speakers and L2 English speakers sharing the speakers’ L1. In other words, are there differences between the study groups that cannot be based on chance. I based my hypothesis on the previous studies, discussed in chapter 3.

| W   | X   |
|---|---|
| Q1.Listen to the audio file and write down th | Q1. F4. Content Word Count = 4. Sarah Perry (was a) veterinary nurse. |
| Sir Perry was a returning nurse.              | 2   |
| Sir Perry was a returnery nurse               | 2   |
| Sir Perry is a nurse                          | 2   |
| Sir Perry was a veteran nurse                 | 2   |
| Sir terry must return a nurse                 | 1   |
| Sir Perry must a returner a nurse             | 2   |
| Sir Perry was a veterinary nurse              | 3   |
| Sir Perry was a retinary nurse                | 2   |
| ...was a... nurse                             | 1   |
| Sir Perry was a — nurse                       | 2   |
| Sir Perri must over turn nurse                | 2   |
| Sir Perry was a returnary nurse               | 2   |
| Sir perry was a veterinary nurse.             | 3   |

Figure 8: The content word counting in the intelligibility section. The study participants’ answers are on the left under W, and the transcription of the phrase and the counted numbers are on the right, under X.

## 4.6 Content analysis

As I wanted to gain further understanding of the rating task answers, I analyzed the open comments utilizing theory-driven content analysis. Overall, I followed the guidelines of the content analysis process (Eskola 2018; Eskola and Suoranta 1998; Vuori n.d. -a; Tuomi and Sarajärvi 2018) and consulted a step-by-step process in thematic analysis (Braun and Clarke 2006). I combined my method using these sources and divided the analysis into four steps.

The steps were familiarization with the data and creating the coding frame, searching for explicitly mentioned features and reducing the content, searching for

implicitly mentioned features and reducing the content, and clustering the features into categories and dividing the features depending on if they helped or hindered the comprehension of the listener.

As explained by Eskola (2018), it is important to familiarize oneself with the data before the subsequent steps. Consequently, this was my first step. As I exported the comments from Question Pro to an Excel file, I was able to avoid the iteration as the comments were already in written form. I read the comments several times and continued to revisit them in the steps that followed.

Eskola (2018) writes that after the familiarization with the data, there are at least two options to continue: inductive, content-based approach, wherein the whole data is attempted to perceive without an underlying theory<sup>6</sup>, and a second approach, which is building a coding frame guided by a theory. As my study focused on how the features of speech affect the intelligibility and comprehensibility, I chose the segmental and suprasegmental features as my coding reference and followed the second approach.

The second step was to search for explicit mentions of features and reduce the original comments into concise forms. For example, one comment stated that “overall it was understandable but I had to listen very carefully to get it as some words were not understandable”. I compressed this into “Some words were not understandable and very careful listening was needed”. I followed this through all the comments.

The third step was reading the comments seeking out implicitly expressed features. One comment stated that “I understand the audio file in the previous exercise much better, now that I have heard the context. It became easier to ‘fill in the gaps’ regarding the words I missed the first time”. As this comment did not state any relevant features at the semantic level (there was no mention of segmental or suprasegmental features explicitly), I moved on to the latent level, seeking if some relevant information was stated implicitly. Even though there were none about pronunciation features, I interpreted that the *context* helped to “fill in” the missing words. Thus, I compressed the comment to “The context helped to ‘fill in’ the missing words”. I categorized it for later consideration under “context”. Again, I followed this through all the comments.

The fourth, and consequently the last step was to determine whether the features helped or impeded comprehension. I read through the reductions and labeled them into “helping” and “impeding”. Lastly, I went through all the comments again in order to infer if they mentioned evaluations on the overall comprehensibility. The process and the coding frame are illustrated in Figure 9 on the next page.

---

<sup>6</sup> Similarly to Braun and Clarke (2006), Eskola (2018) states that the much seen “emerging” of points of interest from the data is effectively impossible, as is the inductive approach if taken literally since the researcher themselves are the ones who choose the points from the data and ultimately they are affected by some theories on how they choose them.

| BE  | BF   | BG               | BH              | BI               |
|---|--|------------------|-----------------|------------------|
| Q7. F2. Open comments   | Q7. F2. Reduction of the comment   | Q7. F2. Category | Q7. F2. Helping | Q7. F2. Impeding |
| I understand the audio file in the previous exercise (about penicillin) much better, now that I have heard the context. It became easier to 'fill in the gaps' regarding the words I missed the first time. | Context helped to "fill the gaps" regarding the missed words in intelligibility section. | Context          | Context         |                  |

Figure 9: The initial coding frame in Excel software. Under the title “BE” is the original comment and under “BF” its reduced form. Under “BG” is the initial category and subsequently “BH” and “BI” show whether the feature depicted was helping or impeding to the listener.

Initially, there were 156 open comments altogether from both study groups, but after the initial steps of familiarization, I found that the majority of them were not useful for the purposes of this research. First, I carefully examined every comment and removed the ones that were clearly irrelevant, such as “no”, “nope”, “Mika Häkkinen” or “she just sounds depressed or something”. After this, the subsequent steps described left 54 open comments for the analysis.

Hence, after all the steps, I had two categories, and these were “features that helped comprehension” and “features that impeded comprehension”. This last stage of the coding frame is depicted in Figure 10, found below. The appearance of the categories is reported in detail in section 5 and their relevance is discussed in section 6. Next, I will discuss my research strategy and epistemology.

| Sample number, speaker, and <i>n</i> | Features that aided comprehensibility              | Features that impeded comprehensibility                                |
|--------------------------------------|--|--|
| S7. F2, <i>n</i> = 12                | Context  | Speech sounds, words, rhythm   |
| S8. F3, <i>n</i> = 8                 | Context, speech sounds, words, word-stress, rhythm | Words, word-stress, sentence-stress, intonation, rhythm                |
| S9. F1, <i>n</i> = 12                | Context, word-stress, sentence-stress, intonation  | Speech sounds, words, word-stress, rhythm                              |
| S10. F2, <i>n</i> = 6                | Context, word-stress, rhythm                       | Speech sounds, words, intonation, rhythm                               |
| S11. F1, <i>n</i> = 9                | Context, intonation, rhythm                        | Speech sounds, words, word-stress, sentence-stress, intonation, rhythm |
| S12. CS1, <i>n</i> = 3               | Speech sounds, words, intonation, rhythm           | Words, rhythm  |
| Additional comments, <i>n</i> = 4    | Intonation, rhythm                                 | Speech sounds, rhythm  |

Figure 10: The final coding frame. Depicted are the sample number, the speaker code and *n*, in addition to the division between helping and impeding features.



## 4.7 Research strategy and epistemology

Different approaches to knowledge and how it is viewed are central parts of research (Philosophy of Science n.d.). Hence, the researcher's stance needs to be made clear whether they see their study, for instance in the light of positivism, utilizing only "pure logic and pure mathematics" (Feigl 2021), or whether they favor interpretivism, emphasizing the role of interpretation in "individual human action...or the context of the wider culture" (O'Reilly 2009: 123). In addition, there is a multitude of philosophical schools among these two opposites.

My study is an empirical study. Empirical study means that the study is based on observations, analyzing, and measuring, and the concrete data collected is in the focus of this study (Empirical Research n.d.). Hence, empirical study differs from the *theoretical study*, not utilizing research objects, but focusing on abstract theories and concepts (Theoretical Research n.d.).

Studies can be furthermore divided, for instance into *experimental*, studying the effect of a procedure or a substance in a controlled environment, *descriptive*, seeking answers to *what, who, what kind of, when, where* with broad data, or *causal*, seeking answers to causal relations (Heikkilä 2014: 13–14). My study can be labeled as a *descriptive* study, as I am trying to find the features of pronunciation contributing to intelligibility and comprehensibility, effectively seeking answers to the abovementioned descriptive questions.

Regarding epistemology, I approach knowledge within this study through empiricism and pragmatism. Empiricism views knowledge as something that is gained through experience, as opposed to rationalism, which in turn approaches knowledge through logical deduction from *a priori* (based on abstract knowledge) premises and innate datum (Markie and Folescu 2021). In turn, pragmatism views the world and its viewer as inextricable matters, and approaches knowledge as a derivation from *a posteriori* (based on experience) praxis, emphasizing the practicality of knowledge (Legg and Hookway 2021).

Finally, I view this study through *methodological eclecticism*, which means that I approach my research questions through methods chosen for the sake of practicality, and as such, the ontological roots are not important (Tuomi and Sarajärvi 2018). The authors write that this eclecticism is opposed to *fundamentalism*, which would entail an approach that is first and foremost selected on the basis of how reality is viewed, e.g. positivism mentioned at the beginning of this section.

## 4.8 Ethical considerations

According to Vuori (n.d. -b), responsible conduct of research (RCR) means honesty, and metical and careful approach in all stages of the research. Furthermore, the Finnish National Board on Research Integrity TENK has formed ethical principles for all research in Finland, effectively advocating the abovementioned RCR in addition to taking into account the research participants' rights according to Finnish Constitution, possible risks involved, voluntary participation, clear information about the research, anonymity, management of the personal data and consent to participate (Finnish National Board on Research Integrity TENK 2019). Similar ethical considerations are described by Dörnyei (2007: 67), following the American Education Research Association guidelines.

Considering the abovementioned, on page one of my rating task I explained the reason and background of my study to the participant, what sections the rating task includes, how it is conducted and the approximate time it takes to complete. In addition, I stated that the participation is completely voluntary and the participant can withdraw at any point. Furthermore, I stated that the responses and the data will be strictly confidential and the participants will remain anonymous. Finally, I gave my phone number and email address, should the participant have any questions concerning the rating task. As such, the participants were referred to only by their listener group abbreviations, and no individual information from which an individual could be identified was used or collected. The rating task raw data was downloaded from the secure server of Question Pro platform to a single, password-protected computer and the data was deleted upon the finalization of this thesis.

The consent to use the participant answers in the study was asked in the final section of the rating task by a statement that said, "by clicking 'Done' I give my permission to use this data anonymously in the research" (Appendix 1.). Finally, regarding the use of the copyrighted speech samples, a special license was bought from the International Dialects of English Archive which included the use of the audio samples as well as the transcription for the duration of the rating task (Appendix 2.).

## 5 RESULTS

In this section of my thesis, I present the study results. First, I explain the results in the intelligibility section, and secondly, I review the results in the comprehensibility section and tie the qualitative data to the quantitative.

