System 100 (2021) 102567

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s

Elina Tergujeff

University of Jyvaskyla, P.O. Box 35 (A), FIN-40014, Finland

ARTICLE INFO

Keywords:
Comprehensibility
Accentedness
Oral proficiency
Language assessment
Common European Framework of Reference for Languages
CEFR
English
Finnish
Finland-Swedish

ABSTRACT

Second language comprehensibility and accentedness are highly complex phenomena, and numerous studies have been conducted to better understand these constructs. However, research has seldom addressed L2 comprehensibility and accentedness in relation to speakers' proficiency in the target language. This study explores L2 English comprehensibility and accentedness across three proficiency levels. The speakers were 60 teenaged Finns, half of them speaking Finnish as their L1 and half speaking Finland-Swedish as their L1. Using 20-s speech samples, the comprehensibility and accentedness of the speakers were rated by 34 English-speaking teenagers on a 9-point scale. Comparisons were made regarding comprehensibility and accentedness between L1 Finnish and L1 Finland-Swedish speakers, and the data were also analysed for a possible relationship between the ratings and the speakers' overall oral proficiency in English. The results revealed a surprising difference between the language groups for speakers at proficiency levels B1–B2: Despite an equal proficiency assessment, L1 Finland-Swedish speakers were rated as more comprehensible and less accented than L1 Finnish speakers. Overall, the ratings were found to correlate with the speakers' oral proficiency, but the correlations were rather weak and the steps from one proficiency level to another often had only small effects on comprehensibility and accentedness.

1. Introduction

According to the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001), an independent language user (B2) is characterised by speech that is intelligible and does not require the listeners to strain themselves in order to understand the speaker. Hence, many learners aim to develop their intelligibility as well as ease of understanding as they climb the ladder of language proficiency. At the same time, researchers are working to untangle and understand the related phenomena of intelligibility, comprehensibility and accentedness (e.g., Munro & Derwing, 1995; Trofimovich & Isaacs, 2012). Understanding these phenomena has proved a complex task. The phenomena are partly related yet distinct: For example, a person with a noticeable foreign accent can still be highly comprehensible (e.g., Munro & Derwing, 1995). In addition, the phenomena are affected by various, partly over-lapping factors (see 2.1).

Derwing and Munro (2015) conceptualise *intelligibility* as the degree of match between a speaker's intended message and the listener's comprehension, or, actual understanding. Intelligibility is investigated with objective measures such as listeners transcribing speech samples in standard orthography (e.g., Field, 2005; Gass & Varonis, 1984; Kennedy & Trofimovich, 2008; Munro & Derwing, 1995). The term *comprehensibility* is used for the listener's ease of understanding a speaker, and is often studied with help of subjective

E-mail address: elina.tergujeff@jyu.fi.

https://doi.org/10.1016/j.system.2021.102567
Received 30 September 2020; Received in revised form 25 May 2021; Accepted 26 May 2021
Available online 11 June 2021

listener ratings using numeric scales (e.g., Saito et al., 2016; Saito, Webb, et al., 2016; Trofimovich & Isaacs, 2012). Finally, *accentedness* refers to a listener's perception of a speaker's strength of foreign accent, and is often studied together with comprehensibility, using scalar ratings (see references above). As reaching B2 proficiency on the CEFR scale entails ease of understanding, the speaker needs to have a good level of intelligibility and comprehensibility, whereas accentedness is of less importance.

The number of studies on these phenomena are increasing, but they seldom consider language proficiency as a factor. As such, the present study sets out to explore L2 comprehensibility and foreign accentedness of speakers across three oral proficiency levels. The investigation has a two-pronged focus: Firstly, listeners' perception of the ease of understanding speakers with varying oral proficiency, and secondly, listeners' perception of the strength of the speakers' foreign accent. Exploring the link between overall oral proficiency and comprehensibility/accentedness will add to our current understanding of L2 development and the role of comprehensibility and accentedness within L2 assessment. As for the context of the present study, this research investigates L2 English spoken by two groups of teenaged learners from Finland: L1 Finnish and L1 Finland-Swedish Finns. Previous empirical research and contrastive phonetics have broadened knowledge about the typical challenges that Finns have regarding spoken English (e.g., Lintunen, 2004), but very little is known about how listeners perceive these challenges when realised in speech. Moreover, Finns' English has not previously been subjected to comprehensibility and accentedness ratings.

2. Literature review

2.1. Complexity of L2 comprehensibility and accentedness

Second language comprehensibility and accentedness are highly complex phenomena. Several studies have focussed on deconstructing them to find out which aspects of speech are linked with or explain comprehensibility and accentedness ratings. As for comprehensibility, these investigations point to the importance of suprasegmentals such as speech rate, pausing, stress and intonation. According to Kang et al. (2010), suprasegmental features alone can collectively account for 50% of the variance in comprehensibility ratings. Studies have also discovered that while many speech features are linked with both comprehensibility and accentedness, segmental accuracy is more strongly associated with accentedness.

For example, Trofimovich and Isaacs (2012) discovered that several speech features correlated with both comprehensibility and accentedness in L1 French speakers' English, but segmental accuracy only correlated with accentedness. Saito and colleagues (2016) obtained similar findings regarding L1 Japanese speakers' English: while comprehensibility was linked to all linguistic domains, accentedness was strongly related to pronunciation – especially segmental accuracy. Based on an investigation, in which speech samples were manipulated to combine native suprasegmentals with non-native segments and vice versa, Winters and O'Brien (2013) conclude that segmental cues seem to contribute to accentedness more than suprasegmentals.

The complexity of L2 comprehensibility and accentedness is highlighted by the fact that the aforementioned pronunciation and fluency features are not the only factors in these phenomena. For example, research has shown that grammatical aspects (Isaacs & Trofimovich, 2012; Saito, Trofimovich, et al., 2016), hesitation (O'Brien, 2014), various lexical variables (Isaacs & Trofimovich, 2012; Saito, Webb, et al., 2016) and semantic context (Kennedy & Trofimovich, 2008) also play a role in L2 comprehensibility and accentedness. In addition, task type (Bergeron & Trofimovich, 2017), listening conditions (e.g., Volín & Skarnitzl, 2010), non-native listeners' language proficiency in the target language (Beinhoff, 2014) and attitudes/attitude manipulation (Taylor Reid et al., 2019) can also play a role.

