

Kuronen, M., P. Lintunen & T. Nieminen (toim.) 2017. Näkökulmia toisen kielen puheeseen – Insights into second language speech. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2017 / n:o 10. 277–300.

Melike Uzal, Erkki Komulainen & Olli Aaltonen

University of Helsinki

The effect of some listener background factors and task type that contribute to degree of perceived accent ratings in L2 Finnish

This study evaluated the effect of some listener background factors—the listeners' gender, age, experience of teaching Finnish as a second language, frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area in Finland— and task type on their degree of perceived accent (DPA) ratings in L2 Finnish. The participants were 31 native-Finnish speakers and 40 nonnative speakers of Turkish origin who ranged in age from 7 to 66 as well as 61 Finnish listeners who evaluated the speech samples for a foreign accent using a 9-point scale. Three speech samples were administered (word pairs, the reading-aloud of sentences, and a spontaneous speech task). The results showed that no marked differences were observed, despite differences in the listeners' gender, age, Finnish as a L2 teaching experience, frequency of contact with immigrants speaking Finnish as an L2, and being a native of the Helsinki Metropolitan Area. The results also showed that ratings of native decreased with sample duration and extemporaneity.

Keywords: perceived accent; L2 phonology; listener background factors; task type effect

1 Introduction and previous research on task types used in foreign accent studies

Much research on second language (L2) phonological acquisition has revealed many factors that can affect the perception of foreign accent in a L2 (see Piske et al. 2001 and Moyer 2013 for a review). Of these factors, those that concern characteristics specific to speakers are termed **speaker-internal factors**, whereas those that concern characteristics not specific to speakers are termed **speaker-external factors**. Studies have found that the perception of foreign accent in L2 speech is affected by speaker-external factors, such as **elicitation techniques** and **listener characteristics** (Levi & Winters 2005). For instance, a few studies (Flege 1984; Piper & Cansin 1988; Munro & Derwing 1994) reported that the use of different elicitation techniques (extemporaneous versus read speech) was not a significant factor in relation to the degree of perceived foreign accent. In this respect, Munro & Derwing's (1994) study is revealing because they demonstrated that when extemporaneous narratives the speakers had produced before were compared to the read transcriptions of their extemporaneous narratives, no differences were found in their accentedness ratings. In contrast, Moyer (1999) detected a trend toward a higher number of native ratings for more isolated task items, so that reading isolated words received the lowest accent ratings, implying the highest native-like speech, which was followed by sentence reading, paragraph reading and free speech production. However, as no nonnative speaker scored significantly better on any particular task, this trend was not statistically significant.

Nevertheless, according to a study by Ishida (2004), this trend of more isolated task items receiving the highest number of native ratings was found to be statistically significant in that the discourse-length tasks (the reading-aloud of a paragraph and picture description) were rated to be more nonnative-like than the more isolated tasks (the reading-aloud of words and sentences). That is, based on previous studies by Ishida (2004) and Moyer (1999), one direction emerges: the more complex a task (from isolated words, to sentences, paragraphs, and picture description) is, the more nonnative-like learners' performances become. Again, however, contradictory findings to this generalization have also been reported. For instance, Oyama (1976), Snow & Hoefnagel-Höhle (1982), Thompson (1991), Munro & Mann (2005), and Toivola (2011) determined that spontaneous speech received the least accented ratings, implying the highest native-like speech when compared to read speech. The reason for the results of Munro & Mann (2005), however, might be their use of read speech samples that were artificially-created, pre-scripted, (words, sentences and a paragraph) and developed to be particularly difficult for Mandarin speakers of English. The same applies to the study by Thompson (1991), which incorporated the reading-aloud tasks of sentences and a passage seeded with

difficult sounds, and free oral production regarding the participants' daily routine. Thompson's (1991) results indicated that adult L2 learners were rated as most accented in the reading aloud task of sentences and least accented in the free oral production task. In short, when speech samples are seeded with difficult sounds for nonnative speakers, it inevitably results in a more pronounced accent (Munro & Derwing 1994). Moreover, as Larson-Hall (2006) argues, L2 speakers might purposely avoid the use of problematic sounds when they produce spontaneous speech samples. Another important point is that all of these studies except Snow & Hoefnagel-Höhle (1982) used reading tasks of paragraphs or sentences instead of a repetition task. When reading aloud, differences in reading ability cause the read speech of weaker readers to sound more accented and detection of their accent becomes easier (Piske et al. 2001). Reading aloud also relies on different skills than spontaneous speech and repetition, such as a higher degree of monitoring as well as literacy skills. Another factor may be related to orthographic differences in script or grapheme phoneme correspondence between the first language (L1) and the target language (Schmid & Hopp 2014). Consequently, another reason for the L2 speakers' accent being more marked for read speech samples in these studies might be differences in their reading ability. Munro & Mann (2005) identified read sentences as the most effective sampling type to rate accentedness, and considered extended speaking tasks as more representative of real communication than single words, but of course lexical, suprasegmental, and morpho-syntactic features inherent in such tasks could influence accentedness ratings. Overall, task-based effects for accentedness ratings in L2 speech have been documented in past research, but past research has provided conflicting results concerning the direction of these effects. This strongly suggests a need for additional research. Also, since there is no consensus as to the best sampling type that serves as a research instrument (Piske et al. 2001), this has resulted in the use of different speech samples in previous research to measure accentedness in L2.

1.1 Some listener background factors affecting degree of perceived accent ratings in an L2

Previous studies have also reported that the perception of foreign accent is affected by listener background factors (Levi & Winters 2005). However, due to conflicting research results, there is no consensus on which listener background factors affect listeners' degree of perceived accent (DPA) ratings or on the strength of their impact (Kang 2012). This points to a need for additional research to address the degree to which some listener background factors affect DPA ratings. For instance, McDermott (1986) concluded that listener background factors were not associated with any significant differences in the

overall assessment of foreign accent. Of the many listener background factors McDermott (1986) tested, age, sex and ethnic diversity of the neighbourhood of the listeners predicted accentedness ratings, with younger, male raters who live in ethnically homogeneous all-English-speaking neighbourhoods giving the strictest ratings (Schmid & Hopp 2014). Likewise, Toivola (2011) also found that listener background factors (age, gender, studying, frequency of contact with immigrants, Finnish as a L2 teacher status, being a native of the Uusimaa region in Finland) had no effect on listeners' accent ratings. In contrast, Kang (2008) discovered that accentedness ratings were affected by listeners' native speaker status, exposure to nonnative speakers (NNSs), training status, prior teaching experience, and attitudes toward accented English.

