Title:
Do Native Speakers of North American and Singapore English Differentially Perceive Comprehensibility in Second Language Speech?

Running Titles:
L2 COMPREHENSIBILITY REVISITED

Authors:
Kazuya Saito (Birkbeck, University of London)
Natsuko Shintani (University of Auckland)

Corresponding author:
Kazuya Saito
Birkbeck, University of London
The Department of Applied Linguistics and Communication
30 Russell Square, London WC1B 5DT, UK.
Email: k.saito@bbk.ac.uk

Abstract

The current study examined the extent to which native speakers of North American and Singapore English differentially perceive the comprehensibility (ease of understanding) of second language (L2) speech. Spontaneous speech samples elicited from 50 Japanese learners of English with various proficiency levels were first rated by 10 Canadian and 10 Singaporean raters for overall comprehensibility, and then submitted to pronunciation, fluency, vocabulary, and grammar analyses. Whereas the raters' comprehensibility judgements were generally influenced by phonological and temporal qualities as primary cues, and, to a lesser degree, lexical and grammatical qualities of L2 speech as secondary cues, their linguistic backgrounds did make some impact on their L2 speech assessment patterns. The Singaporean raters, who not only used various models of English but also spoke a few L2s on a daily basis in a multilingual environment, tended to assign more lenient comprehensibility scores due to their relatively high sensitivity to, in particular, lexicogrammatical information. On the other hand, the comprehensibility judgements of the Canadian raters, who used only North American English in a monolingual environment, were mainly determined by the phonological accuracy and fluency of the L2 speech.

Do Native Speakers of North American and Singapore English Differentially Perceive Comprehensibility in Second Language Speech?

Background

Second Language Comprehensibility

It is commonly accepted that English has become a lingua franca worldwide, allowing for communication not only between native speakers and non-native speakers, but also between non-natives speakers themselves. Attaining an adequate speaking proficiency in the target language is considered as a first crucial step for second language (L2) learners of English to achieve their own career- and academic-related goals. Many L2 students (and even teachers) likely strive to speak like native speakers with little accent as an *ideal* goal (e.g., Derwing, 2003). However, it has been shown that it is extremely difficult for even young bilinguals to achieve native-like proficiency in the target language (e.g., Abrahamsson & Hyltenstam, 2009). Given that speaking an L2 with an accent should be viewed as a normal characteristic of L2 speech, certain researchers have emphasized that adult L2 learners should be encouraged to set more *realistic* goals, such as improving comprehensibility, defined as "the listener's perception of how easy or difficult it is to understand a given speech sample" (Derwing & Munro, 2009, p. 478)..

In Derwing and Munro's early work (Derwing & Munro, 1997; Munro & Derwing, 1995), spontaneous speech samples elicited from non-native speakers with a wide range of proficiency levels were judged by native (and non-native) speaking raters based on a 9-point scale for comprehensibility (*1 = easy to understand, 9 = difficult to understand*) as well as accentedness (*1 = little accent, 9 = heavily accented*). The results showed that, although comprehensibility and accentedness partially overlap, they are essentially independent constructs, because not all linguistic errors relevant to accentedness equally hinder successful comprehensibility.

Follow-up studies have further revealed that L2 comprehensibility judgements seem to be sensitive to certain pronunciation and fluency errors. For example, Derwing and Munro (2006) found that comprehensibility was strongly related to segmental errors with high functional load, but irrelevant to those with low functional load. Hahn (2004) showed that native speakers' actual understanding of lectures was negatively influenced by the misplacement of intonation (see also Field, 2005 for word stress). When it comes to fluency, Munro and Derwing (2001) found that while native speakers positively perceived moderately sped-up L2 speech samples (i.e., 10% increase), very fast speech rates likely make negative impacts on their comprehensibility judgements. Kang, Rubin and Pickering (2010) showed that a range of suprasegmental features (e.g., pitch range, intonation, and pauses) accounted for approximately 50% of the variance in native speakers' L2 comprehensibility judgements. Though fewer in number, some L2 vocabulary and grammar research has also shown that native speakers' comprehension of accented speech is negatively related to the number of word choice (Fayer & Krasinski, 1987) and morphosyntactic errors (Varonis & Gass, 1982) included in the tokens.