Additionally, listeners' L1 may affect the perception of comprehensibility and accentedness. Some scholars have suggested that non-native listeners who share the speakers' L1 may find their peers' L2 speech more comprehensible and less accented (Gallardo del Puerto et al., 2015), but contradicting findings have been presented. Lima (2016) did not find a difference between comprehensibility or accentedness ratings given by non-native listeners who share and do not share the speakers' L1. Moreover, Riney et al. (2005) revealed that listeners sharing the speakers' L1 may even be stricter in their ratings. Thus, definitive conclusions about the existence of a shared language benefit cannot be made, but an even more complex picture of L2 comprehensibility and accentedness is formed. To add to the complexity, Tokumoto and Shibata (2011) found that different language groups might value their own L2 accent differently. In their study, L1 Japanese and L1 Korean speakers were rather negative about their accent in English, whereas L1 Malaysian speakers appreciated Malaysian-accented English. As stated above, attitude (manipulation) has been found to affect the perception of comprehensibility and accentedness (Taylor Reid et al., 2019).

2.2. Second language proficiency and comprehensibility/accentedness

Comprehensibility and accentedness are often considered in language proficiency assessment. For example, comprehensibility is among the descriptors of B2 proficiency in the CEFR (Council of Europe, 2001), and the CEFR companion volume with new descriptors for phonological control (Council of Europe, 2018) considers foreign accent at each proficiency level except for the lowest (A1). Consequently, comprehensibility and accentedness can be expected to be in line with spoken language proficiency, even though criteria-based language assessment and subjective ratings are completely different ways of evaluating L2 speech. However, Isaacs and Trofimovich (2012) argue that there are several shortcomings in the way that comprehensibility and pronunciation has been modelled in assessment context. In their view, pronunciation and accent are often conflated and the treatment of pronunciation is inconsistent and vague. Hence, they propose distinct guidelines for L2 comprehensibility scale development to supplement existing assessment criteria (Isaacs & Trofimovich, 2012).

Scholars have seldom taken an interest in the connection between proficiency and comprehensibility/accentedness by considering

speakers' L2 proficiency as a variable in their research. Beinhoff (2014) considered non-native *listeners'* language proficiency in a study that focussed on intelligibility, comprehensibility and accentedness, finding that the listeners' proficiency is an influential factor. Regarding speakers, Isaacs and Trofimovich (2012) and Saito and colleagues (2016) distinguished between three "ability levels" in their investigations of linguistic correlates of L2 comprehensibility and accentedness, but the levels were based on the global ratings received by the speakers instead of proficiency assessment.

One study that considers a criteria-based language proficiency assessment as a speaker variable is by van Maastricht et al. (2016). In this study, L1 speakers of Dutch ($n = 41$) rated three Dutch-speaking groups for comprehensibility, accentedness and nativeness: L1 speakers ($n = 4$), proficient L2 speakers ($n = 4$) and less proficient L2 speakers ($n = 4$). The proficient L2 speakers were determined to have a CEFR level of B2 or above, whereas the less proficient L2 speakers were at level A2 or below. The results reveal that L1 speakers of Dutch were considered more comprehensible and nativelike, and less accented than the L2 speakers. A similar tendency was detected between more and less proficient L2 speakers, but as for comprehensibility, the difference between the L2 groups did not reach statistical significance at 0.05 level ($p = .058$). Accordingly, the study does not offer conclusive results. Hence, the links need to be confirmed in further investigations such as the present study.

2.3. L2 English spoken by Finns

Finnish-speaking and Swedish-speaking Finns' English was chosen for the present study in light of the unique possibility for comparison that they offer. Both Finnish and Swedish are national languages in Finland. Children and youth attend either Finnish- or Swedish-medium education, but all teaching follows the same national core curricula, e.g., the national core curriculum for basic education. As a result, possible differences in the English of these two speaker groups are likely to do with the influence of the L1 (Asunmaa, 2020), rather than differences in education, for example. Swedish-speaking Finns have been previously found to be more successful in their English studies (e.g., Härmälä et al., 2014), which may be due to Swedish being a Germanic language like English, whereas Finnish as a Finno-Ugric language is typologically more distant from English.

Previous research on English spoken by Finns has mainly dealt with Finnish-speaking Finns, who may be prone to issues of comprehensibility and a detectable foreign accent. In a study by Paananen-Porkka (2007), English-speaking listeners evaluated Finnish-speaking adolescents' speech rhythm in English, describing it as "jerky" (p. 288) for some of the speakers. Listeners also characterised some of the speakers as disturbingly hesitant, speaking too slowly and pausing in an unusual way. Similarly, Pihko's (1997) English-speaking listeners drew their attention to L1 Finnish-speakers' prosodic features such as rhythm, slow speech rate and pausing, and to a few segmental features, e.g., the pronunciation of /v/. Morris-Wilson (1999) also found English-speaking listeners intolerant towards /v/ mispronounced as /w/ by L1 Finnish speakers. Even Finns themselves may consider Finnish-accented English relatively difficult to understand, reporting difficulty in focusing and irritation due to fluency issues (Pihko, 1997). In the same study, Finnish speakers' negative attitude towards Finnish-accented English was evident.

2.4. Adult vs. young participants in listener tasks

The present study is focussed on L2 English spoken by teenaged language learners. The choice of speakers derives from the author's longitudinal venture to develop English pronunciation teaching within the educational system of Finland. The work was started by mapping the contents and practices of pronunciation teaching (Tergujeff, 2013). Investigating teenagers' L2 comprehensibility and accentedness in relation to the speakers' language proficiency will provide useful further information that can be utilised in developing teaching and assessment. In addition to teenaged speakers, teenaged listeners were opted for in the present study. This was based on a real-life orientation: teenagers from Finland would most likely use English to communicate with fellow teenagers, for example in contexts of international study exchange and social media.

Choosing to work with teenagers was a conscious risk. Only a few previous studies have used under-aged raters, but researchers have not reported any major problems with their choice of participants. For example, Field (2005) reported no problems regarding teenaged native-speaker participants in an intelligibility test (dictation). Further, Butler (2014) does not discuss any difficulties with their Korean Year 6 students, who took part in an attitudinal questionnaire and intelligibility test. However, Norell (1991) noticed that a teenaged group (13–17-year-olds) left more blanks in an accent identification and comprehensibility rating task than an adult group. They also state that a number of participants had to be discarded due to filling in the answer sheet "out of step" (i.e., the number of the text passage on the tape did not match the number on the questionnaire, p. 14), but it is not clarified in which group these participants were. Similarly, Paananen-Porkka (2007) had to discard two out of 34 teenaged participants, because they had failed to fill in a questionnaire properly.

A few studies have addressed whether listeners' age affects the perception of foreign accent, and found no effect (Toivola, 2011; Uzal et al., 2018). However, these studies only investigated adult listeners. An exception is Scovel (1981), who noticed that 5–10-year-old children's ability to differentiate between a native-speaker accent and a highly proficient non-native speaker's accent increases with age. Scovel's study suggests that children as young as ten can quite reliably identify accents (97% correct). As a result, my reservations regarding teenaged raters did not pertain to their ability to give reliable ratings due to their young age, but rather to whether they would take the task seriously and be able to fully concentrate on it.

