Foreign Accentedness Revisited: Canadian and Singaporean Raters' Perception of Japanese Accented English

Kazuya Saito, Birkbeck College, University of London

Natsuko Shintani, University of Auckland

Abstract
The current study examined how two groups of native speakers—monolingual Canadians and multilingual Singaporeans—differentially perceive foreign accentedness in spontaneous second language (L2) speech. The Singaporean raters, who had exposure to various models of English and also spoke multiple L2s on a daily basis, demonstrated more lenient accent scores than the Canadian raters, who used only North American English as their primary language of communication. Both of them used a similar processing strategy to make such accentedness judgements—drawing exclusively on the phonological aspects of L2 speech without taking into account the appropriate/complex use of L2 lexicogrammar.

Key words: Foreign accent, speech perception, second language pronunciation

Title:

Foreign Accentedness Revisited: Canadian and Singaporean Raters' Perception of Japanese

Accented English

Abstract

The current study examined how two groups of native speakers—monolingual Canadians and multilingual Singaporeans—differentially perceive foreign accentedness in spontaneous second language (L2) speech. The Singaporean raters, who had exposure to various models of English and also spoke multiple L2s on a daily basis, demonstrated more lenient accent scores than the Canadian raters, who used North American English as their primary language of communication. Both of them used a similar processing strategy to make such accentedness judgements— drawing exclusively on the phonological aspects of L2 speech without taking into account the appropriate/complex use of L2 lexicogrammar.

Key words: Foreign accent, speech perception, second language pronunciation

Acknowledgement

Correspondence concerning this article should be addressed to Kazuya Saito, Birkbeck, University of London, Department of Applied Linguistics and Communication, 26 Russell Square, London, WC1B 5DT, UK. Email: k.saito@bbk.ac.uk

Monolingual and Multilingual Raters' Perception of Foreign Accentedness in Second Language
Speech

Examining how native speakers with diverse linguistic backgrounds (e.g., familiarity,
language learning experience) perceive second language (L2) speech could be an important
research initiative for several pedagogical reasons. For instance, our better understanding of the
link between rater backgrounds and behaviours will reveal which linguistic features are
particularly important for non-native speakers to be successfully understood by a wide range of
native speaking individuals. Furthermore, such findings will also provide several implications for
rater training in high-stakes proficiency exams, where all raters should have an explicit
understanding of the source of rater bias to minimize individual differences among potentially
heterogeneous native speaking raters. In previous L2 speech research, foreign accentedness has
been extensively used as a measure for assessing the global quality of language (Piske, MacKay,
& Flege, 2001). The current study was designed to examine how two groups of raters—
monolinguals and multilinguals—differentially process and evaluate L2 accentedness.

Background

Given that English is used as a language of international communication in many
business and academic domains all over the world, attaining adequate L2 oral ability is critical
for helping non-native speakers achieve their future goals. Several previous survey studies have
noted the tendency of many L2 learners to *idealize* accent-free speech (i.e., nativelike accent) as
the end goal of speech learning (e.g., Tokumoto & Shibata, 2011). Yet, attaining such nativelike
L2 proficiency is extremely difficult, even for early bilinguals (Abrahamsson & Hyltenstam,
2009). Accordingly, certain scholars have emphasized the importance of setting *realistic* goals

especially for adult L2 learners (e.g., prioritizing understanding over nativelikeness) (Derwing & Munro, 2009).

To examine the global constructs of L2 speech production, Derwing and Munro conducted a series of studies (e.g., Munro & Derwing, 1995) where native speakers evaluated a range of foreign accented speech on a 9-point scale for comprehensibility (*1 = easy to understand, 9 = difficult to understand*) and accentedness (*1 = little accent, 9 = heavily accented*). The results showed that even heavily accented speech was highly comprehensible, suggesting that comprehensibility and accentedness may tap into two independent domains of spontaneous L2 speech.

More recently, Isaacs and Trofimovich (2012) and Trofimovich and Isaacs (2012) examined the relative impact of linguistic errors on native speakers' judgement of comprehensibility and accentedness in the context of 40 Francophone talkers. Whereas accentedness scores were strongly tied to pronunciation accuracy (particularly segmentals), comprehensibility scores were equally related to pronunciation and such aspects of language as fluency, vocabulary and grammar (see also Saito, Trofimovich, & Isaacs, 2015). Similar results were replicated in follow-up studies with a group of 45 intermediate-level ESL students (Crowther, Trofimovich, Saito, & Isaacs, 2015) and 120 beginner- to advanced-level Japanese learners of English (Saito, Trofimovich, & Isaacs, 2016a).