The following listener background factors are within the scope of the present study: the listener's age, gender, experience of the target language (teacher experience), foreign accent experience in the form of frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area in Finland. Few studies have been conducted on the relationship between listeners' age and their foreign accent ratings (Toivola 2011). To date, it seems that only McDermott (1986), Scovel (1981) and Toivola (2011) have analyzed the effect of listeners' age on accent ratings. McDermott (1986) found listener age to predict accentedness ratings, Toivola (2011) reported no significant effect for listener age and Scovel (1981) discovered that 5-10-year-old children were unable to perceive foreign accent as well as older children and adults. Similarly, very few studies have been also conducted on the effect of listeners' gender on accent ratings. Currently, it seems that only McDermott (1986) and Toivola (2011) have investigated this factor, Toivola (2011) reporting that gender had no effect on ratings. However, several previous studies have analyzed listeners' experience of the target language (limited specifically to teacher experience), and foreign accent experience and the results of these studies are discussed below.

1.1.1 Listeners' experience of being a language teacher and listeners' foreign accent experience

Studies on foreign accent vary in their operationalization of listener's **experience** of the target language. The literature offers various definitions of listeners' experiences of the target language, including formal training in language and linguistics (phonetics), familiarity with foreign-accented speech, or experience of language tutoring and teaching. As a consequence, when comparing the results of different studies, it is crucial to examine how the listeners' experience of the target language has been operationalized. In the context of this study, listener experience of the target language was operationalized as the effect of language tutoring and teaching experience. Hence, the construct

was measured using the dimension of teaching versus non-teaching. Let us now turn to discuss previous studies in which listeners' experience of the target language was similarly defined.

The literature on the effect of listeners' experience of the target language on accent ratings has produced study findings that have varied substantially. For instance, studies that have used both experienced teaching listeners and inexperienced non-teaching listeners have, on the one hand, determined that non-teaching listeners gave harsher accent ratings than teaching listeners (Barnwell 1989; Bongaerts et al. 2000; Kang & Rubin 2009; Hsieh 2011). For example, Hsieh's (2011) study found that when rating the foreign accents of international teaching assistants, the non-teaching American undergraduate listeners were more severe than experienced ESL teachers. On the other hand, Kang (2008) reported that teaching listeners with teaching experience in English as a second language were harsher in their accent ratings. In contrast to these findings, Bongaerts et al. (1997), Kennedy & Trofimovich (2008) and Toivola (2011) did not detect significant differences between the accentedness ratings of teaching and non-teaching listeners.

When it comes to the literature on the effect of listeners' foreign accent experience on accent ratings, Kennedy & Trofimovich (2008) demonstrated that the experience of nonnative speech did not result in harsher ratings of accentedness by seasoned ESL teachers and they did not rate accentedness differently from inexperienced non-teaching listeners who had had little or no contact with L2 speakers of English. Likewise, the teaching listeners in the study by Toivola (2011), who were teachers of Finnish as an L2 with frequent or daily contact with nonnative speakers of Finnish, did not rate accentedness differently from non-teaching listeners who had frequent, rare, or nonexistent contact with nonnative speakers of Finnish. Furthermore, in a study by Munro et al. (2010), native listeners' familiarity with foreign-accented speech through regular contact with nonnative speakers of English was not determined to have had a statistically significant effect on their accent ratings. In contrast, Thompson (1991) and Schmid & Hopp (2014) concluded that perception of a foreign accent depended on the listeners' familiarity with foreign-accented speech. Thompson (1991) found that listeners' linguistic experience had an effect on their ratings so that linguistically experienced listeners with frequent contact with nonnative speakers of English were more lenient in their ratings when compared to linguistically inexperienced listeners who had little or no contact with nonnative speakers of English. It should be noted, nonetheless, that Thompson's (1991) experienced listeners were not L2 teachers but language experts who had elected courses in linguistics. Moreover, Schmid and Hopp's (2014) findings showed that listeners' lower familiarity with foreign accents resulted in more variable and more strongly foreign-accented judgments.

2 The present study

2.1 The purpose of the study and research questions

The present article is part of a study by Uzal et al. (2015) and the purpose of that study was to explore the effect of some speaker background factors —the speakers' age of onset of extensive L2 acquisition (AO), L1 use, L2 use, home use of L1, L1 proficiency, L2 proficiency, the amount of L2 exposure indexed as length of residence in the target language country, age at the time of testing — on the degree of perceived accent ratings for child L2 learners of Finnish. The findings of Uzal et al. (2015) showed that AO was the main determiner of perceived accent, followed by home use of L1, and the amount of L2 and L1 use, confirming the salience of both age-related factor of AO and language experience factors in determining child L2 learners' foreign accent. The focus of this article is the effect of some listener background factors and task type. In short, the present study has two purposes. The first is to discover how some listener background factors affect listeners' perception of a foreign accent. The factors examined in the present study are the listeners' age, gender, Finnish-as-a-L2 teacher status, frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area in Finland. The second purpose is to identify the effect of three different speech samples (word pairs, the reading-aloud of sentences, and a spontaneous speech task) on the degree of perceived accent ratings.

The research questions of this article are the following:

1. What is the contribution of some listener background factors (listeners' gender, age, Finnish-as-a-L2 teacher status, frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area in Finland) to the degree of perceived accent ratings in L2 Finnish?
2. How do sampling effects (speech sample duration and speech sample extemporaneity) affect the degree of perceived accent ratings in L2 Finnish?

Regarding the first research question, due to the inconclusive research results in the literature, no stand is taken on the effects of listener background factors. For the second research question, the hypothesis is that different types of speech samples would cause listeners to give significantly different foreign accent ratings based on previous research. However, for the present analysis, it was preferred not to take a stand on how the effects of sample duration (sample duration scale: word pairs < single sentences < spontaneous speech) and extemporaneity of the speech samples (extemporaneity scale: scripted word pairs = scripted single sentences < extemporaneous sponta-

neous speech) would present themselves, as previous studies have reported conflicting findings.