The average time to complete the rating task was 14 minutes in both listener groups and the total number of participants who completed the task was 67, with an additional 71 dropouts, resulting in a 48.55% completion rate. The questionnaire was open from early December 2021 to late January 2022.

Two listener groups were examined in my research, Finnish speakers of English ( $n = 31$ ) which will be referred to as NNS (non-native speakers of English), and native English speakers ( $n = 36$ ) who are referred to as NS (native speakers).

### 5.1 Intelligibility

Intelligibility section results showed that for the NS group, the samples were more intelligible than they were for the NNS group. The results, depicted in Table 2, showed a statistically significant difference ( $p = .017$ ) between the groups' responses (NS  $n = 186$ , NNS  $n = 216$ ). The NS group had a higher mean percentage (79%) in the correct content word count than the NNS group (71%). The possible reasons and implications are discussed in section 6.

Table 2: The results concerning the intelligibility section. The table presents the studied feature, means for both listener groups in percentages, t-value, and p-value, respectively. The t-test p-value \*  $<.05$ .

| Feature         | NS mean | NNS mean | t-value | p-value |
|-----------------|---------|----------|---------|---------|
| Intelligibility | 79      | 71       | 2.389   | .017*   |

In line with the quantitative results, the qualitative analysis of the intelligibility section showed that all the samples were more intelligible for the NS group. In addition, the segmentals and the word stress seemed to have the strongest effect on misunderstanding in both listener groups. In turn, intonation, rhythm, or sentence stress seemed to have no noticeable effect on intelligibility.

The phrase used in sample 3, “Her efforts were not futile”, was the least intelligible of all the six samples in both groups. The NS group understood 37% of the content words in this sample and the NNS group 26%. Here the most misunderstood words were the first two: “her” and “efforts” which were heard as “hey...”, “therefore...”, “herefore”, “hey fort...” “here fort” or similar, in both listener groups, showing that segmentals and weak and strong vowel assimilation (discussed further in section 6) seem to have a role regarding intelligibility.

The manipulated word stress shift rightward in the word “veterinary” in sample 1, affected its intelligibility in both listener groups. The NS group transcribed it wrong or left it blank 22 times out of 36, and the NNS group 22 times out of 30. The word “veterinary” was heard as “returnary”, “returning”, “retinary” or similar in both groups, effectively showing that the unusual word stress had an effect on its intelligibility.

Another word that was misheard was “Sarah” in sample 1, which was heard as “Sir” in both listener groups. In the NS group, it was misheard or left blank 26 out of 36 times and in the NNS group 14 times out of 30. The result was unexpected, as there was no intention of misaligning this word. The possible reasons are discussed in section 6. Other samples, which focused on intonation, rhythm, and sentence stress, did not show any noticeable loss in intelligibility.

## 5.2 Comprehensibility

Regarding comprehensibility, the difference between the listener groups was found along the line; the samples were more difficult for the NNS group than they were for the NS group in each area compared. The measuring scale in this section ranged from 0 to 100, wherein 0 meant that the participant “strongly disagreed” with a statement about how difficult a certain feature was (e.g. “segmentals made the speaker more difficult to understand”), and 100 meant that they “strongly agreed” with the statement. Hence, the lower the score, the easier the speaker was to understand. Statistically significant differences were found concerning *intelligibility*, *overall comprehensibility*, *segmentals*, and *central message*. In overall, the results were on the lower side of the scale (combined mean NS = 23.45, NNS = 29.43). The results are depicted in Table 3 and are discussed section by section further on.

Table 3: Results of the comprehensibility section. The table depicts the measured feature (scale range from 0 to 100 meaning that the higher the score, the harder it was to understand), means for both listener groups in percentages, t-value, and p-value \* <.05. Asterisk marks statistically significant difference.

| Feature  | NS mean | NNS mean | t-value | p-value |
|--|---------|----------|---------|---------|
| <b>How much effort it took to understand the speaker in overall</b>                    | 22.64   | 29.78    | -2.753  | .006*   |
| <b>The single speech sounds made the speaker more difficult to understand</b>          | 25.72   | 32.38    | -2.142  | .033*   |
| <b>Pronunciation of individual words made the speaker more difficult to understand</b> | 29.08   | 38.02    | -2.743  | .006*   |
| <b>The word-stress made the speaker more difficult to understand</b>                   | 21.09   | 25.47    | -1.540  | .125    |
| <b>The sentence-stress made the speaker more difficult to understand</b>               | 22.03   | 27.66    | -1.884  | .06     |
| <b>The intonation made the speaker more difficult to understand</b>                    | 21.78   | 26.70    | -1.708  | .089    |
| <b>The speech rhythm made the speaker more difficult to understand</b>                 | 28.82   | 30.93    | -0.653  | .514    |
| <b>The central message of the speaker was difficult to understand</b>                  | 16.41   | 24.53    | -2.837  | .005*   |

*Overall comprehension* section had a statistically significant difference between the groups ( $p = .006$ ), wherein the effort mean score was higher with the NNS (29.78,  $n = 184$ ) than with the NS (22.64,  $n = 195$ ). In effect, this means that it was more difficult for the NNS group to comprehend the samples than it was for the NS group.

From the comments relating to overall comprehension (see examples 1 and 2), I inferred that generally, all the speakers were evaluated as understandable, but it took some effort to understand them. Every sample had at least one mention wherein the speaker was evaluated favorably regarding overall oral performance.

*(Example 1) "This reminds me a bit of 'rally English' that is understandable but just takes more effort to understand." - NS, S8*

*(Example 2) "It was quite easy to understand readers pronunciation." - NNS, S7*

*Segmental features* section showed statistically significant difference ( $p = .033$ ), as the mean effort score was higher with NNS (32.38,  $n = 160$ ) than with the NS (25.72,  $n = 189$ ). Again, this means that regarding comprehensibility, the segmental features of Finnish spoken English were harder for the NNS group than they were for the NS group.

Regarding segmentals, some participants (see examples 3 and 4) found the words with the /r/ sounds more difficult to understand, as another pointed out that the pronunciation of the alveolar fricative /z/ as a postalveolar affricative /tʃ/ caused hindrance in understanding. Furthermore, a postalveolar fricative /ʃ/ pronounced as alveolar fricative /s/ caused difficulty for at least to one participant.

*(Example 3) "Zs being heard as 'ts' is difficult, and the word 'veterinary' was obviously hard to pronounce for the speaker." - NS, S11*

*(Example 4) "The 'n' in "unsanitary" sounded more like 'r' and "mess" sounded more like "miss" - NNS, S10*

*Individual words* was another section demonstrating a statistically significant difference ( $p = .006$ ) with the mean difficulty score by the NNS group (38.02,  $n = 172$ ) being higher than by the NS group (29.08,  $n = 187$ ). This also indicates that the comprehensibility regarding the individual words took more effort for the NNS group than for the NS group.

Based on the analysis, I inferred that individual words hindered comprehension by disrupting the word processing and requiring careful listening causing fatigue, as explained by participants in examples 5-7:

*(Example 5) "There were a few words that weren't understood at first which caused distraction in the following words, because my focus was still trying to figure out the word that sounded odd." - NS, S7*

*(Example 6) "... I had to listen very carefully to get it as some words were not understandable. If I had to listen to her for any longer, it would have been a lot of work to stay focussed on the message. -NS, S9*

*(Example 7) "Some individual words were probably the thing that made the speaker most challenging to understand. "Before long, the itchy(?) goose" took me many tries to understand, as did "unsanitary mess..." - NNS, S10*

*Word stress* and *sentence stress* showed no statistically significant difference between the listener groups in the first ( $p = .06$ ) or the latter ( $p = .125$ ). Even so, the NNS group had more difficulty in both word stress (25.47,  $n = 144$ ) and sentence stress (27.66,  $n = 148$ ), than the NS group, whose score was lower regarding both word stress (21.09,  $n = 186$ ) and sentence stress (22.03,  $n = 175$ ). This is along the line with the other results of this study.

From the comments (see examples 8-10) regarding these aspects I concluded that the stressing of the "right" words helped, and if the stress was in the "wrong place" such as on an article, which are not usually stressed, it hindered comprehension.

(Example 8) "...changing of the stressed syllable - as when the speaker says the word *solitary* - makes many words more of a guess than something I would say I understood." - NNS, S9

(Example 9) "...it felt easier to understand as a whole because the rhythm seemed familiar and the 'right' words were being stressed..." - NS, S10

(Example 10) "A few words were more difficult to understand- e.g. *veterinary*, and sometimes the stressed word was an article like 'a' or 'the'." - NS, S11

*Intonation* showed no statistically significant difference ( $p = .89$ ) but the NNS group yet again scored higher (26.70,  $n = 160$ ) in difficulty than the NS group (21.78,  $n = 180$ ).

Intonation was mentioned in the comments both as a hindering and as a helping feature. It seemed to aid comprehensibility when functioning as an "emphasizer" of important information, and hindered comprehensibility when the intonation was perceived as "even". Comments 11–13 exemplify:

(Example 11) "The intonation was very easy to follow....I feel like this was the reason why the central message was also quite easy to grasp." - NNS, S9

(Example 12) "Intonation also affected understanding a bit: in "Then Sarah confirmed that diagnosis", the intonation was rather even which caused me to kind of ignore the sentence at first, even though I found it to be important information." - NS, S10

(Example 13) "Intonation was missing, undstandable but not easy or nice to listen to." - NS, S11

*Speech rhythm* was a feature that did not show a statistical difference between the groups ( $p = .514$ ) yet the NNS group had more difficulty (30.93,  $n = 167$ ) than the NS group (28.82  $n = 183$ ).

Regarding comments about rhythm, it seemed that the pauses at expected places aided comprehension and the slow tempo made the speaker more pleasant to listen to. Furthermore, pauses effectively functioned as dividers of speech units, thus aiding listeners' comprehension. In addition, I inferred that even if the participants were able to understand the utterances at the lexical level, or in other words, they were intelligible to them, the *pausing* in unexpected places caused them to be distracted from the pragmatic sense of the sentence, effectively hindering comprehensibility.