These studies have shed light on how native speakers perceive comprehensibility and accentedness using different kinds of processing strategies and linguistic resources. For L2 comprehensibility, native speakers tend to collect as much phonological, temporal, lexical and grammatical information as possible from accented speech in order to understand their interlocutors' overall message in a *timely* and *efficient* manner (Kang, Rubin, & Pickering, 2010;

Saito, 2015). For L2 accentedness, native speakers exclusively attend to salient problematic

linguistic features entailing much learning difficulty (e.g., segmentals) without processing any

content of the utterances, judging "nativelikeness" in an invariably *fast, effortless*, and *intuitive*

manner (Munro, Derwing, & Burgess, 2010; Saito, 2015).

It is noteworthy, however, that the aforementioned findings were based solely on native

North American English speakers' perceptions of comprehensibility and accentedness. In most

studies, these native speaking raters' judgement scores were averaged across speech samples

with an assumption that they would *similarly* assess the quality of L2 speech based on their

intuition about one English model—North American English. Though limited in number,

however, some previous studies have identified the significant influence of rater background,

especially in terms of their familiarity with particular foreign accented speech. For example,

certain native speakers tend to provide more lenient general proficiency scores, especially when

they have learned the L2 learners' first language (Winke, Gass, & Myfordet, 2013); and/or have

taught non-native speakers in classroom settings (Saito, Trofimovich, Isaacs, & Webb, 2016b).

Furthermore, it has been shown that ESL professionals are likely to better understand the content

of L2 speech (Kennedy & Trofimovich, 2008) and articulate their assessment decisions by using

linguistic terminologies (Isaacs & Thomson, 2013). These studies in turn suggested that native

speakers with various backgrounds use different kind of linguistic information to perceive L2

comprehensibility and accentedness.

To further examine how rater backgrounds impact the process and product of L2 speech

judgement, we conducted an experimental study where two different groups of raters with

different backgrounds—native speakers of Canadian English and Singapore English—evaluated

the L2 comprehensibility and accentedness of Japanese learners of English. By analyzing their

L2 assessment patterns, our objective was to examine another potentially crucial listener factor: raters' daily linguistic experience with several L2s and various kinds of English models. On the one hand, both Canadian and Singaporean raters were comparable in terms of the lack of any ESL/EFL teaching experience and contact with Japanese learners of English. On the other hand, these raters were significantly different in terms of how many languages they use for daily communication purposes. In this project, the Canadian raters were termed as "monolinguals" such that they spoke only North American English as their language of communication from birth onward in the English-dominant area of the country (Vancouver).[1] The Singaporean raters were referred to as "multilinguals" in the sense that they spoke a few L2s in a multilingual environment, where four languages (English, Mandarin, Tamil and Malay) are used as the national languages. As their first language (L1), they used Singaporean English while speaking a few other languages (e.g., Mandarin, Malay, Tamil) on a daily basis, and had exposure to various models of English (General American, Received Pronunciation, Singapore English) while listening to TV and using the internet (for details, see Method).

In our precursor study (Saito & Shintani, 2016), we took an exploratory approach to analyzing how Canadian and Singaporean raters differentially judge L2 *comprehensibility*, defined as how much effort it takes to understand what someone is saying (Derwing & Munro, 2009). In the study, the raters listened to 50 Japanese learners' spontaneous speech samples, and assessed their comprehensibility on a 9-point scale (*1 = easy to understand, 9 = difficult to*

---

[1] In the current project, certain individuals who perceived themselves as monolinguals were carefully chosen as the Canadian raters. Whereas these raters must have had ample opportunities to be exposed to many other accents over the course of their lives in a multicultural city (i.e., Vancouver), they reported that they used only English as their main language of communication (with little use of the other official language—French). In this regard, the Canadian raters could be considered to be "less multilingual" than the Singaporean raters who regularly listened to diversely accented English as well as used multiple languages in various social settings.

understand). The results provided two crucial findings. First, a paired-sample *t*-test demonstrated that Singaporean raters assigned significantly lower and more lenient comprehensibility scores ($M = 4.0$, $SD = 1.5$) to L2 speech samples than the Canadian raters did ($M = 4.7$, $SD = 1.2$). Second, a multiple regression analysis showed that whereas the raters generally based their comprehensibility judgements on various linguistic domains of L2 speech, spanning pronunciation, fluency, vocabulary and grammar, the Singaporean raters drew more on the lexicogrammar element than the Canadian raters did. To provide a rough estimate of the ratio of variance (in their comprehensibility judgement) explained by linguistic resource, the results are summarized in Table 1.

