Detection of non-native speech in a familiar language and in an unfamiliar language

Melike Uzal
University of Helsinki

Erkki Komulainen
University of Helsinki

Mehmet Akif Kılıç
Istanbul Medeniyet University

Olli Aaltonen
University of Helsinki

Abstract

This study explores the question of whether native and non-native listeners, i.e. natives familiar with the language they are judging and non-natives who are not, manage to distinguish a foreign accent from a native accent in the speech of native speakers (NSs) and nonnative speakers (NNSs). Participants included 21 speakers (11 NSs and 10 NNSs who were native Turkish speakers) as well as two listener groups that consisted of 61 Finnish listeners (FLs), and 10 Turkish listeners (TLs) without Finnish experience. This study compares accent ratings by these two listener groups that evaluated the 21 spontaneous speech samples for foreign accent using a 9-point scale. The results showed a very significant difference between the listener groups for the NSs but no significant difference for the NNSs. The difference between the FL and the TL groups was because the FLs managed to distinguish the NSs from the NNSs, but otherwise these two listener groups exercised statistically similar ratings. Therefore, these results demonstrate that the listeners' familiarity with Finnish, the target language, hence listeners' native speaker status strongly affect ratings of foreign accents, since native listeners could distinguish the NSs, whereas non-native listeners could not. The results suggest that listeners' familiarity with the target language plays a much more

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

SKY Journal of Linguistcs 20 (2017) 109–136

profound role in accent detection than their familiarity with the accent language. Moreover, the results show that contrary to previous research, in the absence of listeners' familiarity with the target language, it is much more challenging to detect a foreign accent. The results also showed that speech rate correlated with the judgments provided by the TLs but not with the judgments provided by the FLs. This result raises the possibility that there are salient universal features of non-native speech such as speech rate that even non-native listeners unfamiliar with the language they are judging utilize while judging a foreign accent.

Keywords: L2 speech, accent detection, L2 listener, L1 listener, familiarity with the accent language, familiarity with the target language, listeners' native speaker status

1 Introduction and background

The overall rating for degree of nativeness is often termed *global foreign accent* (Major 2007). Extensive research has demonstrated that global foreign accent correlates with a number of linguistic phenomena. Some of these include segmentals (Major 1987; 2001; Flege et al. 1995; González-Bueno 1997; Riney & Takagi 1999; Munro et al. 1999; Riney et al. 2000; Bunta & Major 2004), syllable structure (Magen 1998), and prosody (e.g. Anderson-Hsieh et al. 1992; Jilka 2000). Foreign accent is found to correlate also with subsegmental information including deviances observed in voice onset time difference in stop consonants, formant frequencies and vowel durations in vowels, and suprasegmental information including deviations in prosodic phenomena such as stress, phrasing, rhythm and intonation, as well as temporal aspects of speech such as segmental length, tempo, loudness, juncture and pitch differences (Moyer 2013; Schmid & Hopp 2014). For instance, Toivola (2011: 3) found that both temporal aspects of speech, such as speech rate, the number of pauses, the duration of pauses, and the number of single deviant phonetic segments contributed to the perceived degree of non-native accent in the speech of 10 Russian L2 learners of Finnish. Likewise, Trofimovich & Baker (2006: 2) found that two suprasegmentals (the duration of pauses and speech rate) were the variables that contributed the most to the perception of foreign accent by native listeners. In addition, non-native speakers often speak more slowly than native speakers, and previous studies (e.g. Munro & Derwing 1998; 2001) have shown temporal aspect of speaking rate to be the variable that contributes to the perceived degree of accentedness; listeners rate more slowly produced speech to indicate more accent than faster speech. Moreover, as Schmid & Hopp (2014: 4) state, native listeners detect a

foreign accent from features such as disfluency markers of filled pauses and repetitions (Lennon 1990) and hesitation (Dewaele 1996). Therefore, as Flege (1981: 445) observes, it seems that native listeners base foreign accent judgment on a combination of segmental, subsegmental, and suprasegmental differences that distinguish the speech of native speakers from that of non-native speakers. Furthermore, Munro & Derwing's studies (Munro & Derwing 1995; Derwing & Munro 1997) on intelligibility have shown that poor accent ratings are associated with phonetic, phonemic and grammatical errors, as well as problems with intonation.

Scovel (1995: 170) described some very broad general paralinguistic and phonetic features (e.g. the speed or the fundamental frequency of the voice listeners hear, some voice quality settings such as low voice, phonetic data such as the speaker's degree of retroflection, tone and pitch of voice) that affect the native listeners' decisions about whether or not the voice they heard was accented. His study raised the possibility that such paralinguistic features are *universal* and that these salient paralinguistic features are very telling of accentedness. All of the previous studies mentioned above are related to the current study because it also provides information on whether foreign accent judgment made by both native listeners and non-native listeners is based on speech rate, as has been found in previous research.

In second language acquisition research, it is commonly accepted that the most reliable distinguisher of accentedness is a native speaker of the language in question (Major 2007: 540), meaning that a listener's native speaker status has a profound effect on second language (L2) perception. This is because researchers have assumed that non-native speakers cannot be reliable listeners, since most fall short of native production and competence in the L2 themselves (Major 2007: 540). Munro et al. (2006: 114), however, have challenged the view that the ratings of native listeners are more valid than those of non-native listeners. When it comes to empirical evidence, only a small number of studies (e.g. Flege 1988; Riney et al. 2005; MacKay et al. 2006; Munro et al. 2006; Kang 2008) have compared the ratings of native listeners and non-native listeners to determine the effect of listeners' native speaker status on ratings of accentedness (Munro et al. 2006: 126). Surprisingly, with the exception of Kang's study (2008), all found that the ratings of native listeners and non-native listeners familiar with the target language were quite similar. In contrast, Kang (2008: 184) found that non-native speaker (NNS) listeners were harsher than the native speaker (NS) listeners in accentedness ratings.

There are a few studies (Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) that have suggested that even non-native listeners lacking familiarity with the language spoken (L2), i.e. the target language, were able to distinguish a foreign accent accurately from a native one, albeit with a lower success rate than that of native listeners. According to the above-mentioned studies, non-native listeners may have performed this successful accent detection by utilizing general markers of non-nativeness such as speech rate and sentence/utterance duration as an estimate of fluency – a clue from the speech (Bond et al. 2008: 7). For instance, in Major's (2007: 539) study, American English listeners unfamiliar with Portuguese could detect an English accent in Portuguese. Major (2007: 552) interpreted this finding to mean that native and non-native listeners have similar abilities in rating foreign accents and that their L1s and L2s do not dramatically affect the ratings. Voice quality, for example, has been suggested as a potential marker of non-nativeness, though its role in L2 production has not been thoroughly investigated yet (see e.g. Esling 2000). Articulatory effort and carefulness are other potential markers (Weber & Pöllmann 2010: 541). All of the studies mentioned above are related to the current study because it also shows whether non-native listeners lacking familiarity with the target language can distinguish a foreign accent accurately from a native one.

1.1 Aims of the present study

This study has been set with two aims. The first aim was to discover how some listener background factors such as the listeners' familiarity with the accent language and native speaker status, i.e. familiarity with the language spoken (L2)/ the target language spoken, affected listeners' perception of foreign accent. The native language of the non-native speaker is termed *accent language*. One of the aims of the study was to find out whether the listeners' familiarity with the accent language gave them any advantage of detecting accentedness. Due to this reason, none of the listeners were informed that the non-native speakers of Finnish were native speakers of Turkish. With this kind of experimental design, it is important that non-native listeners do not know beforehand that they will hear their own mother tongue as an accent language.