The comments 14–17 demonstrate these aspects:

(Example 14) "... the central message took me a longer time to understand. The last two sentences were difficult because of the pause in "she administered [pause] either" and after that "Her efforts were not futile." I thought at first that they went such as: she administered. Either (,) her efforts were not futile... [cut]" and that made no sense to me in any way. - NNS, S8

(Example 15) "The speech rhythm was more difficult to follow as the speaker made pauses in the phrases in odd places (e.g. "at an [pause] old zoo", "in North Square, near [pause] the Duke Street Tower[?]?"). This made focusing challenging. - NNS, S11

(Example 16) “Words were pronounced correctly, but the pauses made it more difficult to understand.” – NS, S10

*Central message*, finally, was a feature that showed a statistically significant difference ( $p = .005$ ), as the mean effort score was higher with the NNS (24.53,  $n = 159$ ) than with the NS (16.41,  $n = 169$ ). This, in turn, effectively shows that the central message was harder for the NNS group to comprehend.

From the comments concerning the central message (examples 17–20), I concluded that *context* was in a key role especially regarding the understanding of the *central message*, as through context participants were able to “fill in” the words that they did not comprehend or even to replace a wrongly heard word with a more appropriate one.

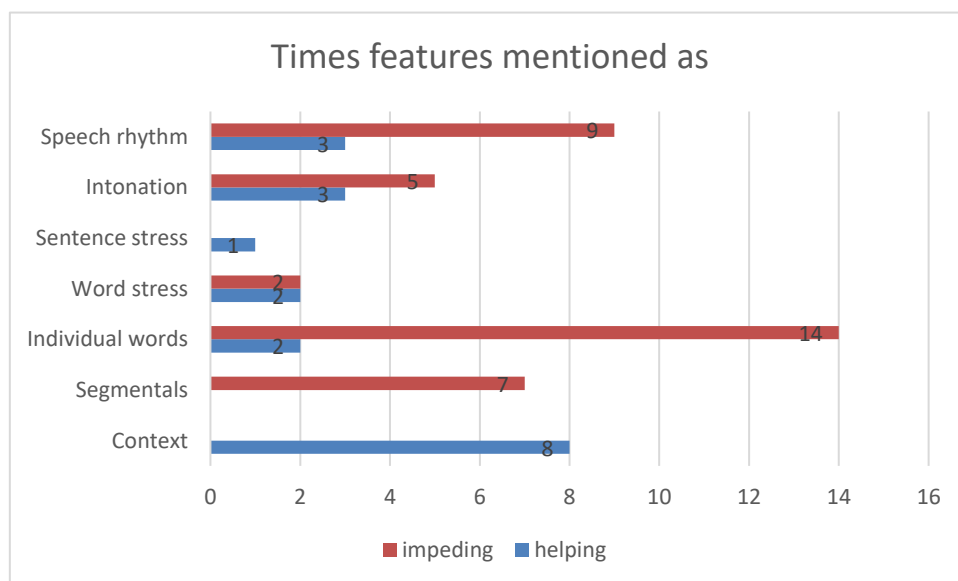
(Example 17) “I understand the audio file in the previous exercise (about penicillin) much better, now that I have heard the context. It became easier to ‘fill in the gaps’ regarding the words I missed the first time.” – NS, S7

(Example 18) “...it was easier to understand e.g. “either” now that it was in context, unlike in the short out-of-context clip that I now realise I transcribed incorrectly.” – NNS, S7

(Example 19) “...I also hear ‘venison’ as the last word, though from context I’m assuming it’s ‘medicine’” – NS, S10

Finally, as described in Table 4, I concluded that the context was the most helpful aspect regarding comprehensibility. In turn, individual words was the most impeding feature, followed by speech rhythm and segmentals.

Table 4: The times feature was mentioned in the comments as impeding or as helping regarding comprehensibility.





In conclusion, all along the line regarding both intelligibility and comprehensibility, it was more difficult for the NNS group to understand the samples than it was for the NS group. Therefore, there was no shared language benefit found in this study. The possible reasons are discussed in section 6.

## 6 DISCUSSION

For reasons not entirely unambiguous, the results in both sections of intelligibility and comprehensibility showed that the samples were more understandable for the NS group than they were for the NNS group. In overall, the results were on the lower side on the 0–100 difficulty scale, averaging between 20 and 30 in both groups.

Along the line, the native speakers deemed every sample more intelligible than the non-native speakers. The study results also show that there was no shared language benefit, as the Finnish listener group had lower comprehensibility results overall, along the line. This is in contrast to Brent and Bradlow (2003) and Foote and Trofimovich (2018), who found a shared language benefit in their studies.

Furthermore, the study results indicate that native and non-native speakers sharing the speakers' L1 may rely on different things regarding intelligibility and comprehensibility, as there were statistical differences regarding intelligibility, overall comprehensibility, segmentals, and perceiving the central message of a passage. Cutler and Butterfield (1992) have found that native speakers of English rely at least partially on strong and weak forms of speech sounds when segmenting the words and this may offer one reason regarding the differences between the groups. Moreover, results from Connell et al. (2018) show that native English speakers tend to rely as well on the word-stress to differentiate words, offering another explanation for the differences.

Especially segmentals and word stress seemed to have the strongest effects regarding intelligibility. In terms of word stress, the first sample utilized the phrase "Sarah Perry was a veterinary nurse", which was manipulated by shifting the word stress in the word "veterinary" to the right, on the second syllable. The stress was realized through pitch, length, and intensity. Hence, instead of the stress that would be normally on the first syllable "VET-e-ri-na-ry" it was on the second, "vet-E-ri-na-ry". Even though the word stress is a distinctive feature in English and not in Finnish (see section 2.3), both listener groups had significant difficulties in understanding the word. The result is in line with the previous studies (see section 3.2), which have found word

stress as a significant factor in terms of intelligibility. Furthermore, the results are in line with Isaac and Trofimovich (2019) who also found word stress to be a significant factor in terms of comprehensibility. As such, Field (2005) has found the intelligibility to be affected more when the stress was shifted to the right, as was the case with this particular word.

In addition, the words in the mental lexicon are accessed at least partially by relying on word stress (see section 2.3). This may offer one possible reason why the word “veterinary” was mostly unintelligible. Moreover, aligned with Field (2005), the result was expected; Field wrote that the shifting of the stress may affect intelligibility and it did so in this study. Finally, regarding the first sample, the mishearing of the word in focus might have been further hindered by the lack of context, as the other words in the phrase did not provide clues about the word “veterinary”. This, in turn, is in line with Sajavaara and Dufva (2001) who state that listeners are usually able to decipher the word’s meaning through context, effectively disregarding single phonological errors. Consequently, the participants were indeed able to decipher the word “veterinary” from the context in the comprehensibility section, which utilized the same phrase but with a longer sample, providing more context.

One feature that affected intelligibility in both listener groups was segmentals. In sample 3, the speaker uttered “Her efforts were not futile” with many deviations in segmentals. The speaker pronounced the word “her” as [hø:], instead of /hɜ:/, followed by “efforts” as [efo:ts] instead of /efə(r)ts/, and the final context word “futile” was pronounced as /fju:til/ instead of /fjutəl/ as it is in American Standard English or /fju:tail/ in Standard Southern British English, to name a few alternatives.

The L1 Finnish effect on the segmentals on their pronunciation was the highest on this speaker, whose level of English was estimated to be B1. However, whether their level of English and the influence of L1 are related was not examined in this study. Nevertheless, sample 3 was the least understood both regarding intelligibility and comprehensibility. As stated by Jenkins (1995), a single phonological error may not hinder comprehension but several might do so. As this sample had several peculiarities in the segmentals, possibly influenced by the L1 Finnish, this could provide one reason for the listener misunderstandings.

Riney et al. (2005) have found that the NS and NNS speakers rely on different cues when differentiating the words and their meanings. This study’s results are in line with theirs, yet I did not concentrate on *how* they differ. What is clear, however, is that the L1 certainly affects the pronunciation and the perception of L2.

Considering the sounds not found in Finnish phonology, which were anticipated to cause hindrance, were, for instance, the word “goose” pronounced as [ku:s] instead of /gu:s/. However, even though it was noted, it did not affect intelligibility. A

possible reason might be that there was no adjacent word to which it could have been mistaken for. Hence, in line with Lintunen (2015), it did not cause hinderance.

Another impeding feature in this sample may have been the lack of pause in between the words “her” and efforts” and their subsequent assimilation to an utterance of [hə:efo:ts]. As noted by Romero-Trillo and Newell (2012) the rhythm (including the concept of pauses) is a feature used by native English speakers when processing the meaning. As there was no pause between the words, the listeners might have perceived it as a non-initial word or were unable to distinguish it as a separate word, resulting in transcriptions such as “hey...” herefore”, therefore” or similar. Hence, also speech rhythm affected both intelligibility and comprehensibility.

Other possible reasons for the low intelligibility and comprehensibility of sample 3 could be the attuning to strong and weak syllables and word stress. Zielinski (2018) has found that native English speakers listen to stress patterns regarding the strong and weak syllables. They found that misalignment on the standard stressing considering the syllables influenced understanding. This effect can be seen in the assimilation of the first two content words in sample 3, leading to the utterance [hə:efo:ts]. The emphasis, which is usually on the first syllable of the word “effort” was now on the word “her”, and as such, could provide a reason why the sentence was highly unintelligible. The importance of segmentals is also pointed out by Saito et al. (2016) and thus this study’s results are in line with theirs.

Segmentals caused impedance in other samples as well. Participants pointed to the alveolar fricative /z/ as a postalveolar affricative [tʃ] in sample 4 to cause additional effort, the palatoalveolar fricative /ʃ/ realized as a laminal alveolar fricative [s] in sample 5, and the “j sounds” and “super strong r’s” in overall to cause hindrance. These factors seemed to not affect intelligibility but hindered the comprehensibility with processing time delays and additional attention requirements.

Regarding functional load, the abovementioned segmentals bear a high-mid-range to low functional load. Following Brown’s (1988: 604) predictions, the /s/ and /ʃ/ appear as number 7 on Brown’s FL scale ranging from 1 to 10, wherein zero equals low FL and 10 high FL. The difference was realized, for instance in the word “itchy” (kutiava) pronounced as [itsi] instead of /itʃi/, causing impedance in comprehensibility. In turn, by “J” sounds, I inferred that the participants meant the difference between the postalveolar fricatives /dʒ/ and /tʃ/, not found in Finnish phonology. Consequently, this may be the reason why the Finnish speaker did not distinguish them in their speech. The difference is marked as number 3 on Brown’s (1988) scale. Again, it may contribute more to comprehensibility than to intelligibility, as it was noted but the participants were able to make comments about the right word.