3. Aim and research questions

The present study explores L1 Finnish and L1 Finland-Swedish speakers' L2 English comprehensibility and accentedness across

three proficiency levels (A2, B1, B2). The aim is to obtain new knowledge about the comprehensibility and accentedness of Finns' English, and the relationship between these and the speakers' general spoken proficiency. The following research questions were set for the study:

1. How comprehensible is L2 English spoken by L1 Finnish and L1 Finland-Swedish speakers, as rated by English-speaking listeners?
2. How foreign accented is L2 English spoken by L1 Finnish and L1 Finland-Swedish speakers, as rated by English-speaking listeners?
3. Are there differences in L2 English comprehensibility/accentedness between L1 Finnish and L1 Finland-Swedish speakers? If yes, what kind?
4. What is the relationship between L2 comprehensibility/accentedness ratings and the speakers' overall L2 oral proficiency?

4. Methodology

To address the research questions, comprehensibility and accentedness ratings were undertaken with speech samples elicited from L1 Finnish and L1 Finland-Swedish speakers. English-speaking listeners rated the speech samples, providing an overall evaluation of the speakers' comprehensibility (RQ1) and accentedness (RQ2). To answer RQ3, the ratings were compared by language group. Before the ratings, the speakers had been assessed for their spoken proficiency in English (see 4.2). Relations between these proficiency assessments and the comprehensibility/accentedness ratings were analysed through Spearman's rho in order to answer RQ4.

4.1. Participants

The present study involved two groups of speakers and one group of listeners. The speakers were L1 Finnish ($n = 30$) and L1 Finland-Swedish ($n = 30$) teenagers, who were Year 9 students, aged 15 to 16. Participant selection was based on the students' English proficiency level, which had been determined in a national assessment (see 4.2). The present study used a total of 60 speakers from three proficiency levels and two language groups: 10 L1 Finnish and 10 L1 Finland-Swedish participants represented proficiency levels equivalent to A2, B1 and B2 on the CEFR scale[1] (Council of Europe, 2001). In addition, selection criteria included an even distribution of male and female participants. The language groups were determined based on background survey data linked to the national assessment. The groups were comprised of students who declared only one language (Finnish or Finland-Swedish) as their mother tongue and the language spoken at home.

As teenagers from Finland are likely to use their English skills for communicating with other teenagers, teenaged listeners were recruited. To avoid the possible influence of listeners' varied English proficiency (Beinhoff, 2014), English-speaking listeners were chosen over non-native listeners. However, the author disliked a strict definition of a native speaker, and chose to use teenagers with various linguistic and cultural backgrounds. Only the following criteria were applied: They attended their education in English and stated to speak English as one of their home languages. This decision was made based on a real-life orientation; the English-speaking people that Finnish youth come across in real life are often from diverse linguistic and cultural backgrounds.

The raters were recruited in a secondary school in the UK. They studied in Years 10 and 12, and their ages ranged from 14 to 17. In total, 38 students took part as raters in the listening task. They also completed a background questionnaire eliciting information about their linguistic background, experience in Finnish-accented and Swedish-accented English and if they had studied Finnish or Swedish. Three participants had frequent contact with Swedish speakers and one participant admitted to not speaking English at home. These four participants were discarded from the analysis. The total number of raters accepted for the present study was 34. All of them completed the comprehensibility and accentedness ratings on all 60 speakers, yielding a total of 4 080 ratings. As there were 11 blank responses, the results were ultimately based on 4 069 ratings.

4.2. Proficiency assessment

The speakers of the present study were among participants in a national assessment on students' English skills in Finland at end of their obligatory schooling (see Härmälä et al., 2014). The assessment was conducted by the Finnish National Agency for Education and the Finnish Education Evaluation Centre, who provided the author a research permit to the assessment data for spoken proficiency. The data consisted of monologic and dialogic speaking tasks that were video-recorded in the participants' schools. Students' proficiency was first assessed in terms of the Finnish application of the CEFR by their own teachers, after which 10% of the assessments were checked by trained raters at the Finnish Education Evaluation Centre.

The main objective of the national assessment was to map student achievement in relation to the learning goals of the national core curriculum. According to the curriculum that was in effect at the time (Finnish National Board of Education, 2004), the goal for Year 9 learners was to reach proficiency level A2.2 in English. The assessment scale did not consider proficiencies above B2.1, but merged the best students into the category "B2.1 or above". This may have been a result of the tasks not enabling a higher proficiency assessment (the tasks were designed to measure language proficiency A1–B1). In other words, participants who may have reached proficiency level B2.2 or even C1 were not apparent in the results of the assessment. As this was problematic for the purposes of the present study, the B2-level speakers' speech materials were re-evaluated by two experts, and none of the speakers were found to have obtained

[1] A finer scale was used in the original assessment, with each level divided into two, i.e., A2.1, A2.2, B1.1, etc.

C1-level proficiency.

4.3. Speech materials

The speech samples used in this study were extracted from video recordings of a semi-spontaneous monologue task (Härmälä et al., 2014). Audio files were first separated from the original video files at the National Education Evaluation Centre, as the author's research permit only applied to the audio material. The speech samples consisted of approximately 20-s samples of the task, in which the speakers were required to speak freely about themselves, their families, hobbies and summer plans. As a result, the samples could not be extracted from the beginning of the speaking task like in many previous studies, as this would have disclosed identifiable information about the speakers. The chosen samples were therefore extracted from the middle or towards the end of the speaking task. Careful consideration was used to avoid decontextualisation and fragmentation of the samples.

The length of speech samples used within comprehensibility and accentedness research varies greatly. Some studies have used speech samples shorter than 10 s (e.g., Munro et al., 2006), whereas others have relied on samples longer than 1 min (e.g., Lima, 2016). Unfortunately, there is no research available on how speech sample length might influence the ratings, but most studies have used 20- or 30-s samples. Twenty-second samples were chosen for the present study, following e.g., Isaacs and Thomson (2013), O'Brien (2014) and Trofimovich et al. (2017). With 60 speech samples to rate, the total duration of the listener task was an important factor in the choice between 20- and 30-s speech samples. Opting for 30-s samples would have increased the total duration by 10 min, which was not considered ideal.

4.4. Procedure

Comprehensibility and accentedness ratings were conducted with the listeners in two groups. This facilitated the presentation of the speech samples in two different orders to minimise the possible effects of fatigue and boredom towards the end of the rating session, and the novelty of the task in the beginning. To the first group of listeners, the speech samples were presented in a randomised order. Then the samples were arranged in four sets, out of which the ones that were presented to the first group of listeners as first and last were presented in the middle to the second group of listeners. Randomising the order for both groups might have resulted in same speech samples being presented among the first or last samples to both listener groups, and was thus avoided. Before the actual ratings, the listeners got to practise with two test samples so that they would know what to expect and how to use the rating scale.