The second aim was to obtain a preliminary assessment on whether markers of non-nativeness are language-specific or language-independent. That is, this study aimed to explore whether markers of non-nativeness are language-independent, innate and universal, in which case some general

markers would reveal a speaker's native status regardless of the listener's language abilities, or whether they are language-specific markers, in which case only listeners who have formed a native perception of the language could identify a speaker's native speaker status. The varied findings of the studies addressing the effects of the listener's familiarity with the target language spoken on foreign accent ratings (e.g. Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) were one of the motivations of the present study. All of these studies' findings on listener familiarity with the target language spoken have suggested the existence of universal general markers for non-nativeness, which in turn implies that the Turkish listeners in the present study unfamiliar with Finnish would be expected to make use of these universal general markers of non-nativeness, enabling them to detect accentedness. This study explores whether this prediction, supported also by Roy C. Major from Arizona State University (personal communication, 2016), correlates with the findings. If the findings of the current study show that non-native listeners with no familiarity in the language they judge use speech rate as a clue to judge accentedness, it will also provide further empirical proof for the existence of universal general markers of non-nativeness.

1.2 Research question and hypotheses

The study sought to discover the degree to which some listener background factors (listeners' NS status, hence familiarity with the target language spoken, and familiarity with the accent language) affect the degree of perceived accent ratings in L2 Finnish. In this study, listeners possessing varying familiarity with the target language spoken were asked to assess the degree of perceived foreign accent in Finnish spoken by native speakers and non-native speakers. One item of interest was how accurately native Turkish listeners (non-native listeners) lacking familiarity with Finnish could detect Finnish L2 speakers in Finnish speech samples. Non-native listeners unfamiliar with the target language spoken are not familiar with the native accent and therefore might not be expected to form reliable, accurate and valid perceptions of that language. Thus, it was expected (the null hypothesis) that in the absence of any familiarity with the target language spoken, the non-native listeners would be unable to identify a foreign accent reliably and accurately, even though they had excellent familiarity with the accent language of Turkish. Indeed, as Major (2007) observes, in theory the idea of asking listeners to rate foreign accents in an

unfamiliar language is strange, even ludicrous. Unlike Major's study (2007), studies by Bond et al. (2008) and Weber & Pöllmann (2010) did not have listeners with excellent familiarity in the accent language and no familiarity with the target language to allow any direct comparisons with the TLs in the present study. In line with previous research on accent detection, it was expected (the alternate hypothesis) that the TLs' excellent familiarity with the accent language (linguistic experience) might help them differentiate between native and non-native speakers of Finnish. Therefore, if the TLs in this study are successful in detecting L2 speakers of Finnish, it will be possible to argue that the TLs' excellent familiarity with the accent language enabled them to distinguish between native and non-native speakers of Finnish in a reliable manner.

Compared with previous studies, this study is perhaps the first to include listeners who had excellent familiarity with the target language (they shared the same L1) and who had no familiarity with the accent language, namely the Finnish listeners (FLs). Also, this study is perhaps the first to employ non-native listeners to judge the degree of perceived accent from spontaneous speech samples in a language unfamiliar to them. To date, only Major (2007), Bond et al. (2008) and Weber & Pöllmann (2010) have used non-native listeners unfamiliar with the language they rated; however, as for their speech sample choices, Major (2007) used read passages of varying durations (M = 22 s) and both Weber & Pöllmann (2010) and Bond et al. (2008) used read speech of sentences. In contrast, this study used spontaneous speech samples of the same duration for both NSs and NNSs.

1.3 Significance of the study

First, the study will provide knowledge of how foreign accent (in this case Finnish with a Turkish accent) is perceived by both Finnish native speakers and non-native speakers unfamiliar with Finnish, which will contribute to the accent detection literature regarding perceptual studies of accented speech. Second, since non-native speakers unfamiliar with Finnish are native speakers of Turkish, it will be further explored whether the listeners' excellent familiarity with the accent language has an effect on their accentedness ratings. In addition, a debated issue in accent-rating studies is whether the listeners' foreign accent ratings indicate something about the speech itself because they are influenced by its acoustic and phonological properties, or whether they indicate something about the listener and

therefore vary with the listeners' language experience (Weber & Pöllmann 2010). Researchers should have an understanding of the factors that figure in listeners' accentedness judgment and, in particular, how properties of speech and the characteristics of listeners influence that accentedness judgment.

2 Methodology

2.1 Speakers

The 21 adult speakers were 11 native speakers (NSs) and 10 non-native speakers (NNSs) of Finnish. All the NSs were native speakers of Finnish from the Helsinki Metropolitan Area with no knowledge of Turkish, eight female speakers and three male speakers aged 22–39 (M = 27.6). All the NNSs were bilingual NSs of Turkish from a wide variety of Turkish cities; they consisted of five female speakers and five male speakers aged 27–66 (M = 40.2). All 10 adult NNSs were either first or second-generation Turkish immigrants to Finland. To sum up, all 21 speakers resided in the Helsinki Metropolitan Area and spoke standard Finnish (Karlsson 2008).

2.2 Speech samples

The speech samples consisted of pieces of spontaneous speech on a topic chosen by the individual speaker from three options (see Appendix). The speech samples from the 21 speakers were 40-second segments extracted from 1-minute recordings. The reason for using spontaneous speech samples was that since spontaneous speech is the most authentic form of natural speech, the use of spontaneous speech as a stimulus would be the most meaningful way to learn about accentedness. A 40-second piece of spontaneous speech is a stretch long enough for accent characteristics to emerge.

2.3 Listeners

Two main listener groups participated in the study: native listeners, referred to as the Finnish listeners, (n = 61) and non-native listeners, referred to as the Turkish listeners, (n = 10) unfamiliar with Finnish. Thus, this study had two listener groups with respect to both listener familiarity with the accent language and the language being spoken (L2) dimension: the Turkish

listeners who had excellent familiarity with the accent language and no familiarity with the language being spoken (L2) and the Finnish listeners who had no familiarity with the accent language and excellent familiarity with the language being spoken (L2), Finnish being their L1. However, since the number of non-native listeners was much smaller than the number of native listeners, this resulted in more foreign accent ratings given by native listeners than non-native listeners. That is, a total of 1321 foreign accent ratings were rated (61 native listeners rating 21 spontaneous speech productions and 10 non-native listeners rating 21 spontaneous speech productions).[1]

The non-native listeners were monolingual NSs of Turkish with no familiarity in Finnish. All the non-native listeners were from Giresun, in the Black Sea region of Turkey. None of the non-native listeners had studied Finnish or had been to Finland, and none reported familiarity with Finnish (in terms of hearing or recognizing it). Thus, these Turkish non-native listeners were good examples of laymen.