Role of Rater Background

While much attention has been directed towards investigating the linguistic characteristics of comprehensible L2 speech, other veins of researchers have examined in-depth whether and to what degree raters' backgrounds, such as experience and familiarity with particular L2 accents, affect L2 comprehensibility. For example, Winke, Gass and Myford (2013) corroborated the influence of native speaking raters' prior L2 learning experience on their speaking proficiency judgements of TOEFL iBT ® tests. It was shown that raters who had studied the test takers' L1s assigned significantly higher rating scores than those who had not.

Similarly, Kennedy and Trofimovich (2008) showed that raters with more exposure to L2 speech (e.g., ESL teachers) showed more understanding of accented sentence productions, probably because of "their greater knowledge of how L2 speakers' pronunciation differs from that of native speakers" (p. 478). Finally, Isaacs and Thomson (2013) demonstrated that, during L2 comprehensibility evaluations, experienced raters (i.e., ESL professionals) were able to explain exactly why they had arrived at certain rating decisions by pinpointing a range of pronunciation errors (probably thanks to their extensive teaching experience), compared to novice raters (i.e., graduate students from non-linguistic fields).

Taken together, research evidence has shown that raters' L2 backgrounds (including accent familiarity and L2 learning experience) significantly affect their intuitive judgement of L2 speech, arguably because such rater variability inevitably changes the quality of the linguistic representations that raters draw on when making subjective judgements of incoming linguistic input data (Bradlow & Bent, 2008). If raters have more *relevant* backgrounds with accented speech (e.g., learning and teaching experience), they tend to judge L2 audio tokens more leniently as well as analyze their own rating processes more clearly than inexperienced and novice raters. From a practical standpoint, it is important to identify such rater effects on L2 speech assessment in order to help elaborate and provide specialized training to mediate such individual differences among professional raters, especially in high-stake testing environments (e.g., TOEFL®, IELTS). L2 research in this area is needed because these professional raters are expected to demonstrate their fairly consistent and reliable evaluations of test takers' L2 speaking proficiency with little variance in attitudes towards same accented-speech samples (Winke et al., 2013).

In the current investigation, we highlighted another interesting listener facet affecting raters' subjective judgements of L2 speech—raters' L1 (but not L2) experience with various kinds of English models rather than their particular experience with certain L2 accents. Our raters, native speakers of Canadian and Singapore English, assessed the comprehensibility of Japanese-accented English speech. These raters were unique in that they did not differ in terms of the lack of their contact with Japanese learners of English as well as their ESL/EFL teaching backgrounds. Yet, they significantly differed in that Singaporean raters spoke a few L2s as well as used various models of English (General American [GA], Received Pronunciation [RP], Singapore English [SE]) as their L1 in a multilingual environment, while Canadian raters used only North American English on a daily basis in a monolingual environment.

Singapore English

Among the four national languages (English, Mandarin, Tamil and Malay) in Singapore, English serves as the lingua franca in most formal contexts (e.g., school, work, and media). It is also the medium of instruction throughout the educational system (primary school to university). Although the other three national languages as well as other languages/dialects (e.g., Hokkien) are also used in informal conversations or local communities, there has been a radical shift in family languages from ethnic to English in recent decades, due to the strong emphasis on educational success in the country (Cavallaro & Chin, 2014; Gupta, 2006; Gupta & Yeok, 1995). Cavallaro and Chin (2014), for example, reported that the preferred home language has drastically shifted from ethnic languages to English in three decades (10.2% to 32.6% for Chinese speakers, 2.3% to 17.0% for Malay speakers, and 24.3% to 36.6% for Indian language speakers).