Speech samples were presented to the listeners via loud speakers in a regular classroom. No noise interference occurred. The volume was adjusted before the actual ratings. The researcher manually set the pace of presenting the speech samples, making sure that the listeners had enough time to give their ratings, as it was felt that this was a better option than setting a fixed interval between the samples. Estimating a suitable interval would have been difficult, and may not have been long enough for all participants (cf. Jeong et al., 2017).

Listeners heard each speech sample only once, after which they rated the speaker's comprehensibility (i.e., how easy or difficult to understand) and foreign accentedness (i.e., accent strength) on separate 9-point numeric scales. The choice of a 9-point scale followed several previous studies (e.g., Munro & Derwing, 1995), and offers the advantage of allowing enough choice for the rater (see Isaacs & Thomson, 2013, for a discussion on different scales). On the scale, 1 stood for "very easy to understand" and "no foreign accent", and 9 stood for "very difficult to understand" and "very strong foreign accent". In other words, the lower the rating, the more comprehensible and less accented the speaker. Based on O'Brien (2016), rating both comprehensibility and accentedness after hearing the speech sample only once can be considered at least as effective as rating the continua separately.

Listeners gave a pen-and-paper response by circling their ratings on the scale. They were instructed to do the task individually and not to peek at each other's ratings. The rating session took 45 min including instructions and a small break halfway through the ratings to prevent fatigue and boredom. Here, it was considered desirable that the listener test takes less than an hour, even with a break in the middle. During the break, the participants were not allowed to compare ratings, but a general discussion about Finland was led by the researcher.

4.5. Analyses

The possible connection between the ratings given by the listeners ($n = 34$) and the speakers' oral proficiency level were investigated with Spearman's rho. In addition, independent sample *t*-tests were performed to compare ratings between the proficiency levels. Also, differences between ratings received by L1 Finnish and L1 Finland-Swedish participants were tested for statistical significance using the *t*-test. Effect sizes were calculated using Cohen's *d* (Cohen, 1988). Effect size reveals the effect of the difference, whereas the *t*-test alone can only explain whether or not the difference can occur by chance. In interpreting the effect size, Cohen's (1988) guidelines were followed: Small effect ($d = > 0.2$), medium effect ($d = 0.5$–0.8), large effect ($d = > 0.8$). However, the interpretations were considered cautiously when drawing conclusions, because Plonsky and Oswald (2014) have found that Cohen's (1988) labels underestimate the range of effects that are typical within L2 research. They suggest that *d* values around 0.40 should be considered small, 0.70 medium and 1.00 large.

As there were some initial doubts about using teenagers as raters in the present study, rater agreement was investigated with the Intraclass Correlation Coefficient (ICC). Using an absolute agreement definition, the average measures ICC proved .959 for comprehensibility and .917 for accentedness. As Cicchetti (1994) determines ICC values between 0.75 and 1.00 to signify excellent agreement, the results show that the teenaged listeners were remarkably unified in assigning ratings. Thus, the ratings can be deemed reliable.

5. Results

5.1. Comprehensibility

Comprehensibility ratings varied greatly between the speakers. Speaker averages varied from 1.38 to 7.71 on the 9-point scale; A2-level speakers' averages ranged from 3.38 to 7.71, B1-level speakers' from 2.53 to 6.26, and B2-level speakers' from 1.38 to 6.41. The total average was 4.46. Fig. 1 demonstrates the average ratings by proficiency level and language group. In addition, averages for all speakers have been added (see discussion in 5.3). At proficiency level A2, L1 Finland-Swedish speakers were rated as slightly easier to understand than L1 Finnish speakers (average 5.28 vs. 5.53), but the difference was not statistically significant ($t = 1.733$, $p = .084$). At higher proficiency levels, L1 Finland-Swedish speakers were rated as clearly more comprehensible than L1 Finnish speakers: The average ratings at B1 level were 3.56 vs. 5.09, and at B2 level they were 2.94 vs. 4.34. These differences were statistically significant ($t = 11.458$, $p = .000$ and $t = 9.855$, $p = .000$). Difference at the B1 level proved to have a large effect ($d = 0.88$), whereas the effect of the difference at the B2 level was medium ($d = 0.76$).

5.2. Accentedness

Similarly to comprehensibility, speakers' accentedness varied. Average ratings ranged from 2.91 to 6.82 on the 9-point scale; A2-level speakers' averages ranged from 4.85 to 6.82, B1-level speakers' from 3.82 to 6.41, and B2-level speakers' from 2.91 to 6.18. The total average was 5.31 – slightly higher than the total average for comprehensibility (4.46). The ratings are presented in Fig. 2 by proficiency levels and language groups. In addition, averages for all speakers have been added (see discussion in 5.3). Speakers from both language groups received nearly equal average ratings at A2 level (6.06 vs. 5.91), and no statistically significant difference was found ($t = 1.132$, $p = .258$). However, L1 Finland-Swedish speakers received better ratings compared to L1 Finnish speakers at levels B1 and B2: 4.74 vs. 5.69 and 4.13 vs. 5.32. Both differences were statistically significant ($t = 6.805$, $p = .000$ and $t = 8.231$, $p = .000$) and with medium effect sizes ($d = 0.52$ and $d = 0.63$ respectively).

5.3. Ratings and overall oral proficiency

As can be seen in Figs. 1 and 2, speakers with higher proficiency received better ratings for comprehensibility and accentedness on average, i.e., they were easier to understand and perceived as less accented. If we combine the ratings received by L1 Finnish and L1 Finland-Swedish speakers, the average ratings for comprehensibility were 5.40 (A2), 4.33 (B1) and 3.64 (B2); and for accentedness 5.99 (A2), 5.21 (B1) and 4.74 (B2). These results are presented in Figs. 1 and 2, along with language group specific averages per proficiency level. The figures demonstrate visually how the listeners clearly differentiated between A2 and B1-level Finland-Swedish speakers, whereas the differences between proficiency levels are smaller for the Finnish speakers.