2.4 Procedure

The 61 native listeners completed the rating task individually in a soundproof recording studio in the Helsinki Metropolitan Area, Finland, with a total of 21 speech samples (21 speakers × 1 spontaneous speech task recording) presented via headsets. For the 10 non-native listeners, the foreign-accent rating task was completed in a quiet room in Giresun, Turkey, again with each non-native listener individually performing the rating, with a total of 21 speech samples presented via headsets. For the native listeners, a preliminary information form and a short training session were provided before the rating began. To avoid unrelated linguistic factors affecting accent ratings, in the preliminary information form, the native listeners were instructed to ignore all non-phonological speech content and only assess foreign accent. The native listeners were also instructed to use the entire scale while rating the samples and were told to guess if they were uncertain. A short training session consisted of 30 practice speech samples of single sentences (13 from NSs, 17 from NNSs) to help familiarize the

[1] There were eight ratings with negative reaction times and other 162 premature ratings given before the end of the sample. These 170 ratings (25% of all ratings) were excluded from the analyses. All in all, due to the excluded 170 ratings, there were a total of 1321 foreign accent ratings rated (61*21*10=1491) instead of 1491.

native listeners with the rating process and the range of accents; these were not analyzed.

A 9-point scale was used for the rating of accent. The listeners were told that they would hear productions spoken by either non-native or native speakers of Finnish, but they were not informed about the proportion of native and non-native speakers. They were asked to rate each production for the degree of accent by pushing one of the nine buttons representing a scale from 1 (no foreign accent) to 9 (very strong foreign accent). The same procedure was followed with the non-native listeners as it was with the native listeners, the only difference being that the non-native listeners did not participate in the training session. This was because the express aim of the study was to discover if non-native listeners who had never heard the target language could identify a foreign accent. Due to this design, the non-native Turkish listeners did not listen to any speech samples of standard Finnish, even though many of them so requested to have a benchmark for their accent judgements. Both the native and non-native listener rating sessions consisted of one block lasting 16–25 minutes. The runs including speech samples and the speakers were randomized.

2.5 Data analysis

71 listeners (61 FLs and 10 TLs) rated 21 speech samples (11 NSs and 10 NNSs). Each listener contributed to many data-points. Consequently, the data were not independent because they came from the same listener who rated. Statistical analysis of such correlated (nested, clustered) data requires methods that can properly account for the intra-subject correlation of response measurements. If such correlation is ignored, then inferences such as statistical tests or confidence intervals can be grossly invalid. On the other hand, the use of averaged ratings (aggregated data) leads to association indices that are too high (Iversen 1991).

Generalized linear mixed models (GLMM) were formulated by Nelder & Wedderburn (1972) as a way of unifying various other statistical models. These methods are now available in statistical packages such as statistical package for the social sciences (SPSS) (IBM 2016). They handle correlated data structures (O'Dwyer & Parker 2014). They also apply to cases where non-normality of distributions exists and/or the scale is ordinal. The linear mixed model (LMM) is more restricted. It can handle nested data, but it assumes the normal distribution of residual values (Madsen & Thyregod 2010).

The nature of the data obtained in the present study required the use of LMM (because the data was correlated and nested) and GLMM (because in addition to being nested data, non-normality of distributions existed as well). Consequently, analyses of foreign accent ratings were performed using GLMM (response variable ordinal, link logit) and LMM in SPSS 24. Results were practically the same. Thus, the results here were reported from the LMM analyses, since it is better suited to present results showing means and standard deviations. This choice also affected the way the results were depicted in graphical form, namely histograms and scatterplots in this study. Scatterplots use averaged values. Regression lines (X on Y) in them show how strong the depicted associations are.

There were four groups (combinations of FL/TL and NS/NNS) to compare. This was done by analyzing six possible pairwise comparisons using Bonferroni corrected post-hoc tests. The magnitude of the mean difference (effect size) was expressed by Cohen's d (difference divided by the pooled standard deviation) (Cohen 1988; Sawilowsky 2003).

The present situation in which there were more listeners than the spontaneous speech samples presented to be rated is not usual. Due to this, reliability had to be assessed by using several reliability indices. Therefore, there were three different approaches to the reliability of the ratings. First, the mean correlation with other listeners was calculated for each listener. The mean correlation grew if ratings bore similarity. Second, all foreign accent ratings (both NS and NNS) were split into two random groups. Reliability existed if the splits correlated. Third, intra-class correlation (ICC) was calculated (Shrout & Fleiss 1979); this is often used in situations where raters are used. The obvious problem is that it gives values that are too high when the number of raters exceeds the number of the stimuli. The ratio here was quite high, 71 to 21 (3.38). Consequently, in such situations very low actual reliability can produce quite a high reliability coefficient.

Speech rates of the speakers were measured with Praat software. The syllables and silences were segmented and labeled from the recorded speech samples. Syllable and silence durations and numbers were measured automatically with a script, which measures interval durations. Later, the speech rate was calculated manually by dividing the number of syllables by the total duration of a certain speech sample, in syllables/second. The duration of silence was not used here, but the silences had to be labeled as well to find the number of syllables clearly.

3 Results

3.1 Comparison of combined speaker and listener groups

Figure 1 illustrates the relative frequency of accent ratings of the four listener-speaker combinations. Most of the FL group successfully rated the NSs as natives (84%). The mean was 1.25. In contrast, the TL group, who were unfamiliar with Finnish, rated the NSs as having a moderate degree of foreign accent according to the foreign accent rating scale (20% gave 1, M = 4.79). The number of 1s for the NNSs was far less in both listener groups: FLs= 10% (M = 5.64), TLs: = 14% (M = 5.55).

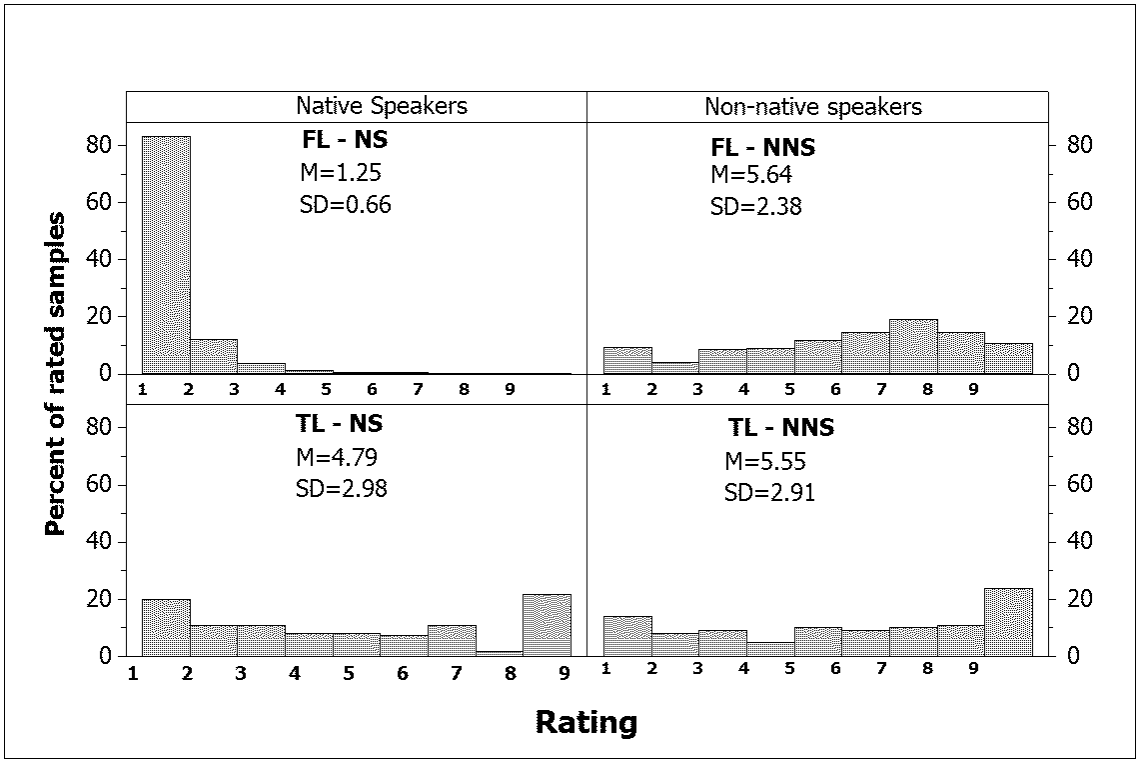

Figure 1. Rating distribution, means and standard deviations according to listener groups