2.2 Methodology

2.2.1 Data analysis

Reliability must always be estimated as it is a necessary (but not sufficient) condition to the validity of measurements. There are no valid measurements without a sufficient amount of reliability. It is a rule of thumb that reliability should be 0.7 or higher (Nunnally & Bernstein 1994). The problem is that there is a variety of reliability indices. Here, four were chosen:

1. parallel measures (items) approach
2. mean correlation between the listeners
3. intra-class-correlation (ICC, two-way mixed, single measurement) and
4. ICC two-way mixed, average measures

Differences in listener/speaker background factors were examined using ordinary one-way analysis of variance (ANOVA). Overall significance was expressed via F-ratio. Explanation of the variance in ratings analysis included initially the linear mixed model (LMM) approach. This analysis was expanded later to include the analysis of the generalized linear mixed model (GLMM) as well. Both methods allow to model the correlative structure of the data. The distribution of the response variable (foreign accent rating) was very strongly positively skewed (see Figure 2). Only GLMM fulfills both of the requirements: structure of the data and the way to treat the response variable distribution as ranked categories and use the link function -2LogLog . This model is usually called ordered categories mixed model regression. LMM and GLMM analyses produced almost the same results. This strengthens the reliability of the results. Also, Cohen's d was used as an index of effect size. It is simply the difference of two means divided by the pooled standard deviation of the groups compared.

2.2.2 Speakers

Speakers from four different groups were selected: child nonnative speakers (NNSs), adult NNSs, child native speakers (NSs) and adult NSs as controls. The part of the speaker data concerning the foreign accent ratings of 19 child NNSs and 20 child NSs as native control speakers comes from Uzal et al. (2015). Unlike Uzal et al. (2015), 32 more speakers were added to this study: 9 child NNSs, 11 adult NSs, and 12 adult NNSs. Thus, of the 71 speakers, 31 were NSs

of Finnish and 40 were NNSs of Finnish. All the native-Finnish controls were born and raised in Finland, lived in the Helsinki Metropolitan Area, and spoke standard Finnish. They consisted of 20 female speakers and 11 male speakers aged 7–39 ($M = 17$ years). All 11 adult NSs were recruited students from the University of Helsinki. All 20 child NSs were pupils at a comprehensive school in the Helsinki Metropolitan Area. The 20 child NSs (age at the time of testing: $M = 11$, $SD = 3$, range = 7–17) were matched to the 28 child NNSs according to their age at the time of testing ($M = 11$, $SD = 3$, range = 7–17). The ANOVA result showed that there was a statistically significant difference for the mean age at time of testing between the 12 adult NNSs (range = 27–66) and the 11 adult NSs (range = 22–39) ($F(1, 21) = 7.39$, $p = 0.013$), the mean being 38 and 28 years, respectively. That is, adult NNSs were significantly older when compared to adult NSs.

The NNSs were all NSs of Turkish from a wide variety of Turkish cities, including 22 female speakers and 18 male speakers aged 7–66 ($M = 19$ years). All 12 adult NNSs were acquaintances of the first author. All 28 child NNSs were pupils at various primary and secondary schools in the Helsinki Metropolitan Area, and most of them had received instruction in Turkish as a mother tongue two hours a week from the first author. To summarize, all 71 speakers resided in the Helsinki Metropolitan Area and spoke standard Finnish and they therefore fulfilled the most crucial requirement for speaker selection (standardized dialect; Long 1993).

2.2.3 Speech samples

The data sources, the speech sample collection and methods in the present study were predominantly the same as in Uzal et al. (2015). The differences were that the present study had 32 more speakers and an additional spontaneous speech sample task unlike Uzal et al. (2015). All speakers were assigned a task involving the repetition of eight sentences, from which five sentences and three word pairs were used for rating. The three word pairs were obtained from the remaining three sentences. The actual foreign accent rating task incorporated five sentences, three word pairs—obtained by extracting each word pair from the remaining three sentences—and spontaneous speech. Therefore, for the rating task, three types of speech samples were obtained from the adults (sentences, word pairs, spontaneous speech) and two types of speech samples (word pairs and sentences) from the children. All of these speech samples differed in duration: word pairs, sentences, and a 40-second passage of spontaneous speech on a topic selected by the speakers from three options. Also, all of these speech samples differed in extemporaneity: read word pairs, read sentences and the extemporaneous speech sample of spontaneous speech. Due to the children's short attention spans,

the scripted speech samples (word pairs and single sentences) were designed to be short and simple and contained the entire phonetic inventory of Finnish.

The speaker recordings for all children were made in empty school classrooms in the Helsinki Metropolitan Area, whereas for all adult speakers, the recordings were made in the soundproof recording studio at the University of Helsinki. A model voice of a female, monolingual, native Finnish adult was recorded in a soundproof recording studio beforehand, and this was presented to the speakers through a computer loudspeaker; this recording represented the spoken standard Finnish norm (Karlsson 2008). The speech materials of the five sentences and the three word pairs have been presented in Uzal et al. (2015). For instance, one of the five sentences presented to repeat was “Kotona on pöytä ja pöytävalaisin”, whereas one of the three sentences presented to repeat was “Kotona on porkkana ja tomaatti”. Then, “kotona on” part was deleted from this sentence and the remaining “porkkana ja tomaatti” word pair was presented for rating. Sentences were presented to speakers simultaneously in both written and aural form to reduce the potential of reading ability biasing accent ratings (Flege et al. 1995). Each sentence was presented once, followed by a silent six-second delay. The six-second delay was followed by a beep, after which the speakers repeated each sentence once. This six-second delay was intended to minimize the possible effect of direct imitation (Tench 1996). If the speakers could not produce a sentence or had forgotten the sentence altogether, the model was presented as many times as necessary to obtain a production without speech irregularities. Thus, the speakers had the opportunity to correct their productions; however, they were not allowed to practice beforehand. After this, unlike Uzal et al. (2015), an additional task of producing spontaneous speech was assigned to all the 11 NS adults and to 10 of the 12 NNS adults¹. The spontaneous speech instructions have been presented in Uzal et al. (2017). The speakers were instructed to discuss one of the three subjects (or make a subject up themselves). The speakers’ spontaneous speeches were recorded for 1 minute and the first 40-second segments were presented for rating. One of the three options for them to discuss was to tell their weekend or their daily routine (e.g. what do you usually do, when, with whom, for how long, what is interesting about it, etc.?). All the recordings were made on a Marantz PMD 660 digital audio recorder with a power microphone. The duration of the recordings ranged from 15 to 25 minutes for each speaker.

¹ As two NNS adults reported that they could not produce spontaneous speech due to their very poor oral language skills in Finnish, these speakers’ spontaneous speech samples were missing. That is, these two speakers who were both beginner level Finnish speakers were not recorded because they informed that they do not trust their oral language skills in Finnish enough to produce spontaneous speech in Finnish.