TABLE 1

The results suggest what factors influenced the perception of comprehensibility by two different groups of the raters. The Canadian raters were almost entirely influenced by pronunciation while the Singaporean raters were partly influenced by lexicogrammar. In the current study, we will report on how the Canadian and Singaporean raters evaluated the other construct of L2 speech—foreign accentedness. Following the definition of Trofimovich and Isaacs (2012), L2 accentedness refers to how closely the linguistic profiles of an utterance approach those of a native speaker. By comparing accentedness scores with those of L2 comprehensibility judgements in the precursor study, we will aim to explore how rater backgrounds (monolinguals and multilinguals) can impact the linguistic correlates of native speakers' intuitions of L2 speech.

Method

Speech Materials

Talkers. The L2 speech data originates from our unpublished corpus of audio recordings by 200+ Japanese learners of English in Montreal and Vancouver completing various speaking tasks, such as word reading, sentence reading, and picture description tasks (Saito, 2011). To ensure a wide range of oral proficiency, 50 Japanese learners were carefully selected for the current investigation based on three affecting variables in adult L2 speech learning: (a) all of the participants reported highly frequent use of L2 (they considered English as their main language of communication at home and/or work at the time of data collection); (b) their length of residence in Canada varied widely from 1 month to 11 years; and (c) their age of arrival in Canada was substantially different, from 19 to 40 years. For a detailed discussion of the relationship between the quality and quantity of L2 input, age of acquisition and adult L2 oral proficiency development, see Piske et al. (2001) and Saito (2015).

TABLE 2

Material preparation. The Japanese talkers' spontaneous speech was elicited via a timed picture description task, whereby the talkers were asked to describe seven pictures with five seconds of planning time per photo. To elicit a certain length of spontaneous speech data without excessive hesitations and dysfluencies, each picture contained three key words in English as hints (these key words were written below the photo). The first four pictures were used as practice, and the remaining three pictures were included for the final analyses.

The three target pictures referred to (a) a table left out in a driveway in heavy rain (keywords: rain, table, driveway); (b) three men playing rock music with one singing a song and the other two playing guitars (keywords: three guys, guitar, rock music); and (c) a long stretch of road under a cloudy blue sky (keywords: blue sky, road, cloud). The keywords were intentionally chosen to push Japanese learners to use problematic segmental and syllable structure features

and show their pronunciation abilities (e.g., English /r/-/l/ contrast). The same materials were

used in the precursor research (Saito & Shintani, 2016).

Approximately 10 seconds (range: 7.2 to 12.7 sec) from the beginning of each

participant's description were extracted and stored as a single WAV file. As such, each

participant contributed 30 sec of spontaneous speech, comparable to previous L2 speech research

(e.g., Murno & Derwing, 1995 for 15 sec, Trofimovich & Isaacs, 2012 for 30 sec). In total, 150

speech samples were created from 50 Japanese learners (50 talkers × 3 pictures).

L2 Accentedness Judgements

Canadian raters. A total of 10 native speakers of North American English were

recruited at a university in Vancouver, Canada. They were undergraduate students majoring in

business and psychology. According to the language background questionnaire, they reported

little familiarity with Japanese-accented English (M = 1.3) on a 6-point scale (*1 = Not at all, 6 =*

Very much). While three of the raters reported beginner or intermediate knowledge of French,

they did not use it for daily communication purposes. The other raters considered themselves to

be monolinguals as they spoke only English on a daily basis. They had never taught English in a

professional setting prior to the project.

Singaporean raters. We recruited 10 native speakers of Singapore English who were

undergraduate students majoring in non-linguistic subjects (e.g., business, humanities). Their

familiarity with Japanese-accented English was quite limited on a 6-point scale (M = 1).

Different from the Canadian raters who used only North American English for everyday

purposes, the Singaporean raters reported their use of multiple languages other than English

(e.g., Mandarin, Malay, Tamil). As summarized in Table 3, while they used English as the most

dominant language in home, school and social settings (68.0-82.5%), they also spoke the other

national languages for daily communication purposes (11-21%).