The LMM analysis was carried out to evaluate the statistical significance and the magnitude of the mean differences described above. Three groups (FL-NNS, TL-NS and TL-NNS) were quite similar to each other in their accent rating means, but differed very clearly from FL-NS group. The pairwise post-hoc comparisons (LMM) of these four groups are shown in Table 1. Three pairs were not significantly different from each other and all their pairwise calculated effect sizes were small. These three groups, however, all differed very significantly statistically (p <.001) from the FL-

NS mean. These three effect sizes (*d*) ranged from 1.24 to 1.53 and could be characterized as very large (Cohen 1988; Sawilowsky 2003). The comparisons confirmed that the listeners' familiarity with Finnish, the target language, was a very strong factor in the ratings since native listeners could distinguish NSs, whereas non-native listeners could not. Figure 1 and Table 1 show that the TLs did not identify and distinguish the NSs (M = 4.79) from the NNSs (M = 5.55). There was a tendency in the right direction, though. This difference in ratings, however, was statistically not significant (p = .440), and the effect size was very small (d = .27) (Sawilowsky 2003). The LMM analysis clearly supports the fact that the non-native listeners were not able to distinguish the accent difference between the NSs and NNSs.

Table 1. Post hoc pairwise LMM analyses of four groups and their effect sizes

Post hoc pairs		**Difference**	***p****	**Cohen's *d***
FL – NS	FL - NNS	-4.39	<.001	1.53
FL – NS	TL – NS	-3.54	<.001	1.24
FL – NS	TL – NSS	-4.30	<.001	1.50
FL – NNS	NS – TL	-0.85	=.980	0.30
FL – NNS	TL – NNS	0.09	=.999	0.03
TL – NS	TL - NNS	-0.76	=.440	0.27

* = Bonferroni corrected

3.2 Reliability analysis

Table 2 shows reliability results of listener ratings carried out in three different ways. The harshest index was the mean correlation between the listeners. As Table 2 shows, TLs' reliability was low for NNSs (.31), for NSs (.14) and combined (.23). FLs' values were high except FL-NS value (.19). These NS samples were very homogenous, which explains this low value. The split procedure was perhaps the most realistic index of reliability. There was a low degree of reliability for TLs (.58 for NNS, .51 for NSs and .67 for both NSs and NNSs combined). The split total in the FL-group was .93, which is very high. All ICC values expressed high degrees of inter-rater reliability, even for TLs (ICC = .68) for NNSs, which is the lowest ICC of TLs, which is still a good inter-rater agreement. All the ICC values, however, were unrealistically high because of the nature of the data at hand. This bias typically presents itself when the number of raters is greater than the number of items to be rated. The values in mean correlations and ICC were not comparable either, since there were 61 FLs

as raters and only 10 TLs. The split method seems to be fair and dependable, i.e. FL ratings had a high reliability, whereas TL ratings had only some reliability. TL value did not reach the level (.7) which is a typical lower limit in split-type situations (Nunnally & Bernstein 1994).

Table 2. Three reliability indices

Mean correlation

	FL	TL	Total
NNS	.84	.31	.62
NS	.19	.14	.11
Total	.92	.23	.72

Split

	FL	TL	Total
NNS	.55	.58	.48
NS	.74	.51	.83
Total	.93	.67	.89

ICC

	FL	TL	Total
NNS	1.0	.68	.99
NS	1.0	.89	.99
Total	1.0	.74	.99

3.3 Speech rate and accent ratings

Due to the stimuli being of the same duration for all speakers, 40-second spontaneous speech samples, it was not possible to measure varying utterance durations and their correlation with accent ratings as in Major (2007), Bond et al. (2008) and Weber & Pöllmann (2010). Instead, speech rate was measured in Praat to see whether mean accent ratings and speech rate correlated, which would imply the use of speech rate as a universal clue to accentedness, hence non-nativeness. Multiple factors can influence foreign accent ratings, one of which might be a slower speaking rate: it might be perceived as less fluent and, therefore, more accented. To find out whether speech rate and accent ratings had any relationship with each other according to listener groups, both Pearson r analysis and LMM analyses were conducted. These two analyses were done to provide comparability among the results of previous studies because previous accent detection studies used Pearson r values to examine the effect of speech rate on accent ratings. The values of these two analyses are shown in Figure 2. In Figure 2, x-axis shows speech rates of speakers calculated in syllables/second ranging from 1.84 to 4.87, whereas y-axis shows foreign accent ratings

ranging from 1 to 9. Figure 2 shows that for the NSs one of the Pearson r correlations was significant (r = -.63, a moderately strong correlation), meaning that the slower the NSs' speech rate, the higher the accent score (the worse and harsher accent rating) was given by TLs. This same relationship in LMM analysis shows a lower value (-.22)**.** Likewise, for the NNSs, the Pearson r correlation was very significant (r = -.89), meaning that the slower the NNSs' speech rate, the higher the accent score was given again by TLs. The same relationship in LMM was likewise of lower value (-.47). Regression lines (X on Y) indicated the same thing. Lines from FL ratings were horizontal while corresponding lines from TL ratings showed a clear negative slope. Thus, LMM results, Pearson r correlations and regression lines show that TLs made use of speech rate, i.e. fluency, as a cue to native speaker status for both the NSs and the NNSs.