2.3 Listeners

The source of the listener data concerning the foreign accent ratings of 19 child NNSs and 20 child NSs as native control speakers is from Uzal et al. (2015). The requirement for the 61 native listeners recruited was that they be monolingual native speakers of Finnish and residents of the Helsinki Metropolitan Area. These listeners were subdivided into two groups according to their background education and teaching experience: non-teaching listeners (34) and teaching listeners (27). This ensured a balance in listeners' sensitivity to accents (Moyer 2007). Thus, the non-teaching listeners were NSs of Finnish with no experience of linguistics and phonetics, while the teaching listeners were linguistically experienced NSs of Finnish and were all teachers of Finnish as an L2. In addition, all the non-teaching listeners were required to be monolingual native speakers of standard Finnish, natives of the Helsinki Metropolitan Area, and residents of the Helsinki Metropolitan Area. All teaching listeners were also required to be residents of the Helsinki Metropolitan Area, but it sufficed that they were monolingual native speakers of Finnish so they did not have to be natives of the Helsinki Metropolitan Area.² Therefore, 17 of the 27 teaching listeners grew up outside the Helsinki Metropolitan Area, and 9 stated that they spoke Finnish with a regional dialect. This meant that there were 17 listeners who were not natives of the Helsinki Metropolitan Area and 44 listeners who were natives of the Helsinki Metropolitan Area. A majority of the listeners (52 of 61) spoke standard Finnish. In addition, all 61 listeners spoke Finnish as their L1 and had studied Swedish and English as their L2s. All listeners also reported normal hearing on the preliminary information form.

The gender distribution of the listeners was rather uneven (46 females, 15 males), primarily because the majority of teaching listeners (26 out of 27) were female.³ However, the 34 non-teaching listener group also had only 14 male listeners. The listeners' age distribution was also uneven: 13 listeners

² Speaking standard Finnish was not used as a selection criterion for the teaching listeners because they were recruited through an email sent to the Association of Teachers of Finnish as a Second Language and most importantly because it was very challenging to recruit such listeners in the first place. The ratings were all made in the University of Helsinki's soundproof recording studio, and due to this logistics of travelling, all teaching listeners were residents of the Helsinki Metropolitan Area. Consequently, 9 of the 27 teaching listeners reported that they spoke Finnish with a regional dialect. We were thus aware of the possibility that their regional accent background might affect their accent ratings. However, the ANOVA findings revealed that these 9 teaching listeners did not give statistically different accent ratings when compared to the other 18 teaching listeners (mean ratings = 3.0 and 2.5, $F(1, 25) = 3.463$, $p > 0.05$).

³ The gender and age distributions were uneven because it was extremely challenging to recruit teaching listeners, whom were recruited by sending an email to the Association of Teachers of Finnish as a Second Language. As a result, we received answers from 26 female L2 Finnish teachers and one male teacher. This imbalance in the gender distribution of the teaching listeners also led to the age distribution being uneven.

were aged between 19 and 24, 16 listeners were aged between 25 and 29, 12 listeners were aged between 30 and 34, and 20 listeners were aged between 35 and 64. The ages of the listeners ranged from 19 to 64, and their mean age was 35. The mean age of the non-teaching listeners was 27 (range 19–44), whereas the mean age of the teaching listeners was 42 (range 25–64). In other words, the teaching listeners were older than the non-teaching listeners, and the ANOVA result showed that the difference was statistically significant ($F(1, 59) = 27.11, p < 0.001$).

The listeners' foreign accent experience in the form of frequency of contact with immigrants was operationalized as the frequency of hearing the non-native Finnish speech of immigrants in their daily lives. The four reply options on the preliminary information form given to the listeners were as follows: never, rarely, often and daily. All the teaching listeners had contact with immigrants speaking Finnish as an L2 either daily ($n = 15$) or often ($n = 12$). A majority of the 34 non-teaching listeners reported having contact with immigrants either rarely ($n = 24$) or never ($n = 3$). Only a few non-teaching listeners stated that they had contact with immigrants either daily ($n = 4$) or often ($n = 3$). Moreover, as expected, the difference between the teaching listeners and the non-teaching listeners in terms of their frequency of contact with immigrants was statistically significant ($F(1, 59) = 44.685, p < 0.001$).

2.3.1 Procedure

All listeners completed the rating task individually in a soundproof recording studio, where a total of 589 speech samples (71 speakers \times 9 target recordings = 5 sentences + 3 word pairs + 21 adult speakers \times 1 spontaneous speech recording) were presented through headsets.⁴ Prior to the listeners' rating task, they were provided with a modified version of Toivola's (2011) preliminary information form (see Appendix A) and a short training session. First, they filled the preliminary information form and then read the instructions for the foreign accent listening test. To avoid unrelated linguistic factors affecting the DPA ratings, before the rating began, the listeners were instructed by the preliminary information form to ignore all nonphonological speech content and assess only DPA.

Since the main objective was to capture listeners' unguided holistic per-

⁴ Two different softwares were used for the rating task because it was recognized that using the Praat software, some of the samples were accidentally rated before they were fully heard. These 170 ratings were excluded from the analyses. 35 listeners did the rating task using the Praat software (Boersma & Weenink 2013), whereas the remaining 26 listeners performed the same rating task using the Presentation software (<https://www.neurobs.com/>) to prevent such premature ratings. The Presentation software forced the listeners to listen to each sample until the end before giving their rating, which they were asked to confirm before advancing to the next sample.

ceptions of foreign accent, Scovel's (1969) definition of foreign accent was adopted to guide listeners in that direction (see Appendix A). This holistic definition of foreign accent was used to avoid listeners from becoming confused during the foreign-accent-rating task with specialist linguistic terminology for the aspects of speech they were requested to rate. The listeners were also instructed by the preliminary information form to use the entire scale when rating the samples and were told to guess if they were uncertain. To help familiarize the listeners with the rating process and the range of accents, there were 30 practice speech samples of sentences (13 from Finnish children, 17 from Turkish children). These sentences were not analyzed.

A nine-point scale was used to rate accent. The listeners were instructed that they would hear productions spoken by NNSs or NSs of Finnish. They were requested to rate each production for the degree of perceived accent by pushing one of nine buttons representing a scale from one (*no foreign accent*) to nine (*very strong foreign accent*). The ratings were given in a single session that lasted between 60 and 100 minutes in three separate blocks. The listeners were allowed to take a short break in between the blocks and midway through the blocks of sentence and word pair rating. In all three types of rating tasks, the listeners were able to adjust the volume before the rating started; the same sample could be played up to five times, and the ratings could not be changed once given. The different sample types were divided so that the rating task of each block involved one sample type: only word pairs, only sentences, or only spontaneous speech. Runs were randomized within each sampling type, including speech samples and speakers. To balance ordering effects, the three blocks were presented in a randomized order.