The linguistic profiles of SE have unique characteristics. While SE historically developed from British English between the 19th and 20th centuries, it has also extensively incorporated

GA, because of the increasingly strong influence of mass media and the Internet. For example, studies reported that there was increasing rhoticity (a trait of American English) in native speakers of Singapore English, and a distaste for the intrusive-r in non-rhotic accents (C. H. Tan & Gupta, 1992; Y.Y. Tan, 2012). At the same time, recent studies have also described a range of linguistic phenomenon idiosyncratic to SE in terms of phonology (e.g., reduced vowel inventory, monophthongization, and variation in the realisation of interdental fricatives), lexes (e.g., admixture of lexical items from contact languages), and morphosyntax (e. g., transfer from substrate languages, patterns of regularisation (Leimgruber, 2014). As a result, SE can be characterized as a multifaceted system constituting both features commonly used in GA and RP, as well as idiosyncratic features independent of any external standard (for a comprehensive review, see Brown, 1999; Deterding, 2007).

Furthermore, Singaporeans are exposed to a variety of non-native English accents in their daily life. The Filipino accent is commonly heard due to the fact that there are many Filipino expatriates working as domestic workers in Singapore. Accents originating from India, Hong Kong, China, Indonesia, and Malaysia can also be heard among the population (see for example, Lalwani, Lwin, & Li, 2005). It has indeed been shown that native speakers of SE in the multilingual environment show much familiarity with various kinds of L1-accented speech and high sensitivities to distinct accents among the different ethnic groups. A number of studies (e.g., Deterding & Poedjosoedarmo, 2000) reported that Singaporean listeners can determine the ethnicity of a speaker on the basis of just ten seconds of speech. This might reflect inter-ethnic differences in intonation (Lim, 2000) and substantial differences in the acoustic realization of stress among ethnic groups (Y.Y. Tan, 2002).

In summation, Singaporeans tend to have highly multilingual backgrounds thanks to (a) their use of SE, which is somewhat independent of RP and GA; (b) hearing GA and RP through television and the Internet; and (c) being exposed to a wide variety of foreign-accented English speech. In this regard, examining these multilingual speakers' evaluation of L2 comprehensibility (relative to that of monolingual speakers of North American English) would be an ideal testing ground for advancing our understanding of the multifaceted role of rater background in perceived L2 speaking proficiency.

Motivation for Current Study

Recently, certain researchers have begun to examine in-depth the complex mechanisms underlying L2 comprehensibility judgements—what kind of linguistic information raters consciously and/or intuitively take into account while assessing comprehensibility in L2 speech (e.g., Crowther, Trofimovich, Saito, & Isaacs, in press; Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, forthcoming; Saito, Trofimovich, & Isaacs, in press; Trofimovich & Isaacs, 2012). This kind of research has directly investigated the actual linguistic properties of the tokens which raters evaluated for overall comprehensibility. As such, these researchers have aimed to examine which aspects of language—pronunciation, fluency, vocabulary and grammar—*interact* to determine raters' intuitive judgements of high, mid and low level comprehensibility in L2 speech.

For example, Saito et al. (forthcoming) explored how 20 native speaking raters evaluated comprehensibility and accentedness in the spontaneous speech samples of 40 Francophone learners. The results showed that raters paid exclusive attention to segmental accuracy during their accentedness judgements. For their comprehensibility judgements, the raters equally focused on the phonological (segmentals, prosody), temporal (speech rate), lexical (appropriateness, richness, abstractness, sophistication), grammatical (accuracy, complexity) and

semantic (sense relations) aspects of language. Such findings were subsequently replicated in follow up studies with 120 beginner-to-advanced Japanese learners of English (Saito et al., in press) as well as 45 intermediate ESL students with various L1 backgrounds (Crowther et al., in press).