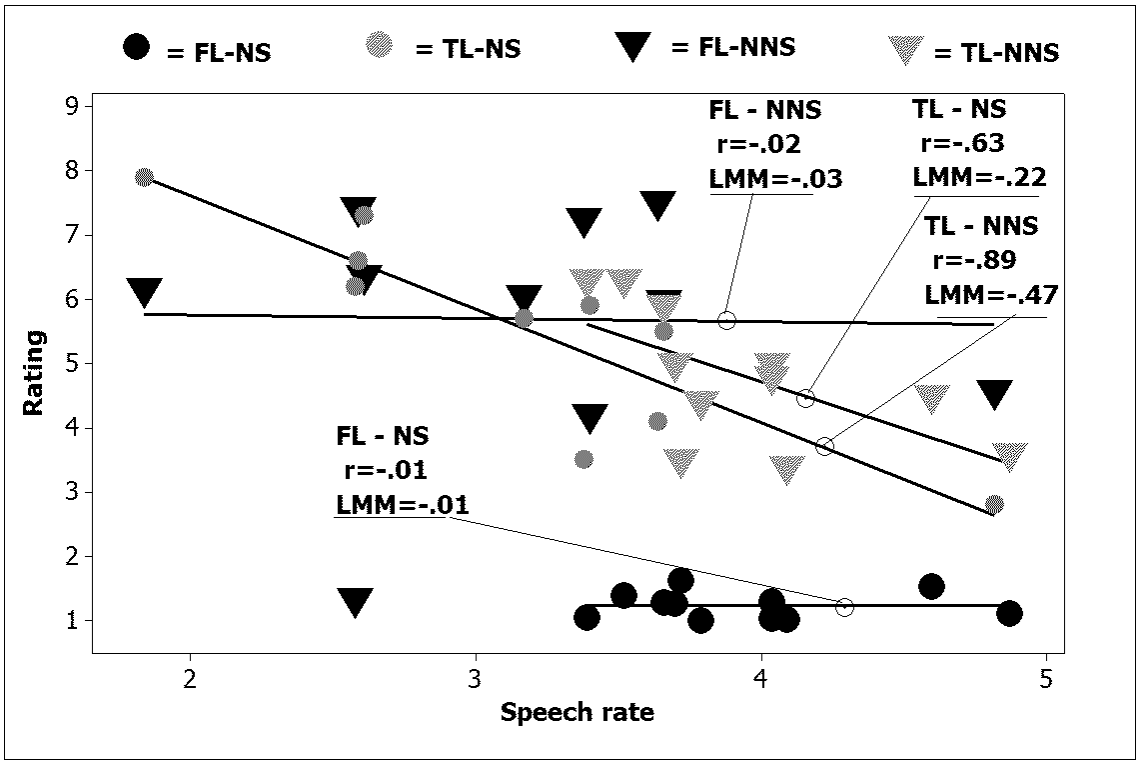

Figure 2. A scatterplot of speech rate versus foreign accent rating

As shown in Figure 2, TLs relied slightly more on speech rate when they rated NNSs when compared to NSs (LMM analysis values of -.22 for NSs vs. -.47 for NNSs). Indeed, high correlations between speech rates of both NSs and NNSs and mean accent ratings for TLs showed that they used speech rate as a basis for foreign accent judgement. Only TLs were found to make use of speech rate. They might have relied on these universal

perceptual features thinking that they do not have any linguistic information to rely on, such as their L1 or L2, as high correlations between speech rate and accent ratings made clear. It is important to note that LMM-analysis coefficients were considered more correct than Pearson r correlations because when counting correlations with mean values as was done here, a part of variance is lost when using aggregated values.

Figure 3 depicts the agreement in ratings between FLs and TLs using regression lines and correlations coefficients.

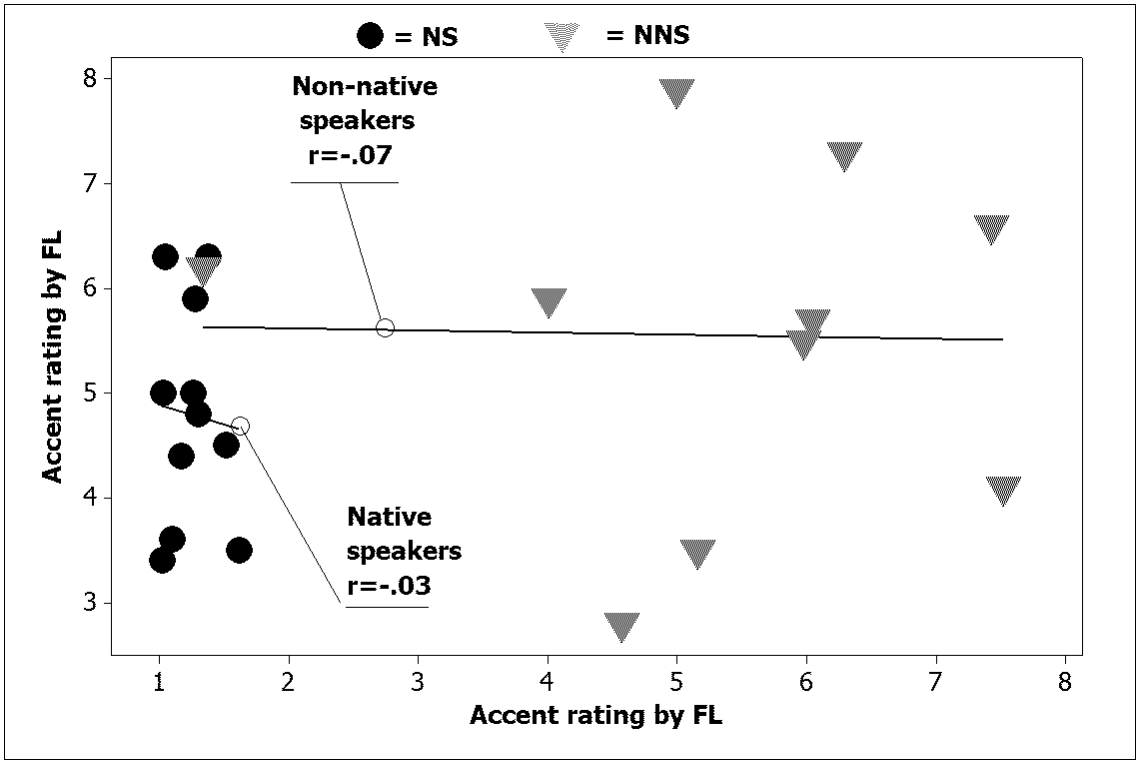

Figure 3. A scatterplot of foreign accent ratings FLs versus TLs by NSs and NNSs

In Figure 3, it can be seen immediately that in both speaker groups the agreement was close to zero. Regression lines (FL on TL) go almost horizontally or have a slight negative slope, and the corresponding correlation coefficients are very small and insignificant, i.e. when it comes to agreement of ratings between TLs and FLs the finding was that there was no agreement.

4 Discussion

4.1 The effect of listeners' familiarity with the language spoken

It was hypothesized that, in the absence of any familiarity with the target language spoken, listeners would be unable to detect accentedness accurately. The present findings confirm this prediction, showing that listeners' familiarity with the target language spoken is a prerequisite for detecting accentedness. The only listener background factor that affected the perception of foreign accent was listeners' NS status, hence listeners' familiarity with Finnish, the target language. The TLs, who had no familiarity with Finnish, were unable to distinguish between native speakers of Finnish and Turkish non-native speakers of Finnish, whereas the FLs managed to identify them extremely accurately. This finding agrees with Kang (2008: 196) that even non-native listeners familiar with the target language, in Kang's case L2 English, had different perceptions (were more stringent) than native listeners in accentedness ratings. Therefore, this study concludes that in the absence of any familiarity with the target language (L2), accent detection might be inaccurate even if the listeners' L1 is the same as that of the non-native L2 speakers; i.e. even the listeners' excellent familiarity with the accent language, Turkish, did not help them to detect accentedness correctly.

The findings of the present study on listener familiarity with the language spoken (L2) – that TLs unfamiliar with Finnish were unable to distinguish between the native and non-native speakers of Finnish – is still in line with the findings of Bond et al. (2008) and supplement the accent detection literature as well by showing that non-native listeners unfamiliar with the target language spoken indeed use speech rate as a clue while making accentedness judgments. However, in contrast to Bond et al.'s (2008) non-native listeners unfamiliar with the target language spoken whose accent ratings were above the level of chance, this study's TLs' accent ratings were not accurate. In what follows, two reasons are discussed that could account for some of the differences in rating patterns between Bond et al. (2008) and the present study. The first reason is that, in the present study, a much more representative sample of non-native listeners was used than in Bond et al. (2008). The 10 TLs in this study represented laymen, and the 61 FLs (some of them Finnish-as-a-second-language teachers and some students at the University of Helsinki) represented educated people with a general familiarity in other languages. The FLs were very reliable, accurate and successful at rating the degree of

Turkish foreign accent in the Finnish speech samples, and the TLs, who had no idea how Turkish-accented Finnish would sound, or for that matter any other accented language in Finnish, were not so reliable and accurate. The TLs had studied German and/or English (world languages) as second or foreign languages at school, but those two languages were of no use to them when judging accentedness in an unknown language, such as Finnish. Similar to the present study's FLs, in Bond et al. (2008), the American listeners were college students who represented educated people with familiarity and experience in dealing with other languages. That is, there was a difference in rating patterns between the present study and Bond et al. (2008) study because TLs represented laymen who were not experienced in dealing with other languages and thus were a more representative sample when compared to American listeners in Bond et al. (2008) who were all college students and relatively experienced in dealing with other languages.