3 Results

3.1 Reliability analysis

To investigate inter-rater consistency, four procedures were chosen. The results of reliability analyses are best summarized by producing a pictorial view of them. Since there was only one spontaneous speech sample per a speaker, split-approach could not be used. Instead, the 9 items were treated as (a) parallel measures. Correlation of the foreign accent ratings between the listeners gave the second way (b) mean correlation. It has the advantage that the reliability of individual listeners is available. Intra-class-correlation (ICC) is widely used in studies using ratings. Therefore, (c) ICC-single measure and (d) ICC-average measure were calculated. ICC-single measure was useful, but ICC-average measure gave unrealistically high values and did not discriminate between items. These features can be seen in Figure 1. As seen in Figure 1, spontaneous speech sample type as stimulus got the highest reliability, but the

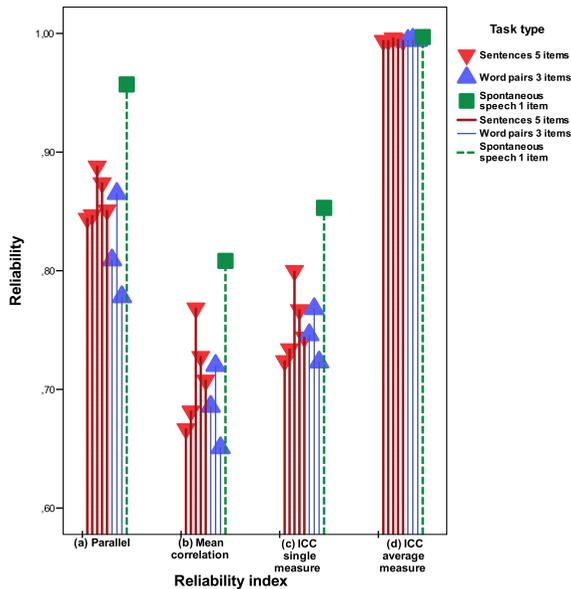


FIGURE 1. Four reliability indices of listeners' accent ratings according to three task types.

reliability differences among three task types were not big. The difference between word pairs and sentences was not very clear, but sentences had slightly higher reliability compared with word pairs. The listeners were more accurate, reliable and consistent in their rating of spontaneous speech samples compared with rating word pairs and sentences. The general conclusion was that all sample type measurements had high enough reliability to ensure validity in further analyses.

3.2 The effect of listener background factors on accent ratings

To assess the effects of the listeners' background factors (research question one) on accent ratings, both LMM and GLMM analyses were conducted. Initial analysis was LMM and final analysis was GLMM. Since GLMM analysis is slightly more correct approach to use with this kind of data at hand, GLMM findings were reported here. As shown in Table 1, the accent ratings were not affected by the listeners' gender, age, Finnish-as-a-L2 teacher status, foreign accent experience in the form of frequency of contact with immigrants or being a native of the Helsinki Metropolitan Area (all p values = n. s.). These five listener background factors' effect sizes (Cohen's d) ranged from 0.01 to 0.15, and could be all characterized as very small (Cohen 1988; Sawilowsky 2003). It was therefore justified to ignore the distinction between teaching and non-

TABLE 1. Statistical significances and effect sizes of listener characteristics on accent ratings (n = 61).

Variable		M	SD	GLMM overall p	Post hoc pairs	Cohen's d of pairwise differences
Gender	Female	2.48	2.10	=0.868		-0.04
	Male	2.57	2.18			
Teaching	No	2.69	2.22	=0.109		0.14
	Yes	2.47	2.10			
Age	15-24	2.43	2.03	=0.152	1-2	-0.11
	25-34	2.67	2.24		1-3	0.04
	35-63	2.34	1.99		1-4	0.15
Contacts	Never	2.57	1.96	=0.720	1-2	0.02
	Rarely	2.52	2.10		1-3	0.07
	Often	2.41	2.06		1-4	0.01
	Every day	2.54	2.22		2-3	0.05
					3-4	-0.06
Native of H. M. A.	No	2.49	2.10	=0.414		0.02
	Yes	2.54	2.16			

Note. Contacts refers to the frequency of hearing the nonnative Finnish speech of immigrants in listeners' daily lives measured by the four reply options (never, rarely, often, daily) in the preliminary information form given to the listeners. Native of H.M.A. stands for being a native of the Helsinki Metropolitan Area in Finland.

teaching listeners, because it transpired that Finnish as a L2 teaching experience had no significant effect on foreign accent ratings.

3.3 The effect of speech sample types on accent ratings

To assess the effect of task type (research question two) on accent ratings, both LMM and GLMM analyses were conducted. All pairwise post hoc comparisons were Bonferroni corrected. GLMM findings were reported in Table 2. As can be seen in Table 2, two out of three pairwise LMM comparisons had statistically significant differences. The two task types differed statistically very significantly ($p < 0.001$), but their effect sizes (d) ranged from -0.43 (small) to 0.15 (very small) (Cohen 1988; Sawilowsky 2003). That is, the two task types had a very significant effect on the accent ratings. Regarding the effect of sample duration, Table 2 shows that the duration of the sample (sample duration scale: word pairs < single sentences < spontaneous speech) affected the mean accent ratings, as word pairs were rated statistically more nativelike than single sentences ($p < 0.001$), though with a very small effect size. Besides, there

TABLE 2. Statistical significances and effect sizes of task type effects on accent ratings.

Task type	Focus	M	SD	p	pairs	comp.	d
1 Sentences	Dur.	2.60	2.16		1 < 2	< 0.001*	0.15
2 Word pairs	Dur., Ex.	2.28	1.95	< 0.001*	1 = 3	= 0.999	-0.25
3 Spontaneous	Dur., Ex.	3.15	2.73		2 < 3	< 0.001*	-0.43

*. The mean difference is significant at the 0.001 level.