Taken together, the results of these studies suggest that native speaking raters tend to pay exclusive attention to one of the most difficult features of L2 speech—segmental accuracy—when asked to assess accentedness. Thus, their accentedness judgements are invariably *fast, effortless*, and *intuitive* (Munro, Derwing, & Burgess, 2010). While gauging comprehensibility, on the other hand, native speaking raters likely rely on every domain of L2 speech (i.e., pronunciation, fluency, vocabulary, grammar) to arrive at overall meaning in a *timely* and *efficient* way (Munro & Derwing, 1995). With a view of attaining comprehensible (rather than nativelike) speech, the findings suggest that an optimal syllabus should help L2 students reach minimum phonological, temporal, lexical and grammatical requirements via interfacing pronunciation, fluency, vocabulary and grammar teaching perspectives.

Importantly, such pedagogical implications need to be interpreted with caution, because all of the relevant findings have been solely based on North American English raters' notions of comprehensibility. In a global society whereby native and non-native speaking interlocutors use English as an international language of communication, various models of English can be commonly used (Jenkins, 2002). Given that many L2 learners' speech is most likely judged by raters with potentially diverse linguistic backgrounds in an increasingly globalized world, it is important to scrutinize how L2 comprehensibility is perceived by raters who use various models of English in a multilingual context, such as native speakers of SE. It is possible that these multilingual Singaporean raters may perceive comprehensibility L2 English speech in a significantly different manner from monolingual GA raters (for details, see below). Accordingly, the current investigation was designed to address two research questions: (a) Do Canadian and Singaporean raters differentially perceive comprehensibility? and (b) If so, which aspects of language—pronunciation, fluency, vocabulary and grammar—relatively predict their potentially different perceptions of comprehensibility?

Method

The study concerns three parts of data collection and analyses: (a) the collection of the oral L2 data produced by 50 Japanese learners of English; (b) the comprehensibility judgements of the oral data by 20 Canadian and Singaporean raters; and (c) the rating of the oral data and transcripts by five linguistically-trained coders.

Collection of Oral L2 Data

Participants. The L2 speech data originates from our unpublished corpus of audio recordings of +200 Japanese learners of English in Montreal and Vancouver who completed various speaking tasks (Saito, 2011). To ensure a wide range of speaking proficiency specifically for the current investigation, 50 Japanese learners were carefully selected based on (a) their amount of L2 immersion and (b) age of arrival in Canada, both of which have been identified as crucial developmental indices for L2 speech learning (for a review, see Piske, MacKay & Flege, 2001).

As shown in Table 1, their length of residence (M = 2.7 years, SD = 3.1) and age of acquisition (M = 27.8 years, SD = 5.4) profiles were equally distributed from 1 month to 11 years.

According to the individual interviews, all of the participants reported English as their primary language of communication, with a high level of professional and integrative motivation. Whereas most of the beginner-to-intermediate participants (e.g., length of residence < 3 years)

were studying abroad at private language institutes, the intermediate-to-advanced participants (e.g., length of residence > 3 years) were likely residents in Montreal (who had speaking partners and/or conducted business with English speaking customers).

TABLE 1 HERE

Procedure. In the current study, the speech of Japanese talkers with a wide range of L2 proficiency levels was elicited via a timed picture description task. As conceptualized in Saito et al. (in press), the task adopted in the study was carefully designed to elicit a certain length of spontaneous speech data without excessive hesitations and dysfluencies.

First, instead of using a series of thematically-linked images (e.g., Derwing & Munro, 2009), speakers described seven separate pictures (e.g., Munro & Mann, 2005) with three keywords printed as hints. Second, to control for task familiarity, the first four pictures were used as practice and the last three were targeted for analyses. Third, to minimize the amount of conscious speech monitoring (see Ellis, 2005), speakers were given a marginal amount of planning time (i.e., only 5 sec) before describing each picture. In this way, participants were induced to pay equal attention to the phonological, temporal, lexical, grammatical, and discoursal domains of language to convey their communicative intentions (Spada & Tomita, 2010) under time pressure (Ellis, 2005).