Curiously, in sample 4, the speaker uttered the words “bowl of porridge” as [pəʊl of əv pɔːrɪdʒ]. In reference to Brown’s (1988) predictions, the difference between /p/ and /b/ are among the highest considering FL. However, virtually every participant transcribed this sentence right even when the word “bowl” was pronounced with /p/. I speculate, following Flege’s (1995) SML model, that they were able to categorize these sounds as “similar” and deduct the meaning from context. Hence, the context seems to have the possibility to “override” even the most severe FL errors.

As such, the “r sounds” are apical alveolar trills in Finnish whereas they are alveolar approximants in English. This might have caused hindrance again on the comprehensibility but did not affect intelligibility, as the participants were able to understand the word in the comments. Curiously, “her” was heard many a time as “hey” in both listener groups, possibly because of the speaker’s way of pronouncing it without the final “r” as is common in British English. However, the speaker did not pronounce the final word “futile” as British /fjuːtaɪl/ but with a more likeness to the American Standard /fjuːtəl/. The key difference was the missing schwa in the second syllable, resembling more “t” than “d” sound. Nevertheless, the word was understood correctly in both groups. This could be conducted to Flege’s SLM presented in section 3.5. Even if the Finnish /r/ is not in exact correspondence to an English /r/, I presume the participants were able to unconsciously categorize it as “similar sounds”.

Some samples did not prove to be unintelligible, even when they had segmental errors. For instance, sample 4 had the phrase “She ate a bowl of porridge”. Nearly all of the participants were able to understand it correctly, even though the speaker uttered “she” /ʃiː/ (nähdä), as “see”, [siː] (meri). This segmental error was somewhat severe, as it effectively was a completely different phoneme; a laminal alveolar fricative [s] where it should have been a palatoalveolar fricative /ʃ/. The latter sound exists only in the loan words in the Finnish phonology and thus may have been the reason why the speaker did not utter it as it is in English. Referring again to Sajavaara and Dufva (2001), the individual segmental errors may not create a problem, as the meaning is understood through context. Hence, all speakers in both groups transcribed it accordingly.

Based on the results, the more context is provided, the less effect a single phonological error had, as could be seen in the longer passages in the comprehensibility section. Had the intelligibility part taken to the opposite end, providing only one word long samples, the abovementioned [siː] would have almost certainly been transcribed as “see” in the participant responses. Thus, the context has a substantial role in intelligibility.

Regarding the comprehensibility section, the most impeding feature was individual words. They hindered comprehensibility by slowing down the word procession in the speaker and by requiring careful attention in order to understand the

speaker. The result is in line with Munro and Derwing (1995b). Thus, even if a single word does not hinder intelligibility, it may hinder comprehensibility. However, the individual words reported both as a helping and as an impeding feature demonstrate how the study results cannot be generalized but have to be seen in regard to limitations of this study, offering insight and depth, but no conclusive results.

The studying of pronunciation features is made difficult by their interwovenness. It is hard to separate an individual feature from another, as they often fundamentally contribute to each other. Thus, words consist of segmentals but in connected speech segmentals do not appear in a vacuum and the borders are at times difficult to define; is a certain misunderstanding due to a detail in segmentals, or has it more to do with word stress, which could have been also misaligned. Curiously, in contrast to Hahn (2004) and Levis (2018), the misalignments in the sentence-stress did not affect intelligibility or comprehensibility in this study.

The comments that described a feature that I had not taken into account were the most troublesome ones. These were “enunciation” and “familiarity” which were frequented in the answers. As they are not features of neither a speaker nor a listener, I opted to not include them in my study. However, as “context” was the most mentioned among these, and as it was the only one clearly defined and described in the comments, I included it in my analysis. In addition, erring on the safe side, I chose to leave aside comments such as “her pronunciation was a bit off” and “improper grammar”, as I was not sure to which feature they point to.

As such, context seemed to be one of the most relevant features in terms of both intelligibility and comprehensibility. Even though it is not a feature of a speaker nor a feature of a listener, I included it in the results as it was mentioned several times with both listener groups. Through context, listeners were able to mentally add or change the words they did not hear or misheard and infer what the intended word was. However, context is a highly complex phenomenon that has more to do with the field of pragmatics than with phonetics and was therefore left outside the theoretical discussion in this study.

Initially, I wanted to compare the possible similarities and differences between the listener groups with descriptive statistics and by comparing the open comments, but as there were significantly more comments made by the NS group (44) than by the NNS group (10), I chose to represent them in aggregate. Thus, no comparison could be made between the groups based on the qualitative data obtained. Furthermore, the questionnaire did not provide a possibility for open comments in the intelligibility section, and the qualitative data in that section is based on the transcriptions made by the participants. Despite these setbacks, the qualitative data still provided valuable insight into how the features affect intelligibility and comprehensibility.

Improvements could be made by narrowing the research focus to include either the concept of intelligibility or comprehensibility, as opposed to including both. This could reduce the workload, and in turn enable more concentration on the chosen one. Nevertheless, more holistic approaches, such as mine, are needed in order to seek out the relevant features before being able to narrow them. In addition, mixed methods approach and triangulation provide more depth and reliability to the study yet require meticulous dedication, time, and a high amount of work. The challenging way chosen for this study proved to be successful, as with the mixed methods approach I was able to reach the study aims and answer the designated research questions.

In addition, the role of qualitative data could be enhanced. As such, the rating task could be improved by adding a commenting possibility to the intelligibility section. It may provide some insight into the transcription choices made by the participants. The comments could have enabled a more thorough qualitative analysis regarding the intelligibility section, and possibly a more thorough understanding of the features relating to intelligibility. Nevertheless, the answer transcriptions themselves proved to be valuable and suitable for analysis, and the additional commenting possibility could have increased the already high workload excessively.

Moreover, the securing of enough answers is an aspect that could be considered more thoroughly. I closed the rating task after receiving 67 participants and 156 initial comments. After careful exclusion of the comments that I found to be not useful, 54 remained. Out of these, 44 comments were made by the NS group and 10 by the NNS group. This effectively made the comparison between the groups impossible. Hence, I presented the content analysis using the comments in aggregate, which proved to be a successful choice.

Initially, I planned to convey the study by using interview data with Alaskan participants. In the end, this proved to be impossible. In retrospect, the interviews may have added additional depth to the answers and possibly raised some unknown aspects in focus. However, mixed methods approach used provided the desired depth with qualitative data analysis, and enabled the use of quantitative methods, which provided potency to the research question answers.

Other aspects of improvement could have been the background section, which included questions that were not, in the end, analyzed in my study. These were for instance about the English level of the participant, their age, and their place of residence. A more careful questionnaire design is therefore advocated.

In summary, the features of pronunciation are interwoven, often inextricable, and affect each other in complex ways. The context seems to provide the most help, regarding both concepts of intelligibility and comprehensibility. In turn, word stress, individual words, speech rhythm, and segmentals can all cause hindrance in both concepts and thus could be paid more attention to in teaching practices.

## 7 CONCLUSION

In this study, I examined the features of pronunciation that contributed to the concepts of intelligibility and comprehensibility. My focus was on the segmental and suprasegmental features. I built my theoretical framework upon the extensive previous research done on these concepts, considered the relevant pronunciation features, such as intonation and speech rhythm, and contrasted the phonological structures of Finnish and English. Using mixed methods and triangulation in both the data collection and in the subsequent data analysis, I sought to further deepen my understanding of these complex matters. As such, quantitative and qualitative results offered insight by complementing each other and providing useful data for the linguistic field.

My first research question asked what segmental or suprasegmental features have the most effect considering intelligibility and comprehensibility in the L2 English by L1 Finnish speakers. Based on the results of this study, *segmentals* and *word stress* seemed to have the strongest effect in both listener groups. In turn, *intonation*, *speech rhythm*, or *sentence stress* seemed to have no noticeable effect on intelligibility. Regarding comprehensibility, the most helpful aspect was *context*, which was frequented in the open comments and seemed to be one of the key aspects which enhanced comprehensibility. In turn, *individual words*, *speech rhythm*, and *segmentals* were the most impeding features.

My second research question asked if there are differences in the affecting factors between native listeners and non-natives, and if so, what are they. As such, statistically significant differences were found between the listener groups concerning *intelligibility*, *overall comprehensibility*, *segmentals*, and *central message*. Indeed, differences were therefore found, but this study did not aim to provide an answer to their cause.

My third research question was that do L1 speakers of Finnish experience a shared language benefit when listening to their peers' English. As such, the study results show that there was no interlanguage intelligibility speech benefit, or shared language benefit, noticed in this study. This means that having Finnish as a mother



tongue did not provide any advantage for the participants in the Finnish group for understanding Finnish spoken English. Overall, the previous studies on this matter are inconclusive, as some have found L1 to provide an advantage and some have found none. This paper contributes to the latter studies.

Regarding notes on future studies, 30 people filled the rating task using a computer, and 37 of the study participants answered it via a mobile device. This aspect was not considered when creating the task, and while it did not affect the actual content of the study, it may have some unknown streamlining issues regarding the layout or functionality of the questionnaire. This is something that would be beneficial to consider in future studies utilizing internet rating tasks, as most people tend to favor answering through a smartphone.

It is clear that L1 has an effect on the pronunciation features of L2. However, pedagogically, it is difficult to raise one feature to a more relevant status than the other as the features affect each other. Suprasegmental features ultimately consist of segmental features, yet it cannot be inferred that they are more relevant when it comes to comprehensibility. However, based on the results of this study, it seems that intelligibility could be improved by focusing more on the correct word stress and improving segmental clarity. This could be done by explaining the differences between strong and weak syllables and how they relate to understanding to the students. The sentence stress cannot be left aside either: as noted by a participant, if the stress falls onto an article, which are not usually stressed, it may hinder the comprehensibility. Furthermore, the speech rhythm could be emphasized, and especially the importance of pauses, which do not only distinguish words but also divide phrases in the listener's mind.

Even though the importance of context was found to be significant within my study, it is not a feature of speech nor a feature of listening. That being said, how context contributes to understanding the central message of the speaker could be highlighted in education. It can be concluded that regardless of the approach, further studies are needed to enhance our understanding of these complex and highly interesting aspects of what contributes to intelligibility and comprehensibility in oral communication.