TABLE 3

In terms of their L1 English, they also reported high levels of use and exposure to a range

of English models (Singapore English[2] for conversations with friends and family, General

American and Received Pronunciation for watching TV and using the Internet), as summarized

in Table 4. Importantly, the Singaporean raters regularly used English not only with other native

speakers but also with non-native speakers from various other Asian countries.

TABLE 4

Procedure. The raters first familiarized themselves with the picture prompts, and then

received a brief explanation from a trained research assistant in regards to the foreign

accentedness rubric, which they then used to evaluate the linguistic nativelikeness of the L2

speech samples. We explicitly asked the Singaporean raters, in particular, to draw on their own

English as a point of reference instead of referring to any particular model (e.g., General

American, Received Pronunciation, Singapore English). During the listening session, 150 speech

samples were played in a randomized order via the *Praat* software (Boersma & Weenik, 2012).

Upon listening to each sample only once, they assigned comprehensibility (1 = *easy to*

understand, 9 = *difficult to understand*) and accentedness (1 = *little accent*, 9 = *heavily accented*)

scores on a 9-point numerical scale (Derwing & Munro, 2009). In this paper, we report the

results of their accentedness scores. The entire session took approximately 1.5 hours.

---

[2] Singapore English is characterized as a complex phenomenon which includes not only linguistic traits of Received Pronunciation and General American, but also idiosyncratic features in phonology (e.g., reduced vowel inventory) as well as in lexicogrammar (e.g., admixture of lexical items and grammar structures from contact languages). For a comprehensive review of Singapore English, see Deterding, 2007).

Pronunciation, Fluency, Vocabulary and Grammar Analyses

To examine which linguistic factors in the L2 speech tokens predicted the raters' accentedness judgements, the tokens were submitted to linguistic analyses. Following the standard in L2 speech research measuring pronunciation (Piske et al., 2001), fluency (e.g., Bosker, Pinget, Quené, Sanders, & De Jong, 2013), and lexicogrammar (Crossley, Sasbury, & McNamara, 2014), we adopted the expert judgement approach towards analyzing the linguistic profiles of spontaneous speech samples. For the validation of the method, see our earlier study (Saito et al., 2015).

Experienced coders. We recruited five *experienced* native speaking coders (3 males, 2 females) who were graduate students at an English speaking university in Montreal, Canada. All of them were native speakers of Canadian English. Not only had they received professional training on pronunciation, fluency, vocabulary, and grammar analyses, but they also had an extensive amount of English-as-a-second-language teaching experience ($M = 4.0$ years from 2 to 6 years).

Procedure. For both the audio-based (pronunciation/fluency) and transcript-based (vocabulary/grammar) measures, the coders first received information on the linguistic categories that they were asked to rate for (see Supporting Information for training scripts). Next, they practiced the procedure with five speech samples before proceeding to the main dataset.

To provide enough phonological information for the pronunciation and fluency analyses, the three picture descriptions were combined and stored as a single wav file per talker. The raters listened to the audio files (as many times as they wanted until they felt satisfied about their judgements) in a unique randomized order via the MATLAB software. By using a moving slider on a computer screen, they evaluated them for four linguistic categories: (a) segmentals

(substitution, omission, or insertion of individual consonant and vowel sounds); (b) word stress (misplaced or missing primary stress); (c) intonation (appropriate, varied use of pitch moves); and (d) speech rate (speed of utterance delivery)[3]. Their judgement scores were recorded on a 1000-point scale (*0 = nontargetlike, 1000 = targetlike*).

With respect to the vocabulary and grammar analyses, the coders read transcribed speech samples (instead of listening to audio files) to control for the influence of pronunciation/fluency on their judgements. While reading each transcript consisting of three picture descriptions, the raters used a moving slider on a 1000-point scale to judge (a) lexical appropriateness (accuracy of vocabulary); (b) lexical richness (varied and sophisticated use of vocabulary); (c) grammatical accuracy (errors in word order, grammar endings, agreement); and (d) grammatical complexity (use of sophisticated, non-basic grammar). For all of the linguistic analyse above, they were asked to rely on their own variety of English (North American English) as an index for measuring the targetlikeness of the speech samples.