Note. "p" stands for GLLM overall p, "pairs" for post hoc pairs, "comp." for pairwise post hoc comparisons and "d" for Cohen's d of pairwise differences, "Dur." for duration, and "Ex." for extemporization.

was a durational trend to rate single sentences more nativelike than spontaneous speech. However, this trend of durational effect reached statistical significance only between shorter word pairs and longer spontaneous speech samples with a small effect size. Thus, there was a durational trend: the longer the sample type length, the less native the mean accent rating. The same statistical difference between spontaneous speech samples and word pairs meant that the extemporaneity of the sample (extemporaneity scale: scripted word pairs = scripted single sentences < spontaneous speech) affected the mean accent ratings as well. That is, the less extemporaneous the sampling type, the more native the mean accent rating: spontaneous speech samples were rated less nativelike than word pair samples ($p < 0.001$) the effect size being small. Also, the purpose of Figure 2 was to show how extremely skewed foreign accent rating distributions were, which caused means and standard deviations to give false impressions.

4 Discussion

This study evaluated the effect of some listener background factors, such as gender, age, Finnish as a L2 teaching experience, frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area in Finland, and task type on the degree of perceived accent ratings in L2 Finnish. As for the first research question, the results revealed that none of these listener background factors had any statistically significant effect on accent ratings. Thus, it seems that when both speakers and listeners speak the same target language and the same variety—standard Finnish in this study—foreign accent detection is reliable and accurate, and no marked differences can be observed, despite differences in the listeners' gender, age, Finnish as a L2 teaching experience, frequency of contact with immigrants speaking Finnish as an L2, and being a native of the Helsinki Metropolitan Area. The findings established that native

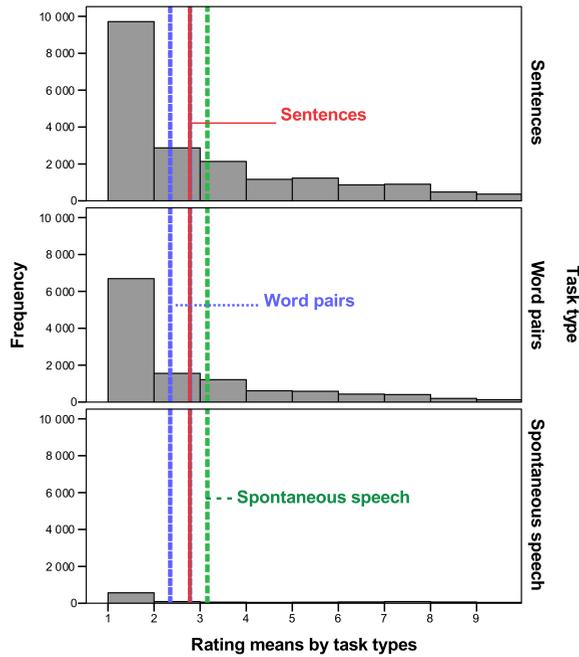


FIGURE 2. Rating distribution and mean foreign accent ratings according to three task types.

Finnish listeners were reliable and consistent in their rating of all three types of speech samples although spontaneous speech had the highest inter-rater reliability. It is interesting that the teaching listeners, who were all accustomed to hearing foreign accented Finnish either daily or often, gave similar ratings to those of the non-teaching listeners, most of whom reported hearing foreign accented Finnish either rarely or never. This finding is in line with some previous studies (Bongaerts et al. 1997; Kennedy & Trofimovich 2008; Toivola 2011, etc.). As the teachers of Finnish as an L2 transpired to be neither more compassionate nor harsher listeners than the non-teaching listeners, the results of this study contradict those of some earlier studies (e.g., for more lenient ratings, Barnwell 1989; Bongaerts et al. 2000; Kang & Rubin 2009; Hsieh 2011; for harsher ratings, Kang 2008). This finding suggests that the teaching and the non-teaching native listeners shared a similar perception regarding the degree of perceived accent despite their different experience of hearing foreign accented L2 speech and the non-teaching group having no experience of teaching Finnish as an L2.

The non-teaching listener group in the present study consisted of university students from different departments, whereas the teaching listener group

was comprised of teachers of Finnish as an L2. As no difference arose in the accent ratings between the non-teaching (student) listener group and the teaching listener group, it can be further concluded that studying per se had no effect on foreign accent ratings. Likewise, Toivola (2011) concluded that even though university students gave more ratings of strongly accented speech compared to non-studying listeners, most of whom were also L2 Finnish teachers, when rating read-aloud speech, this relationship was statistically insignificant. Furthermore, the results of the present study agree with those of Toivola (2011) in that no effect on foreign accent ratings was detected for listener's gender, age, Finnish as a L2 teaching experience, frequency of contact with immigrants, and being a native of the Helsinki Metropolitan Area as well.

As for the second research question, the results showed that perception of a foreign accent depended on the different types of speech samples presented for rating. Thus, it can be concluded that to analyze perceptions of foreign accent, the selection of speech materials is of the utmost importance due to a risk of instrumental bias. It is widely known that task differences affect findings still cited in the L2 phonology literature Moyer (2013). As Moyer (2013) rightly observes, unfortunately, the comparability of studies will continue to be an issue as long as researchers stick to using variable tasks for measuring the same construct, degree of perceived accent. This is a serious methodological problem and this task range problem documented in previous foreign accent studies was seen in this study as well: isolated, artificial and read word pairs were rated more nativelike than extemporaneous spontaneous speech samples which managed to reveal the L2 speakers' real oral performance resulting in more realistic (but harsher) foreign accent ratings. Spontaneous speech task ratings seem to reflect the L2 learners' more comprehensive assessment of oral performance in L2 when compared to word pair ratings. As Moyer (2013) points out, some late L2 learners are only expected to produce an authentic sound within an isolated task, yet they are far less convincing in real conversation where suprasegmentals and pragmatic skills come into play. Likewise, the results of this study revealed that all L2 learners' accent ratings were worse and less native on the spontaneous speech task when compared to the rather isolated task of read word pairs. This finding agrees with the results of several other studies (Moyer 1999; Ishida 2004), but conflicts with some research (as in Oyama 1976; Snow & Hoefnagel-Höhle 1982; Flege 1984; Piper & Cansin 1988; Munro & Derwing 1994; Thompson 1991; Munro & Mann 2005).