## REFERENCES

- Alanen, R. (2011). Kysely tutkijan työkaluna. In Kalaja P., Alanen, R. and Dufva, H. (eds.), *Kieltä tutkimassa. Tutkielman laatijan opas*. Tampere: Finn Lectura, 146–161.
- Anders, L. (2012). Developing Survey Questions. In *Designing & Doing Survey Research* [online], 61–90. SAGE Publications, Ltd. Doi: <https://dx.doi.org/10.4135/9781526402202.n5>.
- Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, 19(3), 251–265. Doi: <https://doi.org/10.1191/0267658303sr222oa>.
- Beinhoff, B. (2014). Perceiving Intelligibility and Accentedness in Non-Native Speech. A Look at Proficiency Levels. *Concordia Working Papers in Applied Linguistics*, 4, 58–72. [http://doe.concordia.ca/copal/documents/6\\_Beinhoff\\_Vol5.pdf](http://doe.concordia.ca/copal/documents/6_Beinhoff_Vol5.pdf). (7 October, 2021).
- Benson, B. (1986). The Markedness Differential Hypothesis. Implications for Vietnamese Speakers of English. In Eckman, F.R., Moravcsik, E.A, Wirth, J.R (eds.), *Markedness*. [https://doi.org/10.1007/978-1-4757-5718-7\\_14](https://doi.org/10.1007/978-1-4757-5718-7_14).
- Bent, T. and Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustic Society of America*, 114(3), 1600–1610. Doi: <https://doi.org/10.1121/1.1603234>.
- Best, C. (1995). A Direct Realist View of Cross-Language Speech Perception. In Strange W. (ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*, 171–204. [https://www.researchgate.net/publication/229078334\\_A\\_direct\\_realist\\_view\\_of\\_crosslanguage\\_speech\\_perception](https://www.researchgate.net/publication/229078334_A_direct_realist_view_of_crosslanguage_speech_perception). (25 February, 2022).
- Bradley, J.S. (1986). Speech intelligibility studies in classrooms. *The Journal of the Acoustical Society of America*, 80(3), 846–854. <https://doi.org/10.1121/1.393908>.
- Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. Doi: <https://doi.org/10.1191/1478088706qp063oa>.
- Brown, A. (1988). Functional Load and the Teaching of Pronunciation. *TESOL Quarterly*, 22(4), 593–606. <https://doi-org.ezproxy.jyu.fi/3587258>.
- Burda, A.N., Hageman, C.F., Scherz, J.A. and Edwards, H.T. (2003). Age and Understanding Speakers with Spanish or Taiwanese Accents. *Perceptual and Motor Skills*, 97(1), 11–20. <https://doi.org/10.2466/pms.2003.97.1.11>.
- Carlisle, R. (1991). The Influence of Environment on Vowel Epenthesis in Spanish/English Interphonology. *Applied Linguistics*, 12(1), 76–95. Doi: <https://doi-org.ezproxy.jyu.fi/10.1093/applin/12.1.76>.
- Catford, J.C. (1987). Phonetics and the teaching of pronunciation. A systemic description of English phonology. In Morley, J. (ed.), *Current Perspectives on Pronunciation. Practices Anchored in Theory*, 87–100. Washington: TESOL.

- Connell, K. Hüls, S. Martínez-García, M.T., Qin, Z., Shin, S., Yan, H. and Tremblay, A. (2018). English Learners' Use of Segmental and Suprasegmental Cues to Stress in Lexical Access. An Eye-Tracking Study. *Language Learning*, 68(3), 635–668. Doi: <https://doi.org/10.1111/lang.12288>.
- Council of Europe. (2020). *Common European Framework of Reference for Languages. Learning, teaching, assessment. Companion volume*. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>.
- Cutler, A. and Butterfield, S. (1992). Rhythmic Cues to Speech Segmentation. Evidence from Juncture Misperception. *Journal of Memory and Language*, 31(2), 218–236. Doi: [https://doi.org/10.1016/0749-596X\(92\)90012-M](https://doi.org/10.1016/0749-596X(92)90012-M).
- Derwing, T. and Munro, M. (2009). Comprehensibility as a Factor in Listener Interaction Preferences. Implications for the Workplace. *Canadian Modern Language Review*, 66(2), 181–202. Doi: <https://doi.org/10.1353/cml.0.0111>.
- Derwing, T. and Rossiter, M. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning* 13(1), 1–17. [https://www.researchgate.net/publication/234570703\\_The\\_Effects\\_of\\_Pronunciation\\_Instruction\\_on\\_the\\_Accuracy\\_Fluency\\_and\\_Complexity\\_of\\_L2\\_Accented\\_Speech](https://www.researchgate.net/publication/234570703_The_Effects_of_Pronunciation_Instruction_on_the_Accuracy_Fluency_and_Complexity_of_L2_Accented_Speech). (7 October, 2021).
- Deterding D. (2010). Norms for pronunciation in Southeast Asia. *World Englishes* 29(3), 364–367. <https://doi.org/10.1111/j.1467-971X.2010.01660.x>.
- Deterding, D. and Lewis, C. (2019). Pronunciation in English as Lingua Franca. In Gao, X. (ed.), *Second Handbook of English Language Teaching*. Cham: Springer. 1–16. Doi: [https://doi.org/10.1007/978-3-319-58542-0\\_41-1](https://doi.org/10.1007/978-3-319-58542-0_41-1).
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics. Quantitative, Qualitative, and Mixed Methods*. Madrid: Oxford University Press.
- Eberhard, M., Simons, G. and Fennig, C. (2021). *Ethnologue. Languages of the World* (24<sup>th</sup> edition). Texas: SIL International.
- Eckman, F. (1985). Some Theoretical and Pedagogical Implications of the Markedness Differential Hypothesis. *Studies in Second Language Acquisition*, 7(3), 289–307. Doi: <https://doi.org/10.1017/S0272263100005544>.
- Empirical Research (N.d.). <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/en/methodmap/strategies/empirical-research>. (15 January, 2022).
- Escudero, P. and Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, 26(4), 551–585. Doi: <https://doi.org/10.1017/S0272263104040021>.
- Eskola, J. (2018). Laadullisen tutkimuksen juhannustaiat. Laadullisen aineiston analyysi vaihe vaiheelta. In Valli, R., Aaltola, J. and Herkama, S. (eds.) *Ikkunoita tutkimusmetodeihin 2. Näkökulmia aloittelevalle tutkijalle tutkimuksen teoreettisiin lähtökohtiin ja analyysimenetelmiin* (5th ed.) Jyväskylä: PS-Kustannus.
- Eskola, J. and Suoranta, J. (1998). *Johdatus laadulliseen tutkimukseen*. Tampere: Vastapaino.

- Feigl, H. (2021). Positivism. *Encyclopedia Britannica*.  
<https://www.britannica.com/topic/positivism>. (25 February, 2022).
- Field, J. (2005). Intelligibility and the Listener. The Role of Lexical Stress. *Tesol Quarterly* 39(3), 399–423. Doi: <https://doi.org/10.2307/3588487>.
- Finnish National Board on Research Integrity TENK. (2019). *The ethical principles of research with human participants and ethical review in the human sciences in Finland*. In Kohonen, I., Kuula-Luumi, A. and Spoof, S-K. (eds.), Tutkimuseettisen neuvottelukunnan julkaisu, 3, 1–72.  
[https://www.tenk.fi/sites/tenk.fi/files/Ihmistieteiden\\_eettisen\\_ennakkoarvio\\_innin\\_ohje\\_2019.pdf](https://www.tenk.fi/sites/tenk.fi/files/Ihmistieteiden_eettisen_ennakkoarvio_innin_ohje_2019.pdf). (15 February, 2022).
- Flege, J. (1995). Second language speech learning. Theory, findings and problems. In Strange, W. (ed.), *Speech perception and linguistic experience. Issues in cross-language research*, Baltimore: New York Press. 233–277.  
[https://www.researchgate.net/publication/333815781\\_Second\\_language\\_speech\\_learning\\_Theory\\_findings\\_and\\_problems](https://www.researchgate.net/publication/333815781_Second_language_speech_learning_Theory_findings_and_problems).
- Flege, J. (1998). Second-language Learning. The Role of Subject and Phonetic Variables. *Speech Technology in Language Learning*, 1–8. [https://www.isca-speech.org/archive\\_open/archive\\_papers/stl8/stl8\\_001.pdf](https://www.isca-speech.org/archive_open/archive_papers/stl8/stl8_001.pdf). (17 February, 2022).
- Flege, J. and Bohn, O-S. (2021). The Revised Speech Learning Model SLM-r. In Wayland, R. (ed.), *Second Language Speech Learning. Theoretical and Empirical Progress*. Doi: <http://doi.org/10.1017/9781108886901.002>.
- Flege, J., MacKay, I. and Meador, D. (1999) Native Italian speakers' production and perception of English vowels. *Journal of Acoustical Society of America*, 106(5), 2973-2987. Doi: <https://doi.org/10.1121/1.428116>.
- Flege, J.E., Munro, M. and MacKay, I. (1995). Factors Affecting Strength of Perceived Foreign Accent in a Second Language. *Journal of the Acoustical Society of America*, 97(5), 3125–31. Doi: <https://doi.org/10.1121/1.413041>.
- Foot, J.A and Trofimovich, P. (2018). Is It Because of My Language Background?. A Study of Language Background Influence on Comprehensibility Judgments. *The Canadian Modern Language Review*, 74(2), 253–278. Doi: <https://doi.org/10.3138/cmlr.2017-0011>.
- Gick, B., Wilson, I. and Derrick, D. (2013). *Articulatory Phonetics*. Malden, Mass.: Wiley-Blackwell.
- Hahn, L. (2004). Primary Stress and Intelligibility. Research to Motivate the Teaching of Suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. Doi: <https://doi.org/10.2307/3588378>.
- Haspelmath, M. (2001). Word Classes and Parts of Speech. In Smelser, N.J. and Baltes P.B (eds.), *International Encyclopedia of the Social & Behavioral Sciences*, 16538–16545. Doi: <https://doi.org/10.1016/B0-08-043076-7/02959-4>.
- Hayes-Harb, R., Smith, B.L., Bent, T. and Bradlow, A. The interlanguage speech intelligibility benefit for native speakers of Mandarin. Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, 36(4), 664–679. Doi: <http://doi.org/10.1016/j.wocn.2008.04.002>.
- Heikkilä, T. (2014). *Tilastollinen tutkimus*. Helsinki: Edita Publishing.