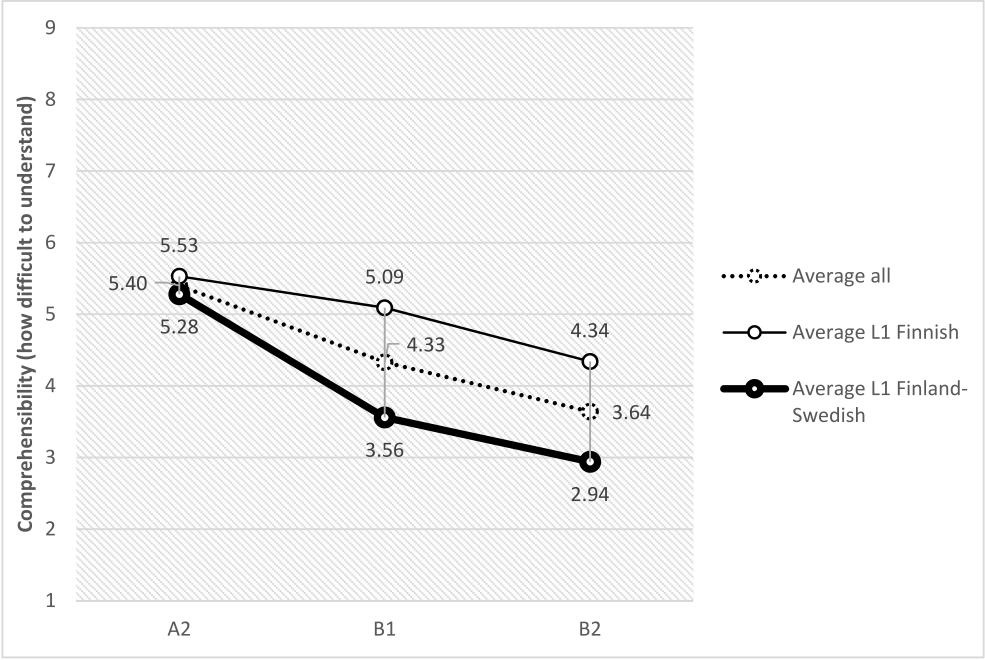

Fig. 1. Comprehensibility ratings received by L1 Finnish ($n = 30$) and L1 Finland-Swedish ($n = 30$) speakers of English at three proficiency levels (1 = very easy to understand, 9 = very difficult to understand).

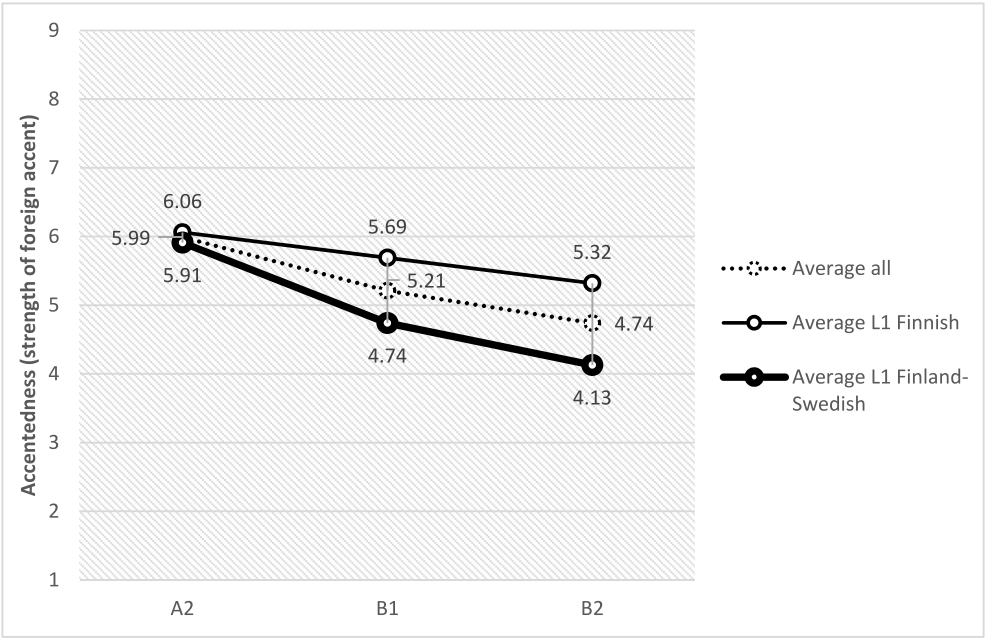

Fig. 2. Accentedness ratings received by L1 Finnish ($n = 30$) and L1 Finland-Swedish ($n = 30$) speakers of English at three proficiency levels (1 = no foreign accent, 9 = very strong foreign accent).

Further analyses revealed that all differences between comprehensibility ratings and the speakers' proficiency levels were statistically significant. However, variation occurred as for effect sizes and between language groups. When the language groups were combined, the effect sizes ranged from small to large: Between A2 and B1 speakers the effect was medium ($d = 0.57$), between B1 and B2 small ($d = 0.36$), and between A2 and B2 (two steps) large ($d = 0.91$). Considering L1 Finnish speakers only, the effect of the difference was small both between A2 and B1 speakers ($d = .25$) and B1 and B2 speakers ($d = .42$). The difference between A2 and B2 speakers proved to have a medium effect ($d = 0.64$). As for the L1 Finland-Swedish speakers, the effect was large between A2 and B1 speakers ($d = .91$), small between B1 and B2 speakers ($d = .31$), and very large between A2 and B2 speakers ($d = 1.25$). These results are presented in Table 1.

Similarly, accentedness ratings were distinct and statistically significant across the three proficiency levels (Table 2). All speakers combined, the difference between A2 and B1, and B1 and B2 speakers' accentedness had small effects ($d = 0.43$ and 0.25), while the effect was medium ($d = 0.68$) between A2 and B2 speakers. As for L1 Finnish speakers, the differences had small effect sizes ($d = .22$, 0.21, 0.44), whereas L1 Finland-Swedish speakers' difference reached a medium effect ($d = .62$) between levels A2 and B1, a small effect ($d = .31$) between B1 and B2, and a large effect ($d = 0.93$) between A2 and B2.

To further test the link between the ratings and proficiency, the data were analysed with Spearman's rho (Table 3), which confirmed a negative, statistically significant correlation between proficiency level and the ratings, concerning both comprehensibility ($r_s = -.345$) and accentedness ($r_s = -.260$). In other words, the higher the proficiency, the lower the rating score (= easier to understand and weaker accent). However, this tendency cannot be deemed strong: correlations between 0.20 and 0.35 are interpreted weak (Cohen et al., 2007, p. 536). The R^2 value for comprehensibility was 0.12 and for accentedness 0.07. These values indicate that only 12% and 7% of the ratings respectively can be attributed to the proficiency level. When analysed separately for the L1 Finnish and L1 Finland-Swedish data, the correlations proved weaker for L1 Finnish speakers. The relationship between proficiency and accentedness was especially weak for the Finnish speakers. Overall, the link to accentedness was weaker than the link to comprehensibility.

Table 1
Differences between comprehensibility ratings received by speakers with overall proficiency levels A2, B1 and B2.