Unmarked accents and *marked accents* may help in the identification of a foreign accent, regardless of the language. If a listener perceives a marked accent, it is language-specific, whereas if a listener perceives an unmarked accent, it is *language-independent.* Accent refers here to a universal aspect of accent in any given language. Marked and unmarked accents apply regardless of a listener's L1 background, familiarity with the language spoken (the target language, L2) and familiarity with the native accent of that language (L1), i.e. the accent language. First, it is suggested that when an accent is unmarked, it is the type of accent that prevents a listener unfamiliar with both the language being spoken (L2) and the accent language (L1) from identifying the speaker's L1 in the L2 speech. Therefore, it is not so telling and salient to cause listeners to detect the accent in question. Examples could be Finnish with a Turkish accent, as in this study, Turkish with a Finnish accent or even Polish with a Turkish accent. In fact, the findings of this study have clearly shown that even Turkish listeners were unable to detect a Turkish accent in Finnish. The Turkish listeners in this study were ordinary laymen, hence non-native listeners with no linguistic experience and linguistic sophistication, and not everyone knows how a Turkish accent (the Turkish accent supposedly being an unmarked accent universally proposed in this study) sounds in different languages or specifically in Finnish L2, as was the case in this study.

Second, it is suggested that Russian, Chinese, Japanese, French, German, Taiwanese, English, Dutch, Italian, Spanish and some others

might be examples of marked accents. In general, these languages have such strong, distinct, revealing and solid accents that any educated listener familiar with these world languages would recognize the accents of their speakers in any given language. Some accents are so telling that it is easier to identify them. According to this model, it would be easier to detect the L1 (accent language) of German-accented Turkish speakers than that of Turkish-accented Finnish speakers because of the distinct quality of the former accent. Likewise, Turkish, Finnish, Hungarian, Polish, Kyrgyz, and Swedish could be examples of unmarked accents. There is little knowledge at a global level about how these languages sound in terms of accent. Moreover, since both Turkish and Finnish are far from being world languages, listeners with no linguistic experience and linguistic sophistication might fail to detect them as accent languages simply because they have no idea what Turkish or Finnish sound like. At present, this is a matter of speculation, as our study does not attempt to uncover these possible accents but simply addresses this possibility. It was beyond the scope of this study to identify these accents, and this speculation comes as a byproduct of the findings in this study. Adjunct Professor Zinny Bond (personal communication, 2016) from Ohio State University, however, concurred with these suggestions and observed that this analysis makes good sense of both the results of the present study and those of Bond et al. (2008).

The second reason that could account for some of the differences in rating patterns between Bond et al. (2008) and the present study is that in taking the proposition of marked and unmarked accent into consideration, it is suggested that in the present study Turkish might have constituted an unmarked accent in L2 Finnish for the Turkish listeners. Likewise, the possibility of marked and unmarked accent might explain how monolingual American listeners managed to identify a native vs. non-native background with a success rate of 63%, significantly above the level of chance in Bond et al. (2008) study. These American listeners with no familiarity in either the accent language, Russian, and the target language, Latvian, might thus have had some idea of how a Russian accent (supposedly being a universally marked accent) would sound in any given language. Since Russian and Latvian belong to the same language family, intuition suggests that for American listeners it might have been relatively more difficult to distinguish between native speakers of Russian and native speakers of Latvian. Nevertheless, the American listeners in Bond et al. (2008) managed to give the highest evaluations to native Latvians and

distinguished between low and high-proficiency Russians, just as the Latvian listeners did. If monolingual Americans, however, had heard Finnish with a Turkish accent, Turkish with a Finnish accent or Swedish with a Polish accent, their success in distinguishing native speakers from non-native speakers might have been lower, perhaps only at the level of chance. A direction for further studies in accent detection research could be to find out the possible existence of marked and unmarked accents.

Scovel (1995: 175) described that accent detection success depends on native listener sophistication in languages, i.e. their overall linguistic experience and linguistic sophistication. Likewise, Scovel's argumentation (1995) that accent detection success depends on native listener sophistication in languages can be broadened to include non-native listeners as well. Therefore, it can be argued that accent detection success depends on non-native listener sophistication in languages as well exemplified with the findings of this study on non-native listeners judging accentedness in a language unfamiliar to them. In other words, in accent detection studies involving the rating of accents in unfamiliar languages, the addition of one new term is proposed. That term is *listener familiarity with language accents.* This term is the same as Scovel's (1995) final third stage of identification in accent recognition process in which, depending on the overall linguistic experience and linguistic sophistication of the native listeners, they may identify the native language of the accented voices. The only difference is that this term applies to all types of listeners regardless of their native speaker status whereas the term used by Scovel (1995) only applies to native listeners.

4.2 The effect of listener familiarity with the accent language

The findings in this study showed that even excellent familiarity with the accent language did not afford listeners any advantage in reliably detecting accentedness in the absence of familiarity with the target language spoken. The TLs, who had no familiarity with the target language, Finnish, but who had excellent familiarity with the accent language, Turkish (they shared the same L1, i.e. the advantage of sharing an L1 language background with the L2 speaker) did not identify foreignness successfully. Given that this study produced a negative finding with respect to listener familiarity with the accent language, Turkish, its argumentation of stating that listener familiarity with the accent language does not necessarily cause listeners to rate accentedness reliably comes from a position of weakness because it

produced a negative finding. An absence of evidence is not evidence of absence. The failure to produce similar positive findings using different language combinations does not refute previous research and Major's (2007) findings on listener familiarity with the accent language. The first author of this study discussed the findings of the present study with Roy C. Major (personal communication, 2016), and he found them unexpected but observed that this discussion makes good sense of both the findings in the present study and those of his and previous research. Although the findings in this study are unexpected, they are nevertheless compatible with findings from previous studies (Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) on listener familiarity with the accent language and provide fresh insights into the issue. This is because the study findings fill the gap in the accent detection literature by showing that even non-native listeners with the potential advantage of excellent familiarity with the accent language can fail to identify non-native speakers, in the absence of familiarity with the target language. By showing this, the present study also supplements the findings of previous research and adds new pieces to the puzzle of how accent detection occurs.