- Hesse-Biber, S.N. (2010). *Mixed Methods Research. Merging Theory with Practice*. New York: Guildford Press.
- Hypoteesien testaus. (N.d.). Kvantitatiivisen tutkimuksen verkkokäsikirja. *Yhteiskuntatieteellinen tietoaarkisto* [online]. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/hypoteesi/testaus/>. (4 February, 2022).
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Isaacs, T. (2013). Phonology. Mixed Methods. In C.A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, 7, 4427–4434.
- Isaacs, T. and Trofimovich, P. (2012). Deconstructing Comprehensibility. Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. Doi: <https://doi.org/10.1017/S0272263112000150>.
- James M.A. (2012). Cross-Linguistic Influence and Transfer of Learning. In Seel N.M. (ed.), *Encyclopedia of the Sciences of Learning* [online]. [https://doi.org.ezproxy.jyu.fi/10.1007/978-1-4419-1428-6\\_702](https://doi.org.ezproxy.jyu.fi/10.1007/978-1-4419-1428-6_702).
- Jarvis, S. and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.
- Jenkins, J. (1995). *Variation in Phonological Error in Interlanguage Talk*. Unpublished Doctoral Dissertation. University of London, Department of Education. <https://core.ac.uk/download/pdf/111068675.pdf>. (2 November, 2021).
- Jenkins, J. (1998). Which pronunciation norms and models for English as an International Language. *ELT Journal*, 52(2), 119–126. Doi: <https://doi.org/10.1093/elt/52.2.119>.
- Jesney, K. (2004). The Use of Global Foreign Accent Rating in Studies of L2 Acquisition. A report prepared for the Language Research Centre University of Calgary. <https://www.semanticscholar.org/paper/The-Use-of-Global-Foreign-Accent-Rating-in-Studies-Jesney/3e4eaa880f359be90d48eb43df7574d98630926f>. (5 January, 2022).
- Jin, L. (2008). Markedness and Second Language Acquisition of Word Order in Mandarin Chinese. In Marjorie K.M. Chan and Hana Kang (eds.) *North American Conference on Chinese Linguistics*, 1, 207–308. [https://naccl.osu.edu/sites/naccl.osu.edu/files/15\\_jin-l.pdf](https://naccl.osu.edu/sites/naccl.osu.edu/files/15_jin-l.pdf). (21 February, 2022).
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* [online] 38(2), 301–315. Doi: <https://doi.org/10.1016/j.system.2010.01.005>.
- Kang, O., Rubin, D. and Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of English language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. Doi: <https://doi.org/10.1111/j.1540-4781.2010.01091.x>.
- Kuronen, M. (2019). Ääntämisen foneettiset oppimiskohteet. In E. Tergujeff, and M. Kautonen (eds.), *Suullinen kielitaito. Opi, opeta, arvioi*. 59–74. Helsinki: Otava.

- Ladefoged, P. and Johnson K. (2010). *A Course in Phonetics* (6<sup>th</sup> ed.) Boston: USA.
- Lado, R. (1957). *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.
- Legg, C. and Hookway, C. (2021). Pragmatism. In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.) [online].  
<https://plato.stanford.edu/archives/sum2021/entries/pragmatism>. (16 January, 2022).
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley & Sons.
- Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press. Doi:  
<https://doi.org/10.1017/9781108241564.004>.
- Lima, E.F. (2016). Comprehensibility and liveliness in nonnative student oral presentations before and after training. A mixed methods study. *System*, 63, 121–133. Doi: <https://doi.org/10.1016/j.system.2016.10.003>.
- Lintunen, P. Ääntämisen oppiminen ja opettaminen. (2015). In Lintunen, P., Veivo, O., Järvinen, H., Sundman, M., Niitemaa, M. and Pietilä, P. (eds.), *Kuinka kieltä opitaan. Opas vieraan kielen opettajalle ja opiskelijalle*. Helsinki: Gaudeamus.
- Low, E.L. (2006). A Cross-Varietal Comparison of Deaccenting and Given Information. Implications for International Intelligibility and Pronunciation Teaching. *TESOL Quarterly*, 40(4), 739–761. Doi: <https://doi-org.ezproxy.jyu.fi/10.2307/40264306>.
- Major, R. (1987). Phonological Similarity, Markedness, and Rate of L2 Acquisition. *Studies in Second Language Acquisition*, 9(1), 63–82. Doi:  
<https://doi.org/10.1017/S0272263100006513>.
- Markie, P. and Folescu, M. (2021). Rationalism vs. Empiricism. In Zalta, E.N (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.) [online].  
<https://plato.stanford.edu/entries/rationalism-empiricism/#IntuThes>. (15 January, 2022).
- McArthur, T., Lam-McArthur, J. and Fontaine, L. (2018). *The Oxford Companion to the English Language* (2<sup>nd</sup> ed.) [online]. Doi:  
<http://doi.org/10.1093/acref/9780199661282.001.0001>.
- McMahon, A.M.S. (2016). *An Introduction to English Phonology*. Edinburgh: Edinburgh University Press.
- Munro, M. (1995). Nonsegmental Factors in Foreign Accent. Ratings of Filtered Speech. *Studies in Second Language Acquisition*, 17(1), 17-34. Doi:  
<https://doi.org/10.1017/S0272263100013735>.
- Munro, M. and Derwing, T. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45(1), 73–97. Doi: <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>.
- Munro, M. and Derwing, T. (1995b). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, 38(3), 289–306. Doi:  
<https://doi.org/10.1177/002383099503800305>.
- Munro, M. and Derwing, T. [1997]. Accent, Intelligibility, and Comprehensibility. Evidence from Four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. Doi:  
<https://doi.org/10.1017/S0272263197001010>.

- Munro, M., Derwing, T. and Morton, S. (2006). The Mutual Intelligibility of L2 Speech. *Studies in Second Language Acquisition*, 28(1), 111–131. Cambridge University Press. Doi: <https://doi.org/10.1017/S0272263106060049>.
- Munro, M. and Derwing, T. (2006). The Functional Load Principle in ESL Pronunciation Instruction. An Exploratory Study. *System*, 34(4), 520–531. <https://doi.org/10.1016/j.system.2006.09.004>.
- Munro, M. and Derwing, T. (2009). Putting accent in its place. Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490. Doi: <https://doi.org/10.1017/S026144480800551X>.
- National Core Curriculum for Basic Education. (2014). Finnish National Board of Education [online in Finnish]. [https://www.oph.fi/sites/default/files/documents/perusopetuksen\\_opetussuunnitelman\\_perusteet\\_2014.pdf](https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf). (29 October, 2021).
- Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburgh: Edinburgh University Press.
- O'Reilly, K. (2009). Interpretivism. *Key Concepts in Ethnography* [online], 119–124. Doi: <https://dx.doi.org/10.4135/9781446268308>.
- Otos ja otantamenetelmät (N.d.). Kvantitatiivisen tutkimuksen verkkokäsikirja. *Yhteiskuntatieteellinen tietoaarkisto* [online]. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/otos/otantamenetemat/>. (11 January, 2022).
- Patkowski, M.S. (1990). Age and Accent in a Second Language. A Reply to James Emil Flege. *Applied Linguistics*, 11(1), 73–89. Doi: <https://doi.org/10.1093/applin/11.1.73>.
- Philosophy of Science. (N.d.). <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/en/methodmap/philosophy-of-science/philosophy-of-science>. (15 January, 2022).
- Riney, T.J, Takagi, N. and Inutsuka, K. Phonetic Parameters and Perceptual Judgments of Accent in English by American and Japanese Listeners. *Tesol Quarterly*, 39(3), 441–466. Doi: <https://doi-org.ezproxy.jyu.fi/10.2307/3588489>.
- Roach, P. (2009). *English Phonetics and Phonology. A practical course* (4<sup>th</sup> ed.) Cambridge: Cambridge University Press.
- Romero-Trillo, J. and Newell, J. (2012). Prosody and Feedback in Native and Non-Native Speakers of English. In Romero-Trillo, J. (ed.), *Pragmatics and Prosody in English Language*, 117–131. Madrid: Springer.
- Saito, Y. and Saito, K. (2016). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation. The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. Doi: <https://doi.org/10.1177/1362168816643111>.
- Saito, K., Trofimovich, P. and Isaacs, T. (2017). Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness. A Validation and Generalization Study. *Applied Linguistics*, 38(4), 439–462. Doi: <https://doi.org/10.1093/applin/amv047>.
- Saito, K., Webb, S., Trofimovich, P. and Isaacs, T. (2016). Lexical Profiles of Comprehensible Second Language Speech. The Role of Appropriateness,