	A2–B1 mean difference	*t*	*p*	*d*	B1–B2 mean difference	*t*	*p*	*d*	A2–B2 mean difference	*t*	*p*	*d*
All	−1.08	10.408	.000	.57	−0.69	6.544	.000	.36	−1.77	16.771	.000	.91
L1 Finnish	−0.44	3.244	.001	.25	−0.75	5.392	.000	.42	−1.20	8.319	.000	.64
L1 Finland-Swedish	−1.72	11.771	.000	.91	−0.62	4.601	.000	.31	−2.34	16.203	.000	1.25

Table 2
Differences between accentedness ratings received by speakers with overall proficiency levels A2, B1 and B2.

	A2–B1 mean difference	*t*	*p*	*d*	B1–B2 mean difference	*t*	*p*	*d*	A2–B2 mean difference	*t*	*p*	*d*
All	−0.78	7.830	.000	.43	−0.48	4.633	.000	.25	−1.26	12.462	.000	.68
L1 Finnish	−0.37	2.858	.004	.22	−0.37	2.762	.006	.21	−0.75	5.696	.000	.44
L1 Finland-Swedish	−1.17	8.124	.000	.62	−0.61	4.038	.000	.31	−1.77	12.044	.000	.93

Table 3
Correlations (Spearman's rho) and statistical significance between overall oral proficiency and L2 comprehensibility/accentedness in Finns' English.

	Comprehensibility	Accentedness
Proficiency (all)	$r_s = -.345$, $p = .000$	$r_s = -.260$, $p = .000$
Proficiency (L1 Finnish)	$r_s = -.249$, $p = .000$	$r_s = -.170$, $p = .000$
Proficiency (L1 Finland-Swedish)	$r_s = -.432$, $p = .000$	$r_s = -.351$, $p = .000$

6. Discussion

6.1. Comprehensibility and accentedness of Finns' English

The present study obtained new knowledge about how Finns' English is perceived by English-speaking listeners. Unsurprisingly, the ratings revealed huge variation between the speakers, as the speakers' overall oral proficiency in English varied from A2 to B2 on the CEFR scale (Council of Europe, 2001). Speakers tended to receive stricter ratings for accentedness than for comprehensibility, which supports earlier findings: The phenomena are related but distinct, and accented speech can be relatively easy to understand (e.g., Munro & Derwing, 1995).

The two language groups involved in this study received significantly different comprehensibility and accentedness ratings at proficiency levels B1 and B2. Accordingly, speakers with the same proficiency level can be perceived very differently by listeners, in terms of comprehensibility and accentedness, and the difference seems to surface first at the threshold level of independent language user (B1). Why this occurs language specifically in my data is curious and calls for further investigation. The unexpected outcome of the study implies that there is something in the L1 Finland-Swedish B1- and B2-level speakers' English that makes it sound more comprehensible and less accented than L1 Finnish speakers' with equal English proficiency; or vice versa, there is something in L1 Finnish speakers' English that makes it sound less comprehensible and more accented. The latter could include speech features such as segmental inaccuracy, unusual rhythm, dysfluent pausing and slow speech rate, which have been found to disturb listeners in L1 Finnish speakers' English (Morris-Wilson, 1999; Paananen-Porkka, 2007; Pihko, 1997) and to be linked with comprehensibility and accentedness (e.g., Kang et al., 2010; Trofimovich & Isaacs, 2012).

The possible effects of accent attitudes (e.g., Taylor Reid et al., 2019) cannot be completely overruled, even though the listeners indicated not being familiar with Finnish- and Swedish-accented English. The listeners may have had differentiating attitudes towards the accents of the two speaker groups, or they may have been more intolerant towards some specific speech features that were present in Finnish-accented English only, causing the difference in the ratings received by the two speaker groups.

6.2. Link between oral proficiency and L2 comprehensibility/accentedness

As previous research has seldom addressed comprehensibility and accentedness in relation to the speakers' overall proficiency in the target language, the focal point of the present study was to explore this relationship. A study by van Maastricht et al. (2016) found that more proficient L2 speakers are perceived as more comprehensible and less accented than less proficient L2 speakers. However, the result did not reach statistical significance for comprehensibility. When the present study investigated the issue with speakers across three proficiency levels, all results were statistically significant. However, many of the differences were small and so was their effect size. If Plonsky and Oswald's (2014) stricter interpretations were followed, some of the effects could be questioned altogether: differences between accentedness ratings received by B1 and B2-level speakers have effect sizes below 0.40, which Plonsky and Oswald view a suitable cut point for small effect in L2 research. The same applies to comprehensibility of L1 Finland-Swedish speakers between proficiency levels B1 and B2, and L1 Finnish speakers between levels A2 and B1.

It can be concluded that the present study confirmed van Maastricht and colleagues' (2016) observation with a statistically significant result, but uncertainties remain because of effect sizes. It seems that the higher the speaker's proficiency, the more comprehensible and less accented they are, but for the L1 Finnish speakers in particular, the steps from one proficiency level to another are rather small. When it comes to L1 Finnish speakers' accentedness, not even the difference between A2 and B2 speakers (two steps) reach the level of medium effect. At 0.44, it barely reaches the cut point for small effect in the interpretation of Plonsky and Oswald (2014). Correlation coefficients confirmed what was suggested above. Statistically significant correlations were found between proficiency and comprehensibility/accentedness, but the correlations were mainly weak.

This is noteworthy from the viewpoint of language assessment. Language assessment aims to offer information about language learners' proficiency and development to the learners themselves and to other relevant actors such as teachers and employers. Especially in the CEFR assessment (Council of Europe, 2001), the descriptors are very practical, focussing on describing how the learner can operate in the target language. As an example, the global CEFR scale describes a B2-level speaker as being able to speak so that it "makes regular interaction with native speakers quite possible without strain for either party" (p. 24). The criterion makes an obvious reference to comprehensibility and to real-life speakers; the kind that gave very different comprehensibility ratings to B2-level speakers of L2 English in the present study, depending on the speakers' L1.

With this in mind, it seems that the proficiency assessment has failed to fully consider the aspects that affect comprehensibility. The qualitative aspects to be considered for spoken language use at B2 level include a sufficient range of language; a relatively high degree of grammatical control; avoiding errors that cause misunderstandings; the ability to correct one's mistakes; speaking with an even tempo and avoiding noticeably long pauses; and coherent discourse (p. 28). As for phonological control, a B2-level speaker has acquired a "clear, natural pronunciation and intonation" (p. 117). One way of understanding the results of the present study is that the above criteria may not be sufficient when assessing oral proficiency with comprehensibility among the descriptors. It seems as though some speech features are perhaps not considered in the oral proficiency assessment, but nonetheless affect the evaluation of the speakers' comprehensibility and accentedness. As Isaacs and Trofimovich (2012) argue, the treatment of pronunciation in assessment context is often inconsistent and vague, which may partly explain why correlations between proficiency and comprehensibility/accentedness turned out weak in the present study. Considering the complex nature of comprehensibility (see 2.1), evaluating it as part of a proficiency assessment is no easy feat, but guidelines for assessing comprehensibility are offered in Isaacs and Trofimovich (2012).