Results

Accentedness Ratings

The results of the Cronbach's alpha analysis demonstrated relatively high inter-rater agreement for 9-point accentedness ratings among the 10 Canadian raters ($\alpha = .92$) as well as t10 Singapore raters ($\alpha = .91$). By pooling over the monolingual and multilingual raters' judgement scores respectively, two mean rating scores were generated for each of the 50 Japanese talkers, and summarized in Table 5.

---

[3] In this study, we used human rater judgements to rate the fluency aspect of speech samples (*0 = too fast or too slow, 1000 = optimal*). In line with the standard of L2 speech research, the subjective measures of this kind have been found as a more ecologically-valid method to reflect what is called "optimal" speech rate which could be considered neither too slow nor fast (e.g., Derwing, Rossiter, Munro, & Thomson, 2004). In contrast, it has remained unclear to what extent objective measures (e.g., articulation and speech rate) can fully capture the non-linear nature of L2 fluency proficiency (Munro & Derwing, 2001).

TABLE 5

To examine whether the monolingual and multilingual raters differentially perceived foreign accentedness, their scores were submitted to an independent *t* test. The results showed that the monolingual raters assigned significantly higher and stricter accent ratings than the multilingual raters did, $t(98) = -3.069$, $p = .003$.

Pronunciation, Fluency, Vocabulary and Grammar Ratings

Given a relatively high inter-rater reliability among the five coders (Cronbach's alpha = .80-.90), their linguistic scores were also averaged across each of the participants' speech samples, summarized in Table 6. The results of the descriptive statics noted a wide range of proficiency among the 50 Japanese talkers' pronunciation, fluency, vocabulary and grammar performance (*SD* ranging from 119 to 300).

TABLE 6

 Linguistic Correlates of Accentedness

In this subsection, we aim to probe the phonological, temporal, lexical and grammatical influences on the monolingual and multilingual raters' accentedness judgments. To this end, a set of partial correlation analyses were conducted to examine how these linguistic factors were related to the monolingual Canadian raters, and the multilingual Singaporeans, respectively. According to the results (summarized in Table 7), the monolingual Canadian and multilingual Singaporean raters' accentedness judgements were significantly associated with three out of the four pronunciation/fluency categories (segmentals, word stress, speech rate) at a $p < .01$ level (Bonferroni corrected) when their lexicogrammar scores were factored out. Next, another set of partial correlation analyses were conducted to explore to what extent lexicogrammar scores can predict the Canadian and Singaporean raters' accentedness scores when their pronunciation

scores were controlled for. The results showed no significant link between accentedness and

lexicogrammar scores in any rater context ($p < .01$).

TABLE 7

Last, to examine how the significant predictors (i.e., segmentals, word stress, speech rate)

interacted to predict the monolingual and multilingual raters' accentedness scores, we conducted

multiple regression analyses. To avoid problems with multicollinearity, the decision was made to

use only segmentals (pronunciation) and speech rate (fluency) as independent variables. Word

stress was eliminated because it was strongly correlated with segmentals ($r = .82$), and since both

of them conceptually correspond to the same aspect of L2 speech—correct pronunciation of

words. According to the regression models (see Table 8), the monolingual raters' accentedness

scores were mainly explained by pronunciation (segmentals for 70%), and, to a much lesser

degree by fluency (speech rate for 3%). The multilingual raters' accentedness scores were solely

predicted by pronunciation (segmentals for 75%). The results suggested that while the Canadian

raters paid minor attention to fluency information (3%), both Canadians and Singaporeans relied

on segmental information (70% for Canadians; 74% for Singaporeans) as a primary cue for their

foreign accentedness evaluations.

TABLE 8

Discussion

Whereas many researchers have emphasized that L2 oral ability should be analyzed from

various perspectives, such as comprehensibility and accentedness (e.g., Derwing & Munro,

2009), recent studies have begun to analyze what kinds of linguistic errors relatively hinder

native speakers' successful understanding of L2 speech, and their intuitive judgements of

perceived nativelikeness (e.g., Crowther et al., 2015; Isaacs & Trofimovich, 2012; Saito et al.,

2015, 2016a; Trofimovich & Isaacs, 2012). Our study re-examined this topic by focusing on

another affecting variable in L2 speech assessment—the role of rater background (monolinguals

vs. multilinguals).