While it might seem counterintuitive that listener familiarity with the accent language failed to enhance their ability to detect foreignness, it should be noted that previous studies also varied on whether they found a correlation between familiarity with the accent language and accent ratings (e.g. Major 2007; Bond et al. 2008 for studies finding such an effect, Munro et al. 2010 for no effect, and Weber & Pöllmann 2010 for a study that failed to find such a strong effect). It is suggested that there are several reasons for the findings of the present study that the listeners' excellent familiarity with the accent language conferred no advantage in the absence of any familiarity with the target language. First, it is proposed that the phonological properties of particular language pairs (in this case Turkish and Finnish) might make accent detection more difficult. This first suggestion creates the space for the second suggestion: if a language has a marked accent, as in the study case of Russian by Bond et al. (2008), it might be easier to detect it. Likewise, the possibility of Turkish being an unmarked accent, at least on a global level, might have made its detection more difficult for the TLs, who did not realize that it was in fact their own mother tongue which was foreign in the L2 speech of Finnish. Since the TLs had never heard Finnish before it sounded completely foreign, incomprehensible and unmarked.

4.3 Possible cues for accentedness: listeners' rating behavior

The findings of this study showed that there was a statistically significant difference between the ratings of the TLs and those of the FLs. Not only did the TLs fail to distinguish between the NSs and NNSs of Finnish, but their agreement with FL group ratings was also bad or nonexistent, as TLs gave foreign accent ratings relying on speech rate. The findings also showed that foreign accent ratings of FL group were among themselves reliable and the FLs were able to identify a foreign accent, whereas ratings of the TL group were only to some extent reliable. That is, when making their accent judgment on the stream of spontaneous speech there was a big difference in rating behavior. Naturally, the TLs had no possibility of relying on segmental information because they had never heard the target language before. Before the rating began, all the TLs had asked to hear a model voice representing standard Finnish pronunciation, as they wished to have a yardstick on which to base their accent judgment when rating the spontaneous speech samples. Furthermore, they reported difficulty in deciding on the accentedness scores, feeling that they were rating at random with no clue of what Finnish sounded like, which made them feel uneasy and uncomfortable. They had no expectations of how Finnish would sound, so they had no chance of knowing whether the speech signals corresponded to the pronunciation norms of Finnish. It was clear that the TLs were puzzled by the rating task.

Taken together, the findings of this study show how the TLs and FLs perceived the degree of foreign accent in Finnish in fundamentally dissimilar ways, each based on different phonetic parameters. There was a clear disparity and no agreement in the rating strategy between the FLs and TLs, who had never heard Finnish before and thus had no knowledge of its phonological structure. Although they had one source of language information (Turkish) available to them, they did not seem to utilize this familiarity with the L1 of the NNSs. Based on previous research on non-native listeners judging accentedness from languages unfamiliar to them, in the present study apart from the same L1 advantage, one could expect non-native listeners to make use of obvious perceptual cues such as slower tempo, i.e. speech rate (universal, non-linguistic speech characteristics as Bond et al. 2008 term them). The results of the study showed that the non-native listeners had indeed utilized general traces of non-native speech such as speech rate, but their use of speech rate did not improve their ability to identify natives from non-natives. That is, even though general traces of

non-nativeness might exist, and even though TLs used speech rate as a clue to non-nativeness, it did not lead to accurate accent ratings, an important distinction to make. Therefore, one cannot argue that the use of speech rate improved a listener's ability to detect accentedness.

Bond et al. (2008) found non-native listeners with no familiarity in the language they judge use utterance duration to make foreign accent judgment. Thus, Bond et al. (2008) concluded that their non-native listeners lacking familiarity with the accent and the target languages used fluency as a general marker of non-native speech, which they roughly estimated by utterance duration. The present study has found non-native listeners with no familiarity in the language they judged (the target language) use speech rate to judge a foreign accent. All in all, it seems that what is probably observable from speech without knowing the language is speech rate, tempo, utterance duration and fluency as these two studies have shown. Additionally, the reason for these features to signal non-native speech is because previous studies have found non-native listeners with no familiarity in the target language they judged to utilize these features while making their foreign accent judgments. In fact, what the speech rate findings of this study showed was plainly that the non-native listeners made use of speech rate to judge a foreign accent, so slower speech rate signaled more non-native speech to them, whereas a faster speech rate signaled more native speech to them. This finding shows that general traces of non-nativeness (speech rate) might exist. While it might seem counterintuitive that TL's use of speech rate as a cue to non-nativeness failed to enhance their ability to detect foreignness, it should be noted that previous studies also varied on whether they found a correlation between utterance duration and accent ratings. For instance, previous studies that measured utterance duration as an example of general markers for non-native speech noticed that the correlation between utterance duration and accent ratings were either insignificant (Weber & Pöllmann 2010: 540) or weak (Major 2007: 549). Only Bond et al. (2008: 6) found a high correlation between utterance duration and accuracy in identifying speakers as native or non-native for their American listeners. Thus, it seems that only Bond et al. (2008) were justified in concluding that their non-native listeners lacking familiarity with the accent and the target languages used fluency as a general marker of non-native speech.

5 Conclusion

In light of the findings in the present study, the following conclusions can be made about foreign accent detection: If listeners have excellent familiarity with the accent language (it is their L1) and no familiarity with the target language – as the TLs had in this study – they have both their L1 language source (if they can detect this accent language and utilize their familiarity with it) and some possible universal non-segmental information available to them in their use. That is, when non-native listeners consider they have no linguistic information on which to base their accent judgement, it is natural for them to rely on universal non-segmental information, such as speech rate or fluency, as indeed TLs did in the present study. There is a critical point here, however. Because this study demonstrated that the TLs' use of these general markers of non-nativeness, i.e. speech rate, did not improve their ability to detect foreignness, such argumentation is weak. However, the failure of the present study to reproduce the findings of Bond et al. (2008) adds to the accent detection literature by showing that the possible use of these universal non-segmental perceptual cues does not necessarily allow listeners to distinguish between native and non-native speakers accurately. The findings of the present study are in line with those of Bond et al. (2008) in the sense that the TLs based their judgment on some universal perceptual factors, i.e. speech rate. The difference in the findings of this study is that their foreign accent ratings based on speech rate perceptions were inaccurate, as they were unable to distinguish natives from non-natives. The findings of this study have clearly shown that it was challenging for native Turkish-speaking listeners to detect accentedness correctly (the accent language Turkish) when they were unfamiliar with Finnish as the target language.

In contrast, if listeners have excellent familiarity with the target language (it is their L1) and no familiarity with the accent language, as the FLs had in this study, they have their L1 language source, hence both segmental and non-segmental information from their native language, available to them. Thus, the FL had multiple sources of information on which to base their foreign accent evaluation. The acoustic analysis of speech rate by Praat and correlation analyses, however, showed that FLs did not make use of speech rate. All in all, the findings of this study that the TLs relied on non-segmental information such as speech rate agrees with the findings of Riney et al. (2005), Major (2007), Weber & Pöllmann (2010), and Bond et al. (2008). Riney et al.'s (2005: 441) acoustic and

auditory analyses showed that their untrained native listeners relied more on segmentals (vowels and consonants, especially /r/ and /l/), whereas untrained, non-native listeners of American English (native speakers of Japanese) relied more on non-segmental parameters (meaning everything else, including intonation, fluency, sentence duration, and speech rate) to make perceptual judgments. One must keep in mind, though, that the non-native listeners in the study by Riney et al. (2005) were familiar with English, the target language, whereas TLs in this study were unfamiliar with the target language.

The findings given in this study suggest that the commonly accepted view in literature observing that a listener's native speaker status has a strong positive effect on their foreign accent detection success is correct. However, this same view has been challenged by previous research such as Major (2007), Bond et al. (2008), and Weber & Pöllmann (2010). Of course, this raises the further question of why these empirical findings seem to point in the other direction as the everyday observation, and this might be a topic for further research. All in all, the findings of this study suggest that markers of non-nativeness are language-specific because in this study only the Finnish listeners who have formed a native perception of the language could identify a speaker's native speaker status, whereas the Turkish listeners failed to identify non-nativeness. However, the findings also showed that the Turkish listeners made use of some language-independent general markers of non-nativeness such as speech rate to identify non-nativeness.