When the sample duration was examined, word pairs received significantly lower and better accent ratings than single sentences. Also, there was a trend (which did not reach significance) to rate single sentences more nativelike than spontaneous speech. All in all, there was a durational trend that the longer the sample type length was, the less native the mean accent rating

was. These durational results are consistent with the finding of the study by Ishida (2004) and Moyer (1999) in that the trend is toward a higher number of native ratings for the more isolated task items. Furthermore, this durational trend also agrees with Munro & Mann's (2005) durational effect finding that read paragraphs were the least native, followed by sentences, and then words. When the sample extemporaneity was examined, it was found out that there was a statistically systematic difference in ratings for spontaneous vs. read speech as well. That is, this study's findings showed that the less extemporaneous the speech sample was (extemporaneity scale: scripted word pairs = scripted single sentences < spontaneous speech), the more native the mean accent rating was. That is, the extemporaneous spontaneous speech sample was rated less native than scripted read word pairs. This meant that there was a significant difference between extemporaneous spontaneous speech and scripted read speech of word pairs. This effect of extemporaneity contradicts the results of Munro & Mann (2005) and Toivola (2011) on sample extemporaneity, as they reported that the more extemporaneous the sampling type is, the more native the mean accent rating (for instance, spontaneous picture narrations received more nativelike accent ratings than scripted paragraphs in the study by Munro & Mann 2005). To summarize, the findings of the present study support the general tendency in previous foreign accent studies: the longer, and less constrained the speech sample, the stronger the accent is rated (DeKeyser & Larson-Hall 2005).

5 Conclusion

When speaking informally with the listeners, they made it clear that even though they were instructed to ignore all other nonphonological speech content, it was very difficult for them to rate only purely phonological foreign accent while they rated spontaneous speech. The listeners told openly that they just could not simply ignore the mistakes in the choice of words, lexical errors, and grammatical errors and stated that these types of mistakes affected their accent ratings vastly. For instance, one listener provided the following comment: "If someone says 'valmistan lapset' (I make the kids ready), it is clear immediately that the speaker is nonnative". This informal observation agrees with the study by Moyer (1999), in which native listeners were influenced by structures beyond the L2 phonological production in their ratings of authenticity on a six-point scale of nativeness when the speech samples were more than mere word repetition. McDermott (1986) also noticed the same methodological effect pointing out that morphosyntactic or lexical errors can influence accentedness judgments when extemporaneous speech is used. This means that it is important that future researchers be aware of listeners' difficulties in

ignoring other linguistic factors when confronted with spontaneous speech, although it should be purely phonological variables determining accentedness ratings.

In line with Toivola's (2011) findings, teaching status did not affect the results in the present study either. However, a limitation of the present study (similar to Toivola 2011) was the fact that all 16 listeners not native to the Helsinki Metropolitan Area were teaching listeners. Out of 27 teaching listeners, 11 were natives of the Helsinki Metropolitan Area. In order to rule out possible interdependence between teaching and not being native to the Helsinki Metropolitan Area, future studies should recruit native listeners who are non-teaching and not native to the Helsinki Metropolitan Area. This would allow detection of whether not being native to or a resident of the Helsinki Metropolitan Area has an independent effect on foreign accent ratings. Moreover, although the results of both the present study and Toivola (2011) showed that frequency of contact with immigrants and being a teacher had no effect on accent ratings, it is possible that not being a native to the Helsinki Metropolitan Area has an effect on ratings when not combined with teacher status. As Toivola (2011) observes, future research should strive to discover whether native Finnish listeners' language variation affects their foreign accent rating behavior and their native speaker identification success, and thus listeners should be recruited from areas of Finland where native Finnish listeners are seldom in contact with foreigners and rarely hear foreign accented Finnish. For instance, much more heterogeneous reactions to foreign accents could be expected from Finns living in the north of the country, where native Finnish speakers are seldom in contact with nonnative speakers of Finnish. In this study even though most of the 34 nonteaching listeners reported rarely or never having contact with immigrants, the reality of being a resident of the Helsinki Metropolitan Area cannot be underestimated. The proof comes from statistics from public registers which show that at the beginning of 2016, 88,132 residents of Helsinki spoke a foreign language (other than Finnish, Swedish or Sami) as their mother tongue (Hiekkavuo et al. 2017). Almost half of Finnish residents with a foreign mother tongue live in the Helsinki Metropolitan Area, which is home to around 20 per cent of Finland's entire population. Consequently, the Helsinki Metropolitan Area is clearly the center of foreign language speakers in Finland (Hiekkavuo et al. 2017). Therefore, residents of the Helsinki Metropolitan Area are expected to be more used to hearing foreign-accented Finnish compared to native Finnish people from other regions of Finland. Also, as another future research suggestion, as one of the reviewers suggested, it would be interesting to compare the effect of listeners' highly educated status, whether educated people's and ordinary laymen's foreign accent ratings differ, on their perception of a foreign accent.

References

- Barnwell, D. 1989. 'Naive' native speakers and judgements of oral proficiency in Spanish. *Language Testing*, 6 (2), 152–163. DOI: 10.1177/026553228900600203.
- Boersma, P. & D. Weenink 2013. PRAAT. Amsterdam: University of Amsterdam. <http://www.fon.hum.uva.nl/praat/>.
- Bongaerts, T., S. Mennen & F. van der Slik 2000. Authenticity of pronunciation in naturalistic second language acquisition: the case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, 54 (2), 298–308. DOI: 10.1111/1467-9582.00069.
- Bongaerts, T., C. van Summeren, B. Planken & E. Schils 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19 (4), 447–465. DOI: 10.1017/S0272263197004026.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale (Mich.): Lawrence Erlbaum.
- DeKeyser, R. & J. Larson-Hall 2005. What does the critical period really mean? In J. F. Kroll & A. M. B. DeGroot (eds) *Handbook of bilingualism: psycholinguistic approaches*. New York (N. Y.): Oxford University Press, 88–108.
- Flege, J. E. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76 (3), 692–707. DOI: 10.1121/1.391256.
- Flege, J. E., M. J. Munro & I. R. A. MacKay 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16, 1–26. DOI: 10.1016/0167-6393(94)00044-B.
- Hiekkavuo, A., E. Haapamäki, S. Ranto & M. Salorinne 2017. *Population with foreign background in Helsinki 2016*. Helsinki: Helsingin kaupungin tietokeskus.
- Hsieh, C. 2011. Rater effects in ITA testing: ESL teachers' versus American undergraduates' judgments of accentedness, comprehensibility, and oral proficiency. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 9, 47–74.
- Ishida, S. 2004. *Taking pronunciation to task: exploring methodological issues in maturational constraints research*. A paper presented at the Annual Meeting of the American Association of Applied Linguistics, May, 1–4, Portland, Oregon, USA.
- Kang, O. 2008. Ratings of L2 oral performance in English: relative impact of rater characteristics and acoustic measures of accentedness. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 6, 181–205.
- 2012. Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9 (3), 249–269. DOI: 10.1080/15434303.2011.642631.
- Kang, O. & D. L. Rubin 2009. Reverse linguistic stereotyping: measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28 (4), 441–456. DOI: 10.1177/0261927X09341950.
- Karlsson, F. 2008. *Finnish: an essential grammar*. New York (N. Y.): Routledge.
- Kennedy, S. & P. Trofimovich 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience and semantic context. *The Canadian Modern Language Review*, 64 (3), 459–489. DOI: 10.3138/cmlr.64.3.459.
- Larson-Hall, J. 2006. What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɹ/ and /l/ by