It is worth noticing that the CEFR was updated after the proficiency assessments used in the present study were conducted; a companion volume with new descriptors (Council of Europe, 2018) was published in 2018. The companion includes e.g., a new set of descriptors for phonological control, which can help to place the speakers in proficiency levels more easily than before. The new descriptors are substantially more detailed than the previous ones, taking into consideration both segmental accuracy and suprasegmentals in a detailed manner (p. 136). If we consider the above-mentioned example of B2-level speakers being able to communicate without strain for either the speaker or the interlocutor (unchanged in the 2018 update bar the legitimate loss of "native speaker"), the new phonology descriptors are specific about some of the speech features that have been found to correlate with comprehensibility. For example, stress placement is emphasised in the new descriptors, and found to correlate with comprehensibility by e.g., Trofimovich and Isaacs (2012). In sum, assessing oral proficiency with the help of the new descriptors may give results closer to the listener ratings, and the lack of such detailed descriptors at the time of the present study may explain the found discrepancy.

6.3. *Reflections on the study and future directions*

Another approach could be to consider the present study as a test of how well the original assessment (Härmälä et al., 2014) succeeded in evaluating Finnish youth on the CEFR scale – especially from the viewpoint of equality between the two language groups, as there was variation between the groups in how well the proficiency assessments correlated with comprehensibility and accentedness. One could ask if the L1 Finland-Swedish speakers, who are known to succeed better in their English studies, were perhaps required more to reach B1–B2 proficiency. This could explain better comprehensibility and accentedness ratings.

In addition, attitudes towards foreign accent can be different among language groups (Tokumoto & Shibata, 2011), which could have affected the assessment. One has to bear in mind that the proficiency assessments were originally conducted by the speakers' own teachers, who quite likely belong to the same language group as their students. If Finnish-speaking teachers have been more critical towards Finnish-accented English than the Finland-Swedish-speaking teachers towards Finland-Swedish-accented English, a difference might occur. However, 10% of the speakers' tests were checked at the Finnish Education Evaluation Centre (FEEC) by two trained raters. The agreement between teachers and the trained raters at FEEC was high ($r = 0.79$–0.84; Härmälä et al., 2014), which speaks to the reliability of the assessments.

It is also noteworthy that the oral proficiency assessments were conducted based on longer stretches of speech than the comprehensibility and accentedness ratings. Selecting the extracts for the present study plays a role in the outcome of the ratings. As described in 4.3, the present study used 20 s of the speakers' monologue task, avoiding all personal details they may have revealed in the task. It needs to be acknowledged that the extract may not have represented the speaker's overall proficiency, but may have given a better or worse impression. This limitation needs to be considered by not drawing too definite conclusions about the results. Future studies on comprehensibility and accentedness across different proficiency levels could look into the possibility of planning the speech data so that the exact same materials could be used for both oral proficiency assessment and comprehensibility/accentedness ratings.

Considering the complex nature of comprehensibility and accentedness, it should also be acknowledged that no definite explanations for the results can be presented without thorough analyses of the speech samples. The differences between the language groups may lie in lexis, grammatical aspects, fluency, pronunciation, etc. Additionally, the language proficiency assessments are not based on sole comprehensibility and accentedness, but several other aspects of language use have naturally been considered in determining the speaker's proficiency. However, the present study points to an unclear relationship between overall oral proficiency and comprehensibility/accentedness, and raises questions to be tackled in future research. The natural next step is to analyse how the English of these two language groups differs, and to find connections between those speech features and comprehensibility, accentedness and overall oral proficiency. More generally, research-based knowledge is needed on the role of comprehensibility and accent in language assessment, and the possible effects of listener attitudes towards accented speech.

On a final note, it was considered important to use age-matched speakers and listeners in the study, and teenagers were recruited to

act as raters. This was considered a risk, as only a few previous studies have used young raters, and the participants' concentration spans and if they would take the task seriously was unknown to the researcher, who had never worked with the participants before. Fortunately, all concerns proved unnecessary, with no problems occurring in the rating sessions. The answer sheets were filled in correctly and with hardly any blanks (cf. Norell, 1991). Participants gave their ratings independently and without disturbing others. The ICC analysis of the ratings revealed a remarkable rater agreement. Based on the present study, teenagers can participate well as raters in comprehensibility and accentedness studies. However, careful attention should be paid to giving clear instructions, practising with test items and the length of the rating session. Limiting the length of a rating task to the maximum of 60 min can be considered a good rule of thumb.

7. Conclusion

This study set out to explore L1 Finnish and L1 Finland-Swedish speakers' L2 English comprehensibility and foreign accentedness, and their possible link to the speakers' overall oral proficiency in English. The study obtained new knowledge about the comprehensibility (RQ1) and accentedness (RQ2) of Finns' English, and about how they develop across three proficiency levels (see Figs. 1 and 2). Results revealed that comprehensibility and accentedness ratings given by English-speaking listeners correlate with the speakers' proficiency level (RQ4). However, the correlations were mainly weak, and there seems to be a discrepancy between ratings received by the two language groups (RQ3). Despite identical proficiency assessments, L1 Finland-Swedish speakers' English was rated more comprehensible and less accented than L1 Finnish speakers' English at proficiency levels B1 and B2. This implies there is something that differentiates the speaker groups from each other at these levels of proficiency.

Even though the issue needs to be confirmed in future studies, the results imply that speakers with an equal level of proficiency are not necessarily as equally accented or comprehensible. As for accentedness, the result is not surprising nor important, as accent does not necessarily compromise comfortable intelligibility. However, the weak link between overall oral proficiency and comprehensibility is more unexpected and curious from the language learning viewpoint. Therefore, language learners wishing to achieve good comprehensibility are advised to pay attention to speech features that have been found to correlate with the phenomenon, regardless of their overall oral proficiency level. Finally, the constituents of comprehensibility and accentedness could be more carefully considered within language assessment, as has already been implemented in the CEFR update (Council of Europe, 2018).

Author statement

Elina Tergujeff: Conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft, review and editing, visualization, project administration, funding acquisition.

Declaration of competing interest

None.