In our precursor study (Saito & Shintani, 2016), we found that the multilingual

Singaporean raters tended to assign better comprehensibility scores, indicating that they could be

considered more lenient about the comprehensibility of Japanese-accented English speech. This

was arguably because these Singaporean raters took into account not only the pronunciation

information, but also the lexicogrammatical aspects of L2 speech during their L2

comprehensibility judgements. In the current study, we reported the results of their accentedness

judgements, providing two crucial findings regarding the complex relationship between

accentedness ratings, linguistic domains, and rater backgrounds. First, the multilingual

Singaporean raters assigned lower accentedness scores than the monolingual Canadian raters did:

the Singaporean raters were more tolerant of the nativelikeness of Japanese-accented English

speech than the Canadian raters. Second, both groups of raters actually used a similar processing

strategy for accent judgements by exclusively drawing on the phonological information of L2

speech without taking into account the appropriate/complex use of L2 lexicogrammar.

Our findings are consistent with previous L2 speech research showing that raters who are

familiar with foreign accents tend to show more lenient attitudes towards accented speech (e.g.,

Saito et al., 2016b; Winke et al., 2013). In the context of our study, one reason for the leniency of

the Singaporean English raters could have been their familiarity with the pronunciation errors

typical of many non-native speakers. This familiarity is due in part to their frequent contact with

non-native speakers with various L1 backgrounds in a multilingual context. For example, Jenkins

(2000) pointed that non-nativelike pronunciation is often observed in social interactions between

non-native speakers, resulting in little negative impact on their successful communication; such non-core features include mispronunciation of certain segmentals (e.g., interdental fricatives), schwa insertion (but not deletion) in complex syllables, and monotonous (but not wrong) prosody.

It is also important to mention that our definition of accentedness—how closely the linguistic profiles of an utterance approaches those of a native speaker—may have affected the Singaporean raters' lenient judgements relative to the Canadian raters. Whereas the Canadian raters' comparison point may have referred only to North American English, the Singaporean raters' concept of "nativelikeness" could have highlighted a wider range of accents including not only General American but also Received Pronunciation and Singaporean English. Consequently, there is a possibility that Japanese-accented speech was perceived more targetlike in the Singaporean raters' composite nativelikeness model than the Canadian raters' singly-constructed nativelikeness model.

According to the extensive literature on perceptual adaptation, with a sufficient amount of relevant listening experience native speakers can easily adjust their perceptual representations to recognize words and sentences which deviate from the pronunciation norms in a given society, such as acoustically-manipulated ambiguous speech (Norris, McQueen, & Cutler, 2003), deaf speech (McGarr, 1983), and foreign-accented speech (Bradlow & Bent, 2008). In this regard, the current study indicated that the multilingual Singaporean raters could quickly adjust to the Japanese-accented speech (with which they reported little familiarity), arguably because they were accustomed to understanding the L2 speech of many non-native speakers with various L1 backgrounds.

Following this line of thought, the findings of this study further promote our current understanding of the multifaceted role of rater background in the product and process of L2 speech assessment according to which aspects of L2 speech raters are asked to evaluate (i.e., comprehensibility vs. accentedness). For L2 comprehensibility judgements, accumulative experience in L2 accent-decoding helps listeners better capture what non-native speakers intend to say by using every available piece of linguistic information (not only pronunciation but also lexicogrammar).

For L2 accentedness judgements, multilingual raters who are familiar with various models of English may have relatively broad perspectives on nativelikeness (compared to monolingual raters), resulting in "leniency" towards what "nativelikeness" actually means. To make such accentedness judgements, both groups of raters seem to similarly focus only on pronunciation and fluency information instead of processing the actual content of L2 speech. This may allow raters to instantly check for linguistic nativelikeness simply based on how much L2 learners have acquired one of the most difficult domains of adult SLA—L2 pronunciation accuracy and fluency (Munro et al., 2010; Saito, 2015).

Implications

The findings of the current study (i.e., native speakers' intuition of L2 speech) can bring to light several crucial practical implications. By allowing native speakers to have an explicit understanding of their own intuitive and implicit behaviours during L2 speech assessment, for example, we can help them (a) minimize their unwanted individual variability in high-stakes proficiency exams, and (b) approach mutual intelligibility when interacting with other non-native speakers.