The three target pictures depicted a table left out in a driveway in heavy rain (keywords: rain, table, driveway), three men playing rock music with one singing a song and the other two playing guitars (keywords: three guys, guitar, rock music), and 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 particularly problematic segmental and syllable structure features and demonstrate their pronunciation abilities. For instance, Japanese speakers have been reported to neutralize the English /r/-/l/ contrast ("rain, rock, brew, crowd" vs. "lane, lock, blue, cloud") and to insert epenthetic vowels between consecutive consonants (/dəraɪvə/ for "drive," /θəri/ for "three," /səkaɪ/ for "sky") and after word-final consonants (/teɪbələ/ for "table," /myuzɪkə/ for "music") in borrowed words (i.e., Katakana).

Preparation of the stimuli. All speech recording was carried out individually in university labs using a digital Roland-05 audio recorder (44.1 kHz sampling rate with 16-bit quantization). To ensure that all speakers understood the procedure, the researcher (a native speaker of Japanese) delivered all instructions in Japanese. The last three of the seven pictures described by the speakers were used for the main analysis. On average, about 10 sec (range: 7.2 to 12.7 sec) from the beginning of each description were extracted for each speaker. Each participant contributed 30 sec of spontaneous speech, which is comparable to previous L2 speech research (e.g., Derwing & Munro, 1997 for 15sec, Hopp & Schmid, 2013 for 15 sec; Isaacs & Thomson, 2013 for 30 sec). In total, 150 speech samples were created from 50 Japanese learners (50 talkers × 3 pictures).

Comprehensibility Judgements

Participants. To assess the comprehensibility of these L2 tokens, 20 native speaking raters of north-western American English (n = 10) as well as Singapore English (n = 10) without any teaching experience were recruited at English-speaking universities in Vancouver and Singapore.

The Canadian raters were all undergraduate students majoring in business and psychology at the university in Vancouver. According to the language background questionnaire,

their average self-reported familiarity with Japanese-accented English was 1.3 (*SD* = 0.5) on a 6-point scale (*1 = Not at all, 6 = Very much*). They reported no ESL/EFL experience.[1] While three of them reported beginner or intermediate knowledge in French, they did not use it in their predominantly English speaking environment. The other seven raters considered themselves as monolinguals. Since all of the Canadian raters presumably had varied degrees of exposure to French, we used an inclusive definition of monolinguals for these raters in the sense that they lacked any significant amount of extensive experience in second and foreign language learning on a daily basis. These *relatively monolingual* Canadian raters were sharply contrasted with the *highly multilingual* Singaporean raters, who used several languages for communication under multilingual conditions.

Similar to the Canadian raters, the Singaporean raters were undergraduate and graduate students with non-linguistic majors (e.g., business, humanities). They also reported little familiarity and contact with Japanese learners of English (*M* = 1 from 1 = *Not at all* to 6 = *Very much*) as well as no ESL/EFL teaching experience. According to the language background interview, these raters had unique multilingual backgrounds. While all of them considered English as their most proficient and comfortably used language, due to using it dominantly on a daily basis at school, work and home, they reported different levels of proficiency in other languages (Mandarin, Malay, Tamil) (see Table 2). They also reported that their use of L1 and L2 varied according to the context (home vs. school vs. friends) (see Table 3).

TABLE 2 HERE

TABLE 3 HERE

In terms of their use of SE, GA, and RP, as described in Table 4, the raters mainly used Singapore English (*M* = 70.5%, *SD* = 22.3), particularly in conversations with friends and family members; however, they all reported a good amount of exposure to GA and RP, typically through mass media (e.g., TV, Internet).

TABLE 4 HERE

Thus, the two groups of raters were comparable in terms of the lack of familiarity with Japanese-accented English and any relevant experience in ESL/EFL teaching. Yet, they were substantially different in terms of their multilingual backgrounds. While the Canadian raters were characterized as monolingual in nature (focusing on North American English), the Singaporean raters demonstrated not only varied proficiency in languages other than English (Mandarin, Malay, Tamil), but also their use and exposure to a range of English models (SE, GA, RP).