Acknowledgements

The author would like to thank Talia Isaacs for their help with arranging listener data collection, Maria Kautonen for their helpful expertise in the field of language assessment, Riikka Ullakonoja, Henna Heinonen and Maria Kautonen for their insightful comments on an earlier version of the manuscript, and Anastazija Kirkova-Naskova for initial inspiration. The author also wishes to express gratitude to all participants of this study, the Finnish National Agency for Education and the Finnish Education Evaluation Centre for collaboration, and to members of the project Fokus på uttalsinlärningen med svenska som mål- och källspråk (2015–2019) for collegial support and encouragement. The research has been funded by the Academy of Finland together with the University of Jyväskylä (project: Intelligibility, comprehensibility and accentedness of English spoken by Finns 2018–2021; grant number 315980), and the Swedish Cultural Foundation in Finland (2016–2017).

References

Asunmaa, K. (2020). Det går bra i engelskan i Svenskfinland – spekuleringar ur ett finlandssvenskt lärarperspektiv [Success in English for Swedish Finland – speculations from Finland-Swedish learner perspective]. In M. Härmälä, & M. Puukko (Eds.), *Taitotasoja ja suullista kielitaitoa. Näkökulmia englannin kielen oppimistulosten arvioinnin osoittamiin kehittämiskohteisiin* (pp. 73–85). Finnish Education Evaluation Centre. https://karvi.fi/app/uploads/2020/05/KARVI_Artikkelikokoelma_1_2020.pdf.

Beinhoff, B. (2014). Perceiving intelligibility and accentedness in non-native speech: A look at proficiency levels. *Concordia Work. Papers Appl. Linguistics (COPAL), 5*, 58–72.

Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals, 50*(3), 547–566. https://doi.org/10.1111/flan.12285

Butler, Y. G. (2007). How are nonnative-English-speaking teachers perceived by young learners? *TESOL Quarterly, 41*(4), 731–755. https://doi.org/10.1002/j.1545-7249.2007.tb00101.x

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Lawrence Erlbaum.

Cohen, J., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Council of Europe. (2018). Common European Framework of reference for languages: Learning, teaching, assessment. In *Companion volume with new descriptors*. Council of Europe. http://www.coe.int/lang-cefr.
Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.
Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399–423. https://doi.org/10.2307/3588487
Finnish National Board of Education. (2004). *National core curriculum for basic education*. Finnish National Board of Education.
Gallardo del Puerto, F., García Lecumberri, M.-L., & Gómez Lacabex, E. (2015). The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics, 25*(2), 202–224. https://doi.org/10.1111/ijal.12063
Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65–89. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x
Härmälä, M., Huhtanen, M., & Puukko, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013 [Learning outcomes in English at the end of basic education 2013]*. Finnish National Agency for Education.
Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgements of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*, 135–159. https://doi.org/10.1080/15434303.2013.769545
Jeong, H., Thórén, B., & Othman, J. (2017). Mutual intelligibility of Malay- and Swedish-accented English: An experimental study. *Indonesian Journal of Applied Linguistics, 7*(1), 43–53. https://doi.org/10.17509/ijal.v7i1.6857
Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of English language learner proficiency in oral English. *The Modern Language Journal, 94*, 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x
Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*(3), 459–489. https://doi.org/10.3138/CMLR.64.3.459
Lima, E. (2016). Comprehensibility and liveliness in nonnative student oral presentations before and after training: A mixed methods study. *System, 63*, 121–133. https://doi.org/10.1016/j.system.2016.10.003
Lintunen, P. (2004). *Pronunciation and phonemic transcription: A study of advanced Finnish learners of English. Anglicana turkuensia* (vol. 24). University of Turku.
van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Native speaker perceptions of (non-)native prominence patterns: Effects of deviance in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Communication, 83*, 21–33. https://doi.org/10.1016/j.specom.2016.07.008
Morris-Wilson, I. (1999). *Attitudes towards Finnish-accented English*. Doctoral dissertation, University of Stirling.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x
Munro, M. J., Derwing, T. M., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111–131. https://doi.org/10.1017/S0272263106060049
Norell, P. (1991). *Native-speaker reactions to Swedish pronunciation errors in English: Recognition, intelligibility and attitude*. Almqvist & Wiksell International.
O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning, 64*, 715–748. https://doi.org/10.1111/lang.12082
O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition, 38*, 587–605. https://doi.org/10.1017/S0272263115000418
Paananen-Porkka, M. (2007). *Speech rhythm in an interlanguage perspective: Finnish adolescents speaking English*. University of Helsinki.
Pihko, M.-K. (1997). *"His English sounded strange." the intelligibility of native and non-native English pronunciation to Finnish learners of English*. University of Jyväskylä Centre for Applied Language Studies.
Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079
Riney, T., Takagi, N., & Inutsuka, K. (2005). Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly, 39*(3), 441–466. https://doi.org/10.2307/3588489
Saito, K., Trofimovich, P., & Isaacs, T. (2016a). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholiguistics, 37*(2), 217–240. https://doi.org/10.1017/S0142716414000502
Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016b). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition, 19*(3), 597–609. https://doi.org/10.1017/S1366728915000255
Scovel, T. (1981). The recognition of foreign accents in English and its implications for psycholinguistic theories of language acquisition. In J.-G. Savard, & L. Laforge (Eds.), *Proceedings of the 5th congress of L'association internationale de Linguistique appliquée, montréal 1978* (pp. 389–401). Québec: Le Presses de L'Université Laval.
Taylor Reid, K., Trofimovich, P., & O'Brien, M. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition, 41*, 419–442. https://doi.org/10.1017/S0272263118000244
Tergujeff, E. (2013). *English pronunciation teaching in Finland. Jyväskylä Studies in Humanities 207*. University of Jyväskylä.
Toivola, M. (2011). *Vieraan aksentin arviointi ja mittaaminen suomessa [Assessing and measuring foreign accent in Finnish]*. University of Helsinki.
Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes, 30*, 392–408. https://doi.org/10.1111/j.1467-971X.2011.01710.x
Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905–916. https://doi.org/10.1017/S1366728912000168
Trofimovich, P., Kennedy, S., & Blanchet, J. (2017). Development of L2 French oral skills in an instructed setting: A focus on speech ratings. *Canadian Journal of Applied Linguistics, 20*, 32–50.
Uzal, M., Komulainen, E., & Aaltonen, O. (2017). The effect of some listener background factors and task type that contribute to degree of perceived accent ratings in L2 Finnish. In M. Kuronen, P. Lintunen, & T. Nieminen (Eds.), *Näkökulmia toisen kielen puheeseen – Insights into second language speech. AFinLA-e. Soveltavan kieliteteen tutkimuksia* (vol. 10, pp. 277–300). The Finnish Association for Applied Linguistics.
Volín, J., & Skarnitzl, R. (2010). The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication, 52*(11/12), 1010–1021. https://doi.org/10.1016/j.specom.2010.06.009
Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*, 486–507. https://doi.org/10.1016/j.specom.2012.12.006