- Japanese speakers. *Language and Speech*, 49 (4), 521–548. DOI: 10.1177/00238309060490040401.
- Levi, S. V. & S. J. Winters 2005. *External factors affecting the perception of foreign accent in L2 speech*. A poster presented at the 1st ASA Workshop on Second Language Speech Learning, May 14–15, 2005, Vancouver, Canada.
- Long, M. H. 1993. Second language acquisition as a function of age: research findings and methodological issues. In K. Hyltenstam & A. Viberg (eds) *Progression & regression in language: sociocultural, neuropsychological & linguistic perspectives*. Cambridge: Cambridge University Press, 196–221.
- McDermott, W. L. C. 1986. *The scalability of degrees of foreign accent*. PhD thesis. Ithaca (N. Y.): Cornell University.
- Moyer, A. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, 21 (1), 81–108. DOI: 10.1017/S0272263199001035.
- 2007. Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Multicultural Development*, 28 (6), 502–518. DOI: 10.2167/jmmd514.0.
- 2013. *Foreign accent*. Cambridge: Cambridge University Press.
- Munro, M. J. & T. M. Derwing 1994. Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11 (3), 253–266. DOI: 10.1177/026553229401100302.
- Munro, M. J., T. M. Derwing & C. S. Burgess 2010. Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52 (7–8), 626–637. DOI: 10.1016/j.specom.2010.02.013.
- Munro, M. J. & V. Mann 2005. Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics*, 26 (3), 311–341. DOI: 10.1017/S0142716405050198.
- Nunnally, J. & I. Bernstein 1994. *Psychometric theory*. 3rd ed. New York (N. Y.): McGraw-Hill.
- Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5 (3), 261–283.
- Piper, T. & D. Cansin 1988. Factors influencing the foreign accent. *Canadian Modern Language Review*, 44 (2), 334–342.
- Piske, T., I. R. A. MacKay & J. E. Flege 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29 (2), 191–215. DOI: 10.1006/jpho.2001.0134.
- Sawilowsky, S. S. 2003. A different future for social and behavioral science research. *Journal of Modern Applied Statistical Methods*, 2 (1), 128–132. DOI: 10.22237/jmasm/1051747860.
- Schmid, M. S. & H. Hopp 2014. Comparing foreign accent in L1 attrition and L2 acquisition: range and rater effects. *Language Testing*, 31 (3), 367–388. DOI: 10.1177/0265532214526175.
- Scovel, T. 1969. Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, 19 (3–4), 245–253. DOI: 10.1111/j.1467-1770.1969.tb00466.x.
- 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 AILA*. Québec: Le Presses de L'Université Laval, 389–401.

- Snow, C. & M. Hoefnagel-Höhle 1982. The critical period for language acquisition: evidence from second language learning. In S. Krashen, R. Scarcella & M. Long (eds) *Child-adult differences in second language acquisition*. Rowley (Mass.): Newbury House, 93–111.
- Tench, P. 1996. Intonation and the differentiation of syntactic patterns in English and German. *International Journal of Applied Linguistics*, 6 (2), 223–256. DOI: 10.1111/j.1473-4192.1996.tb00096.x.
- Thompson, I. 1991. Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning*, 41 (2), 177–204. DOI: 10.1111/j.1467-1770.1991.tb00683.x.
- Toivola, M. 2011. *Vieraan aksentin arviointi ja mittaaminen suomessa*. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-952-10-7217-8>.
- Uzal, M., E. Komulainen, M. A. Kılıç & O. Aaltonen 2017. Detection of non-native speech in a familiar language and in an unfamiliar language. *SKY Journal of Linguistics*, 20, 109–136. http://www.linguistics.fi/julkaisut/SKY2017/SKYJol30_2017.pdf#page=111.
- Uzal, M., T. Peltonen, M. Huotilainen & O. Aaltonen 2015. Degree of perceived accent in Finnish as a second language for Turkish children born in Finland. *Language Learning*, 65 (3), 477–503. DOI: 10.1111/lang.12126.

Appendix 1: Preliminary information form

1) Age	15-19 30-34 45-49 60-64	20-24 35-39 50-54 65-69	25-29 40-44 55-59
2) Gender	Female Male		
3) Mother tongue	Finnish Swedish Other		
4) Is your hearing normal?	Yes No		
5) Are you a native speaker of Finnish from the Helsinki Metropolitan Area?	Yes No		
6) Are you a Finnish-as-a-second-language teacher?	Yes No		
7) Are you a student?	Yes No		
8) Are you a linguistics/phonetics student?	Yes No		
9) Are you a linguistics/phonetics teacher/researcher?	Yes No		
10) Are you in daily contact with immigrants speaking Finnish as a second language?	never rarely often every day		
11) Dialect background: In my opinion, I speak the	Standard Finnish.		dialect.

The above information will be connected to the foreign accent judgments you give. Individual listener identity information will not be saved. I have read and accept the above information on the connection of the information.

Date _____

Signature and name in block capitals _____

Instructions for the Foreign Accent Listening Test

Read the instructions before starting the listening test!

Turn off your cell phone, please! Thank you!

Your task is to assess the kind of Finnish you hear pronounced by the speakers. Pay attention solely to the speaker's pronunciation. Avoid drawing attention to possible errors in grammar, syntactic errors, word-choice errors, and style errors. If you are unsure, make your best guess and use the whole scale. No speech impediments have been noted in the speech data of any speaker.

Foreign accent is a concept which has no generally accepted, uniform definition. A foreign accent means deviations in the standard pronunciation of second language as compared to native speaker pronunciation (Scovel 1969). Before the actual listening test, you will hear a total of 30 sentences from different speakers. The purpose is to briefly familiarize you with listening to speech samples, different pronunciation and giving ratings.

The actual listening test consists of three blocks:

- a word pairs listening test,
- a sentences listening test,
- a spontaneous speech listening test.

The blocks will come in random order. You can take a short break between blocks as well as in the middle of the word pairs and sentences listening tests. The spontaneous speech listening test is shorter than the others, and it includes no break.