Rater Training

In high-stakes testing contexts, researchers have extensively discussed how to elaborate an optimal procedure to train native judges with a varied amount of relevant experience with foreign accents to demonstrate similar reactions to same L2 speech samples (Winke et al., 2013). As one potential approach, our findings suggest that monolingual raters without much familiarity with L2-accented speech, in particular, should be informed of how other raters with much relevant experience with L2 learning (e.g., ample opportunities to listen to particular foreign accented speech and/or to use multiple languages on a daily basis) are likely to judge the comprehensibility versus nativelikeness aspect of L2 accented speech. It is then important to point out that such experienced raters are generally capable of paying special attention not only to form (pronunciation, diversity), but also to meaning (lemma appropriateness), especially when they are to assess the comprehensibility of L2 speech. In contrast, the role of the rater aspect could be minor in the case of L2 nativelikeness judgements which may require experienced and inexperienced raters alike to make quick, intuitive and automatic decisions solely based on pronunciation information in accented speech.

Mutual Intelligibility

Another crucial topic in L2 speech literature concerns the role of native and non-native speakers in successful L2 communication. Whereas much attention has been directed towards explaining which linguistic errors non-native speakers should work on to improve the overall qualities of their L2 speech (speaker → listener intelligibility), very few studies have examined how native speakers should make efforts to understand accented speech for the purpose of mutual intelligibility (listener → talker intelligibility) (cf. Jenkins, 2000; Saito & van Poeteren, 2012). The findings of the current study suggested that an increase in native speakers' experience with various kinds of foreign accents may be related to their successful understanding of L2

speech. This could be due to more experienced raters' ability to attend to lexicogrammar usage in

L2 speech without being distracted by phonological errors typical of many non-native speakers.

Therefore, one intriguing direction for future research would be investigating whether exposing

inexperienced raters to a wide range of foreign accented speech can increase their familiarity

with non-nativelike phonological patterns, and in turn help enhance their optimal processing of

L2 speech (see Derwing, Munro, & Rossiter, 2002).

Conclusion

In line with previous L2 assessment research, our study showed that native speaking

raters with more experience with diversely accented L1/L2 speech (Singaporean multilinguals)

perceived differently than those without such experience (Canadian monolinguals). The former

raters demonstrated more lenient accent scores than the latter raters. At the same time, both of

them used a similar processing strategy to make such accentedness judgements—drawing

exclusively on the phonological aspects of L2 speech without taking into account the appropriate

and complex use of L2 lexicogrammar. Combined with our precursor research (Saito & Shintani,

2016), the results of the current study provided some knowledge of how monolingual and

multilingual native speakers differentially perceive two different constructs of L2 speech

(comprehensibility and accentedness). Building on the previous listener factor research (e.g.,

Winke et al., 2013), we found that not only raters' familiarity with particular foreign accents, but

also their L1 backgrounds (monolinguals vs. multilinguals), could be a crucial affecting variable

for their L2 assessment patterns.

References

Abrahamsson, N. & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language – listener perception *vs.* linguistic scrutiny. *Language Learning, 59*, 249–306.

Boersma, D., & Weenink, P. (2012). *Praat: Doing phonetics by computer version 5.3.14.* Retrieved from http://www.praat.org

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*, 159-175.

Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707-729.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*. Advance online publication. doi: 10.1093/applin/amt056

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly, 49*, 814-837.

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.

Derwing, T.M., Munro, M.J. & Rossiter, M. J. (2002) Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingualism and Multicultural Development, 23,* 245-259.

Derwing, T.M., Rossiter, M.J., Munro, M.J. & Thomson, R.I. (2004). L2 fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Deterding, D. (2007). *Singapore English*. Edinburgh University Press.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*, 135-159.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34,* 475-505.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Kang, O., Rubin, D., Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English, *Modern Language Journal, 94,* 554-566.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review, 64,* 459–489.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26,* 451-458.

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Munro, M. J., Derwing, T. M., & Burgess, C. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52*, 626-637.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: A

    review. *Journal of Phonetics*, *29*, 191-215.

Saito, K. (2011). *Beginner, intermediate and advanced Japanese learners of English in Canada.*

    Unpublished corpus of second language speech.

Saito, K. (2015). Experience effects on the development of late second language learners' oral

    proficiency. *Language Learning, 65*, 563-595.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate

    linguistic influences on L2 comprehensibility and accentedness: A validation and

    generalization study. *Applied Linguistics*. doi: 10.1093/applin/amv047

Saito, K., Trofimovich, P., & Isaacs, T. (2016a). Second language speech production:

    Investigating linguistic correlates of comprehensibility and accentedness for learners at

    different ability levels. *Applied Psycholinguistics, 37*, 217-240

Saito, K., Trofimovich., P., Isaacs, T., & Webb., S. (2016b). Re-examining phonological and

    lexical correlates of second Language comprehensibility: The role of rater experience. In

    T. Isaacs & P. Trofimovich (Eds.). *Interfaces in second language pronunciation*

    assessment: Interdisciplinary perspectives. Bristol, UK: Multilingual Matters.