- Fluency, Variation, Sophistication, Abstractness and Sense Relations. *Studies in Second Language Acquisition*, 38(4), 677–701. Doi: <https://doi.org/10.1017/S0272263115000297>.
- Sajavaara, K. and Dufva, H. (2001). Finnish-English Phonetics and Phonology. *International Journal of English Studies*, 1(1), 241–256. [https://www.researchgate.net/publication/27338253\\_Finnish-English\\_phonetics\\_and\\_phonology](https://www.researchgate.net/publication/27338253_Finnish-English_phonetics_and_phonology). (2 November, 2021).
- Savolainen, E. (2001). Verkkokielioppi. Suomen kielen äänne-, muoto- ja lauseoppia. *Mitä fonologia on* [online]. Finn Lectura. <https://fl.finnlectura.fi/verkkosuomi/Fonologia/sivu112.htm>. (22 October, 2021).
- Singh, K. (2007). *Quantitative Social Research Methods*. Los Angeles: Sage Publications.
- Snow, C.E. and Hoefnagel-Höhle, M. (1978). The Critical Period for Language Acquisition. Evidence from Second Language Learning. *Child Development*, 49(4), 1114–1128. <https://www.jstor.org/stable/1128751>. (17 February, 2022).
- Suomi, K., Toivanen, J. and Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, phonology, phonotactics and prosody* [online]. <http://urn.fi/urn:isbn:9789514289842>. (29 October, 2021).
- Suzukida, Y. and Saito, K. (2021). Which Segmental Features Matter for Successful L2 Comprehensibility. Revisiting and Generalizing the Pedagogical Value of the Functional Load Principle. *Language Teaching Research*, 25(3), 431–450. Doi: <https://doi.org/10.1177/1362168819858246>.
- Szmigiera, M. (2021). Statista. The most spoken languages worldwide in 2021. <https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/>. (15 March, 2022).
- Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels. A comparison of two L1s. *System*, 100, 102567. Doi: <https://doi.org/10.1016/j.system.2021.102567>.
- Theoretical Research. (N.d.). <https://koppa.jyu.fi/avoimet/hum/metelmapolkuja/en/methodmap/strategies/theoretical-research>. (15 January, 2022).
- Thomas, J.J. and Cook, K.A. (2005). *Illuminating the Path. The Research and Development Agenda for Visual Analytics* [online]. [https://ils.unc.edu/courses/2017\\_fall/inls641\\_001/books/RD\\_Agenda\\_VisualAnalytics.pdf](https://ils.unc.edu/courses/2017_fall/inls641_001/books/RD_Agenda_VisualAnalytics.pdf). (23 February, 2022).
- Tuomi, J. and Sarajärvi, A. (2018). *Laadullinen tutkimus ja sisällönanalyysi* (6<sup>th</sup> ed.) Helsinki: Tammi.
- Valli, R. (2018). Vastausten tulkinta määrällisessä tutkimuksessa. In Valli, R., Aaltola, J. and Herkama, S. (eds.) *Ikkunoita tutkimusmetodeihin 2. Näkökulmia aloittelevalle tutkijalle tutkimuksen teoreettisiin lähtökohtiin ja analyysimenetelmiin* (5<sup>th</sup> ed.) Jyväskylä: PS-Kustannus.
- Valli, R. and Perkkilä, P. (2018). Sähköinen kyselylomake ja sosiaalinen media aineistonkeruussa. In Valli, R. and Aarnos, E. (eds.), *Ikkunoita tutkimusmetodeihin. 1, Metodien valinta ja aineistonkeruu. Virikkeitä aloittelevalle tutkijalle* (5<sup>th</sup> ed.) Jyväskylä: PS-Kustannus.



- Vuori, J. (N.d. - a). Laadullinen sisällönanalyysi. In Vuori, J. (ed.), *Laadullisen tutkimuksen verkkokäsikirja* [online].  
<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/analyysitavan-valinta-ja-yleiset-analyysitavat/laadullinen-sisallonanalyysi/>. (14 January, 2022).
- Vuori, J. (N.d. - b). Tutkimusetiikka ihmistieteissä. In Vuori, J. (ed.), *Laadullisen tutkimuksen verkkokäsikirja* [online].  
<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/tutkimusetiikka/tutkimusetiikka-ihmistieteissa/>. (15 February, 2022).
- Wardhaugh, R. (1970). The Contrastive Analysis Hypothesis. *TESOL Quarterly*, 4(2), 123–130. <https://doi.org/10.2307/3586182>.
- Wells, J.C. (1982). *Accents of English 3. Beyond the British Isles*. Cambridge: Cambridge University Press.
- Wiik, K. (1998). *Fonetiikan perusteet* (2<sup>nd</sup> ed.) Helsinki: WSOY.
- Wilkerson, M. (2013). The Sound of German. Descriptions of Accent by Native and Non-Native Listeners. *Die Unterrichtspraxis. Teaching German*, 46(1), 106–118. Doi: <https://doi.org/10.1111/tger.10132>.
- Whitman, R.L. and Jackon, K.L. (1972). The Unpredictability of Contrastive Analysis. *Language Learning*, 22(1), 29–41. <https://doi-org.ezproxy.jyu.fi/10.1111/j.1467-1770.1972.tb00071.x>.
- Zielinski, B.W. (2008). The listener. No longer the silent partner in reduced intelligibility. *System*, 36, 69–84. Doi: <https://doi.org/10.1016/j.system.2007.11.004>

# APPENDICES

## Appendix 1. The rating task form



JYVÄSKYLÄN YLIOPISTO  
UNIVERSITY OF JYVÄSKYLÄ

**Survey on Intelligibility and Comprehensibility of Finnish spoken English** E

**Dear Research Participant,**

I warmly welcome you to participate in this survey on the intelligibility and comprehensibility of Finnish spoken English. The survey is a part of my upcoming Master's Thesis in the University of Jyväskylä, Finland.

**INSTRUCTIONS:**

In this survey, you will hear audio files and be asked to either transcribe them (writing down what you hear) or evaluate them with 0–100 scale answers.

**To complete this survey successfully, please use either headphones or make sure that you are situated in a quiet place or surroundings with the possibility to listen to the audio samples.**

There are three (3) sections and **18 questions in total**.  
The survey will take approximately **15 minutes to complete**.

-----

Your participation in this study is completely voluntary. However, if you feel uncomfortable answering any questions, you can withdraw from the survey at any point. Your survey responses will be strictly confidential and data from this research will be reported in the aggregate.  
The participants will remain anonymous and your information will be coded and remain confidential.

If you have questions at any time about the survey or the procedures, you may contact me, Jarno Laitinen, at +358442890666 or by email at [jarno.o.laitinen@student.jyu.fi](mailto:jarno.o.laitinen@student.jyu.fi).

Thank you for your time and support, I very much appreciate your participation!

Sincerely,  
Jarno

Please start with the survey now by clicking on the **Start** button below.

**Start**

## Section 1 - Intelligibility

Transcribing audio samples.

In this section, you will hear six (6) short audio files and be asked to transcribe them.



Next

### Q0. Test question for your audio settings.

Please put your headphones on or make sure you have your sound system enabled.

Can you hear the following phrase clearly?



If **yes**, you may proceed to the actual survey starting from the next page.

If **no**, please adjust your audio settings (volume or connection) accordingly and return to this question.

Yes

No



Next

### Q1.

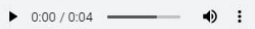
Listen to the audio file and **write down the phrase** as you hear it



**Important! Please listen to the file only once** as repeated listening may cause incomparable results.

### Q2.

Listen to the audio file and **write down the phrase** as you hear it



**Important! Please listen to the file only once** as repeated listening may cause incomparable results.

### Q3.

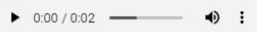
Listen the audio file and **write down the phrase** as you hear it



**Important! Please listen to the file only once** as repeated listening may cause incomparable results.

Q4.

Listen to the audio file and **write down the phrase** as you hear it



**Important! Please listen to the file only once** as repeated listening might cause incomparable results.

Q5.

Listen to the following audio and **write down the phrase** as you hear it



**Important! Please listen to the file only once** since repeated listening may cause incomparable results.

Q6.

Listen to the final audio sample of this section and **write down the phrase in it**



**Important! Please listen to the file only once** as repeated listening may cause incomparable results.



Next

## Section 2 - Comprehensibility

In this second section, you will hear six (6) slightly longer audio samples (around 20 seconds each) and rate their comprehensibility with seven (7) statements on a scale of 0-100 (strongly disagree – strongly agree)



Next

Q7.

Listen to the following audio sample and then **answer the questions**.

Note! You may listen the audio sample as many times as you like.



Next Question

On a scale of 0–100, **how much effort it took from you to understand the speaker?**  
The larger the number, the more the effort it took.

0 = *It took me no amount of effort at all*  
100 = *It took me the highest amount of effort*



On a scale of 0–100...

No answer = *neither agree or disagree*  
0 = *strongly disagree*  
100 = *strongly agree*

**the single speech sounds** (i.e., vowels, syllables, consonants) made the speaker more difficult to understand

**the pronunciation of individual words** made the speaker more difficult to understand

**the word-stress** (i.e., OBJECT vs. Object) made the speaker more difficult to understand

**the sentence-stress** (the word that was STRESSED in a sentence) made the speaker more difficult to understand

**the intonation** (falling/rising of the pitch) made the speaker more difficult to perceive

**the speech rhythm** (i.e., pauses, pace, tempo) made the speaker more difficult to understand

**the central message** of the speaker was difficult to understand

Strongly disagree

Strongly agree



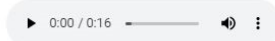
Do you have **any comments** on the sample?



Next

Q8.

**Listen to the following audio sample and answer the questions**  
Note! You may listen the sample as many times as you like.



Next Question

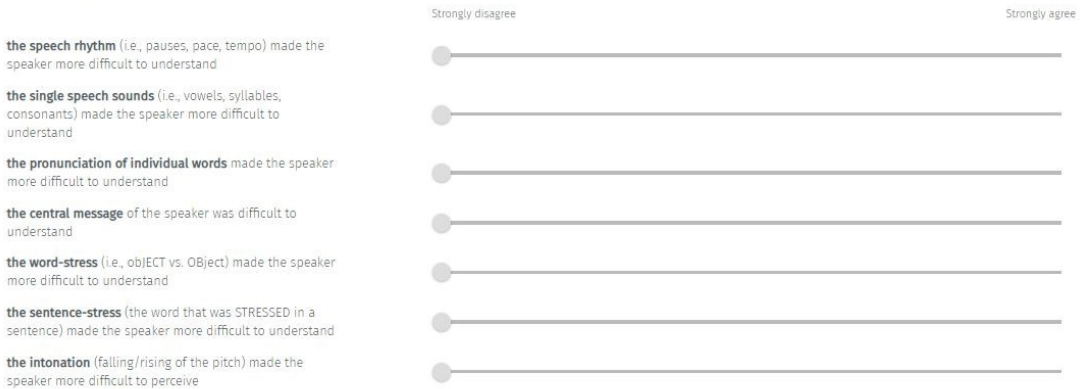
On a scale of 0–100, **how much effort it took from you to understand the speaker?**  
The larger the number, the more the effort it took.

*0 = It took me no amount of effort at all*  
*100 = It took me the highest amount of effort*



On a scale of 0–100...

*No answer = neither agree nor disagree*  
*0 = strongly disagree*  
*100 = strongly agree*



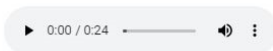
Do you have **any comments** on the sample?



Next

Q9.

**Listen to the following audio sample and answer the statements**  
Note! You may listen the audio file as many times as you like.



Next Question

On a scale of 0–100, **how much effort it took from you to understand the speaker?**  
The larger the number, the more the effort it took.

*0 = It took me no amount of effort at all*  
*100 = It took me the highest amount of effort*



On a scale of 0–100...