Abbreviations

L1	first language
L2	second language
GLMM	generalized linear mixed model
ICC	intra-class correlation
LMM	linear mixed model
M	mean
NS	native speaker
NNS	non-native speaker
FL	Finnish listener
TL	Turkish listener
SPSS	statistical package for the social sciences

Appendix

The spontaneous speech instructions

Discuss one of these subjects (or make a subject up yourself). Your reply should be only 1-minute long (reply in Finnish).
A. Describe your weekend or your daily routine: What do you usually do, when, with whom, for how long, what is interesting about it, etc.?
B. Describe one significant experience in your life: Who was included? How old were you then? How did this affect you?
C. Describe a person in your life who means a lot to you: How do you know this person? Why is she/he important in your life?

References

Anderson-Hsieh, Janet & Johnson, Ruth & Koehler, Kenneth. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language learning* 42(4). 529–555.

Bond, Zinny. September 5th 2016. (Personal communication.)

Bond, Zinny S. & Stockmal, Verna. & Markus, Dace. 2008. A note on native and non-native accentedness judgments. *Ohio university working papers in linguistics and language teaching*. 1–8.

Bunta, Ferenc & Major, Roy. 2004. An optimality theoretic account of Hungarian ESL learners' acquisition of/e/and/æ. *IRAL: International Review of Applied Linguistics in Language Teaching* 42. 277–299.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* (2nd ed.) Hillsdale (NJ): Lawrence Erlbaum.

Derwing, Tracey M., & Munro, Murray J. 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19(01), 1–16.

Dewaele, Jean-Marc. 1996. Les phénomènes d'hésitation dans l'interlangue française: analyse de la variation interstylistique et interindividuelle [The phenomena of hesitation in the French interlanguage: analysis of the interstylistic and interindividual variation]. *Rassegna Italiana da Linguistica Applicata* 28. 87–103.

Esling, John H. 2000. Crosslinguistic aspects of voice quality. In Raymond D. Kent & Martin J. Ball (eds.). *Voice quality measurement*, 25–35. San Diego: Singular Publishing Group.

Flege, James E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America* 84(1). 70–79.

—— 1981. The phonological basis of foreign accent: A hypothesis. *Tesol Quarterly* 15(4). 443–455.

Flege, James E. & Munro, Murray J., & MacKay, Ian R. A. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16. 1–26.

González-Bueno, Manuela. 1997. Voice-onset-time in the perception of foreign accent by native listeners of Spanish. *IRAL-International Review of Applied Linguistics in Language Teaching* 35(4). 251–268.

IBM Corp. Released 2016. *IBM SPSS statistics for Windows, version 24.0*. Armonk, NY: IBM Corp.

Iversen, Gudmund R. 1991. *Contextual analysis.* Sage university paper series on quantitative applications in the social sciences, No. 7. Beverly Hills, CA: Sage Publications.

Jilka, Matthias. 2000. Testing the contribution of prosody to the perception of foreign accent. *New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second-language speech,* 199–207. Klagenfurt, Austria: University of Klagenfurt.

Kang, Okim. 2008. Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness. *Spaan fellow working papers in second or foreign language assessment* 6. 181–205.

Karlsson, Fred. 2008. *Finnish: An essential grammar.* London: Routledge.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40. 387–417.

MacKay, Ian R. A. & James E. Flege & Satomi Imai. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27(02). 157–183.

Madsen, Henrik & Thyregod, Poul. 2010. *Introduction to general and generalized linear models*. Boca Raton: Chapman & Hall/CRC Press.

Magen, Harriet S. 1998. The perception of foreign-accented speech. *Journal of Phonetics* 26(4). 381–400.

Major, Roy C. 1987. English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15(2). 197–202.

—— 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. New York: Routledge.

—— 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29(04). 539–556.

—— August 25th 2016. (Personal communication.)

Moyer, Alene. 2013. *Foreign accent: The phenomenon of non-native speech.* Cambridge: Cambridge University Press.

Munro, Murray J. & Derwing, Tracey M. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech* 38(3). 289–306.

—— 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48(2). 159–182.

—— 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language acquisition* 23(04). 451–468.

Munro, Murray J. & Derwing, Tracey M. & Flege, James E. 1999. Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics* 27(4). 385–403.

Munro, Murray J. & Derwing, Tracey M. & Burgess, Clifford S. 2010. Detection of nonnative speaker status from content-masked speech. *Speech Communication* 52(7–8). 626–637.

Munro, Murray J. & Derwing, Tracey M. & Morton, Susan L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28(01). 111–131.

Nelder, John A. & Wedderburn, Robert W. M. 1972. Generalized linear models. *Journal of the Royal Statistical Society* 135(3). Series A (General). 370–384.

Nunnally, Jum & Bernstein, Ira. 1994. *Psychometric theory* (3rd ed.). New York: New York: McGraw-Hill.

O'Dwyer, L. M. & Parker, C. E. 2014. *A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study*. Washington, DC: Regional Educational Laboratory Northeast & Islands.

Riney, Timothy J.& Takada, Mari & Ota, Mitsuhiko. 2000. Segmentals and global foreign accent: The Japanese flap in EFL. *Tesol Quarterly* 34(4). 711–737.

Riney, Timothy J. & Takagi, Naoyuki. 1999. Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49(2). 275–302.

Riney, Timothy J. & Takagi, Naoyuki & Inutsuka, Kumiko. 2005. Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39(3). 441–466.

Sawilowsky, Shlomo S. 2003. A different future for social and behavioral science research. *Journal of Modern Applied Statistical Methods* 2(1). 128–132.

Schmid, Monika S. & Hopp, Holger. 2014. Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing* 31(3). 367–388.

Scovel, Thomas. 1995. Differentiation, recognition, and identification in the discrimination of foreign accents. In J Archibald. (ed.), *Phonological acquisition and phonological theory,* 169–195.

Shrout, Patrick E. & Fleiss, Joseph L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86(2). 420–428.

Toivola, Minnaleena. 2011. *Vieraan aksentin arviointi ja mittaaminen suomessa* [The assessment and measurement of foreign accent in Finnish]. Helsinki: Helsingin yliopisto.

Trofimovich, Pavel & Baker, Wendy. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1). 1–30.

Weber, Andrea & Pöllmann, Katja. 2010. Identifying foreign speakers with an unfamiliar accent or in an unfamiliar language. *New sounds 2010: Proceedings of the sixth international symposium on the acquisition of second language speech*, 536–541. Poznan: Adam Mickiewicz University.

Contact information:

Melike Uzal
Faculty of Arts
Department of Modern Languages
Speech Sciences
P.O. Box 9, FI-00014
University of Helsinki, Finland.
e-mail: melike(dot)uzal(at)helsinki(dot)fi

Erkki Komulainen
University of Helsinki
e-mail: erkki(dot)komulainen(at)helsinki(dot)fi

Mehmet Akif Kılıç
Istanbul Medeniyet University
e-mail: makilic(at)yahoo(dot)com

Olli Aaltonen
University of Helsinki
e-mail: olli(dot)aaltonen(at)helsinki(dot)fi