Saito, K., & van Poeteren, K. (2012). Pronunciation-specific adjustment strategies for

    intelligibility in L2 teacher talk: Results and implications of a questionnaire study.

    Language Awareness, 21, 369-385.

Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards

    pronunciation. *World Englishes, 30,* 392-408.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility.

    Bilingualism: Language and Cognition, *15*, 905-916.

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in

    rating oral performance. *Language Testing, 30*, 231-252.

Table 1

Summary of Variance in L2 Comprehensibility Judgements Explained by Linguistic Factors

	Canadian raters	Singaporean raters
Pronunciation and fluency information	79%	64%
Lexicogrammar information	5%	16%
Other information	16%	20%

Table 2

Length of Residence and Age of Arrival Profiles for 50 Japanese Speakers

Length of residence	*n*	Age of arrival	*n*
0-0.5 years	13	19-24 years	13
0.5-1 years	8	24-28 years	12
1-3 years	11	28-30 years	9
3-5 years	8	30-32 years	8
5-11 years	10	32-40years	8
Total	50	Total	50

Table 3

Mean Ratio of Singaporean Raters' Language Use

	At home		At school		With friends	
	M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
English	68.0	17.7	82.5	13.8	74.0	14.3
Mandarin	21.5	21.1	11.0	6.1	18.5	11.3
Other Chinese dialects	5.0	7.1	2.5	4.2	3.0	3.5
Malay	2.5	6.3	3.0	6.7	2.5	4.2
Tamil	3.0	6.3	1.0	3.2	2.0	4.2

Table 4

Mean Ratio of Singaporean Raters' English Use

	M (%)	*SD* (%)
General American	17.0	10.3
Received Pronunciation	12.5	14.2
Singapore English	70.5	22.3

Table 5

Summary of Accentedness and Linguistics Scores of 50 Japanese Learners

	M	SD	Range
Monolingual raters	6.7	1.4	2.2-8.6
Multilingual raters	6.0	0.9	3.3-7.8

Note. *1 = little accent, 9 = heavily accented*

Table 6

Summary of Linguistics Scores of 50 Japanese Learners

A. Pronunciation and fluency

	M	*SD*	*Range*
Segmentals	354	147	70-840
Word stress	429	119	240-810
Intonation	326	134	120-770
Speech rate	463	198	100-830

B. Lexicogrammar

	M	*SD*	*Range*
Lexical appropriateness	714	125	410-930
Lexical richness	387	192	60-860
Grammatical accuracy	482	300	80-890
Grammatical complexity	294	162	60-750

Note. *0 = non-targetlike, 1000 = targetlike*

Table 7

Partial Correlations between the Pronunciation and Lexicogrammar Variables and

Accentedness Judged by Monolingual and Multilingual Raters

Linguistic variables	Accentedness (Monolingual raters)	Accentedness (Multilingual raters)
Segmental errors[a]	.75*	.81*
Word stress[a]	.62*	.64*
Intonation[a]	.36	.26
Speech rate[a]	.43*	.38*
Lexical appropriateness[b]	.16	.08
Lexical richness[b]	.02	.12
Grammatical accuracy[b]	.31	.26
Grammatical complexity[b]	.06	.15

Note. *α < .01 (Bonferroni corrected). [a]Variables partialled out from each correlation include lexical appropriateness and richness, and grammatical accuracy and complexity. [b]Variables partialled out from each correlation include vowel/consonant errors, word stress, intonation, and speech rate.

Table 8

Results of Multiple Regression Analyses Using the Segmental and Temporal Factors as

Predictors of Accentedness

Predicted variable	Predictor variables	*Adjusted R²*	*R² change*	*F*	*p*
Accntedness	Segmentals	.70	.70	116.17	*p* < .0001
(Canadian raters)	Speech rate	.73	.03	67.369	*p* < .0001
Accentedness (Singaporean raters)	Segmentals	.74	.74	146.282	*p* < .0001

Note. The variables entered into the regression equations included segmentals and speech rate.