No answer = neither agree nor disagree

0 = strongly disagree

100 = strongly agree

**the central message** of the speaker was difficult to understand

**the intonation** (falling/rising of the pitch) made the speaker more difficult to perceive

**the sentence-stress** (the word that was STRESSED in a sentence) made the speaker more difficult to understand

**the word-stress** (i.e., OBJECT vs. Object) made the speaker more difficult to understand

**the speech rhythm** (i.e., pauses, pace, tempo) made the speaker more difficult to understand

**the pronunciation of individual words** made the speaker more difficult to understand

**the single speech sounds** (i.e., vowels, syllables, consonants) made the speaker more difficult to understand

Strongly disagree

Strongly agree

Six horizontal Likert scale sliders, each with a grey circle at the 0 position. The sliders are aligned with the seven categories of speech difficulty listed on the left.

Do you have **any comments** on the sample?

A large, empty text input field with a light grey background and a small cursor icon at the bottom right.

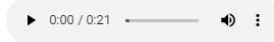


Next

Q10.

Listen to the following audio and answer the questions.

Note! You may listen the audio sample as many times as you want.



Next Question

On a scale of 0–100, **how much effort it took from you to understand the speaker?**

The larger the number, the more the effort it took.

0 = It took me no amount of effort at all

100 = It took me the highest amount of effort

Low amount of effort

High amount of effort

Effort it took to understand the speaker

A single horizontal slider with a grey circle at the 0 position, representing the effort required to understand the speaker.

On a scale of 0–100...

*No answer = neither agree nor disagree*

*0 = strongly disagree*

*100 = strongly agree*

**the sentence-stress** (the word that was STRESSED in a sentence) made the speaker more difficult to understand

**the word-stress** (i.e., obJECT vs. OBJect) made the speaker more difficult to understand

**the single speech sounds** (i.e., vowels, syllables, consonants) made the speaker more difficult to understand

**the pronunciation of individual words** made the speaker more difficult to understand

**the intonation** (falling/rising of the pitch) made the speaker more difficult to perceive

**the speech rhythm** (i.e., pauses, pace, tempo) made the speaker more difficult to understand

**the central message** of the speaker was difficult to understand

Strongly disagree

Strongly agree



Do you have **any comments** on the sample?

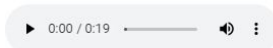


Next

Q11.

Listen to the following audio sample and answer the questions.

Note! You may come back and listen the sample as many times as needed.



Next Question

On a scale of 0–100, **how much effort it took from you to understand the speaker?**

The larger the number, the more the effort it took.

*0 = It took me no amount of effort at all*

*100 = It took me the highest amount of effort*

Low amount of effort

High amount of effort

Effort it took to understand the speaker





On a scale of 0–100...

*No answer = neither agree nor disagree*

*0 = strongly disagree*

*100 = strongly agree*

**the word-stress** (i.e., obJECT vs. Object) made the speaker more difficult to understand

**the single speech sounds** (i.e., vowels, syllables, consonants) made the speaker more difficult to understand

**the intonation** (falling/rising of the pitch) made the speaker more difficult to perceive

**the central message of the speaker** was difficult to understand

**the speech rhythm** (i.e., pauses, pace, tempo) made the speaker more difficult to understand

**the pronunciation of individual words** made the speaker more difficult to understand

**the sentence-stress** (the word that was STRESSED in a sentence) made the speaker more difficult to understand

Seven horizontal sliders, each with a grey circle at the start and a grey line extending to the right. The sliders are positioned to the right of the corresponding text items.

Do you have **any comments** on the sample?

A large, empty text input field with a light grey background and a thin border. A small cursor icon is visible at the bottom right corner.

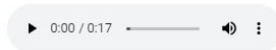


Next

**Q12. Final question of the section.**

**Listen to the following audio sample and answer the statements**

Note! You may listen to the audio as many times as you like.



Next Question

On a scale of 0–100, **how much effort it took from you to understand the speaker?**

The larger the number, the more the effort it took.

*0 = It took me no amount of effort at all*

*100 = It took me the highest amount of effort*

A horizontal slider with a grey circle at the start and a grey line extending to the right. The text 'Low amount of effort' is above the start and 'High amount of effort' is above the end. Below the slider, the text 'Effort it took to understand the speaker' is aligned with the start.

On a scale of 0–100...

No answer = neither agree nor disagree

0 = strongly disagree

100 = strongly agree

**the central message** of the speaker was difficult to understand

**the pronunciation of individual words** made the speaker more difficult to understand

**the single speech sounds** (i.e., vowels, syllables, consonants) made the speaker more difficult to understand

**the sentence-stress** (the word that was STRESSED in a sentence) made the speaker more difficult to understand

**the word-stress** (i.e., obJECT vs. Object) made the speaker more difficult to understand

**the intonation** (falling/rising of the pitch) made the speaker more difficult to perceive

**the speech rhythm** (i.e., pauses, pace, tempo) made the speaker more difficult to understand

Seven horizontal sliders, each with a grey circle at the 0 position, for rating the difficulty of understanding various aspects of speech. The sliders are aligned with the text on the left.

Do you have **any comments** on the sample?

A large, empty text input field for providing comments on the sample.



Next

**Section 3 - Background information.**

**In this final section of the survey, five (5) questions of relevant background information will be collected.**



Next

**Q13.**

What is your age?

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- Above 64

**Q14.**

Which country do you live in?

**Q15.**

In which state/city do you live in?

**Q16.**

Are you a native speaker of Finnish or English?

- Yes, Finnish
- Yes, English
- No, neither

**Q17.**

Do you speak any other languages? You may select more than one option.

- Finnish
- English
- Swedish
- Spanish
- France
- German
- Swahili
- Vietnam
- Norwegian
- Russian
- Estonian
- Japanese
- Chinese
- Other
- None

**Q18.**

Do you have an experience with Finnish spoken English, regarding intelligibility and/or comprehensibility, that you would like to share? Or do you wish to share an additional comment on some sample? Feel free to do so.

By clicking "Done", I give my permission to use this data anonymously in the study.

Thank you very much for completing this survey,  
I really appreciate your significant contribution and investment.

All the best,  
Jarno

## Appendix 2. The text used in audio samples

### COMMA GETS A CURE

Well, here's a story for you: Sarah Perry was a veterinary nurse who had been working daily at an old zoo in a deserted district of the territory, so she was very happy to start a new job at a superb private practice in North Square near the Duke Street Tower. That area was much nearer for her and more to her liking. Even so, on her first morning, she felt stressed. She ate a bowl of porridge, checked herself in the mirror and washed her face in a hurry. Then she put on a plain yellow dress and a fleece jacket, picked up her kit and headed for work.

When she got there, there was a woman with a goose waiting for her. The woman gave Sarah an official letter from the vet. The letter implied that the animal could be suffering from a rare form of foot and mouth disease, which was surprising, because normally you would only expect to see it in a dog or a goat. Sarah was sentimental, so this made her feel sorry for the beautiful bird.

Before long, that itchy goose began to strut around the office like a lunatic, which made an unsanitary mess. The goose's owner, Mary Harrison, kept calling, "Comma, Comma," which Sarah thought was an odd choice for a name. Comma was strong and huge, so it would take some force to trap her, but Sarah had a different idea. First she tried gently stroking the goose's lower back with her palm, then singing a tune to her. Finally, she administered ether. Her efforts were not futile. In no time, the goose began to tire, so Sarah was able to hold onto Comma and give her a relaxing bath.

Once Sarah had managed to bathe the goose, she wiped her off with a cloth and laid her on her right side. Then Sarah confirmed the vet's diagnosis. Almost immediately, she remembered an effective treatment that required her to measure out a lot of medicine. Sarah warned that this course of treatment might be expensive—either five or six times the cost of penicillin. I can't imagine paying so much, but Mrs. Harrison—a millionaire lawyer—thought it was a fair price for a cure.

*Associated text files are copyright and used by special arrangement with the International Dialects of English Archive at <https://www.dialectsarchive.com>. Copyright 2000. Douglas N. Honof, Jill McCullough & Barbara Somerville.*

### Appendix 3. The phrases used in intelligibility section

| <b>Speaker code</b> | <b>Sample phrase (content words bolded, auxiliaries within brackets)</b>      | <b>Context word count</b> |
|---------------------|---|---------------------------|
| <b>F4</b>           | <b>Sarah Perry</b> (was a) <b>veterinary nurse</b> .                          | 4                         |
| <b>CS</b>           | Even so, (on) <b>her first morning</b> , she felt stressed.                   | 8                         |
| <b>F1</b>           | <b>Her efforts</b> (were not) <b>futile</b> .                                 | 3                         |
| <b>F1</b>           | She ate (a) <b>bowl</b> (of) <b>porridge</b> .                                | 4                         |
| <b>F4</b>           | When she got there, <b>there</b> (was a) <b>woman</b> (with a <b>goose</b> ). | 7                         |
| <b>F2</b>           | Either five (or) <b>six times</b> (the) <b>cost</b> (of) <b>penicillin</b> .  | 6                         |

## Appendix 4. The passages used in comprehensibility section

Sample 7, speaker F2:

“Sarah warned that this course of treatment might be expensive – either five or six times the cost of penicillin. I can’t imagine paying so much, but Mrs. Harrison – a millionaire lawyer – thought it was a fair price for a cure.”

Sample 8, speaker F3:

“First she tried gently stroking the goose's lower back with her palm, then singing a tune to her. Finally, she administered ether. Her efforts were not futile.”

Sample 9, speaker F1:

“Before long, that itchy goose began to strut around the office like a lunatic, which made an unsanitary mess. The goose's owner, Mary Harrison, kept calling, "Comma, Comma," which Sarah thought was an odd choice for a name. Comma was strong and huge, so it would take some force to trap her, but Sarah had a different idea.”

Sample 10, speaker F2:

“Once Sarah had managed to bathe the goose, she wiped her off with a cloth and laid her on her right side. Then Sarah confirmed the vet’s diagnosis. Almost immediately, she remembered an effective treatment that required her to measure out a lot of medicine.”

Sample 11, speaker F1:

“Well, here's a story for you: Sarah Perry was a veterinary nurse who had been working daily at an old zoo in a deserted district of the territory, so she was very happy to start a new job at a superb private practice in North Square near the Duke Street Tower.”

Sample 12, speaker CS:

“When she got there, there was a woman with a goose waiting for her. The woman gave Sarah an official letter from the vet. The letter implied that the animal could be suffering from a rare form of foot and mouth disease, which was surprising, because normally you would only expect to see it in a dog or a goat.”