Procedure. To ensure the comparability of findings with the previous literature, we carefully followed the rating method used in previous research (e.g., Derwing & Munro, 2009; Isaacs & Trofimovich, 2012). The comprehensibility session took place individually at quiet rooms at both universities. After familiarizing themselves with the picture prompts, the raters

[1] One Canadian participant pointed out that he had occasionally helped his/her non-native speaking friends with their academic assignments (e.g., tutoring, proofreading). Since his volunteer experience here focused on both content and language, we did not count this as professional ESL/EFL teaching experience.

were presented with the speech samples in a randomized order via the *Praat* software (Boersma & Weenik, 2012). To simulate the raters' intuitive judgements of comprehensibility in real conversations, they were asked to listen to each picture description only once and to rate comprehensibility on a 9-point numerical scale (1 = *easy to understand*, 9 = *hard to understand*). Since the dataset represent a wide range of L2 oral proficiency (from beginners to experts), the raters were always reminded to use the entire scale as much as possible (from 1 to 9). The comprehensibility scores they chose and assigned to each speech sample were digitally recorded via *Praat*, and later transferred to an excel spread sheet. Given that the entire session lasted for approximately 1.5 hours, they took a five-minute break halfway through to avoid fatigue.

Pronunciation, Fluency, Vocabulary and Grammar Analyses

To examine what kinds of linguistic errors in L2 speech led the raters to make certain comprehensibility judgements, the Japanese talkers' spontaneous speech samples were also coded by linguistically-trained raters for the eight audio-based (segmentals, word stress, intonation, speech rate) and transcript-based (lexical appropriateness and richness, grammatical accuracy and complexity) categories elaborated and validated in the precursor study (Saito, Trofimovich, & Isaacs, in press, forthcoming).

Participants. In line with the definition of linguistically-trained raters by Isaacs and Thomson (2013), five native speaking coders (3 males, 2 females) were recruited based on their linguistic and pedagogical experience. All of them were graduate students in the Department of English at a university in Montreal, and had received training in phonetics and phonology as well as previously taught English in ESL/EFL settings (M = 4.0 years from 2 to 6 years).

Procedure. First, three picture descriptions by the same participant were combined and stored in a single WAV file with a total mean length of 30sec (range: 20.1-36.8sec). This was done to provide the trained raters with sufficient phonological information for their pronunciation and fluency judgements. The raters listened to and evaluated the 50 speech samples using the following phonological and temporal 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). All speech samples were played in a randomized order via the software MATLAB. The raters were able to listen to each speech file as many times they needed to, and used a free moving slider on a computer screen to assess each of the linguistic qualities of the speech samples. If the slider was placed at the leftmost end of the continuum, marked with a frowning face (indicating very negative), it was recorded as "0"; if it was placed at the rightmost end of the continuum, marked with a smiley face (indicating very positive), it was recorded as "1000". The slider was initially placed in the middle of each scale. The raters were encouraged to use the entire scale (0-1000) to assess beginner to advanced Japanese learners.

Second, the 50 speech samples were transcribed, edited and cleaned up by modifying certain audio-related errors, such as mispronunciations of given target words (e.g., rock music spoken as lock music, table spoken as devil) based on the contextual information of the pictures (outside spoken as ought side was transcribed as outside, lonely spoken as lawn Lee was transcribed as lonely) as well as orthographic markings of pausing (e.g., uh, um, oh, ehh). The modified written transcripts were assessed for the following lexical and grammatical dimensions: (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). Similar to the audio-based measures, three picture descriptions were delineated with dots and

stored in a transcript file per talker. The raters read each file displayed on a computer screen via MATLAB, and provided vocabulary and grammar ratings by using a moving slider on a 1000-point scale (*0 = non-targetlike, 1000 = targetlike*).

As reported in Saito et al. (forthcoming), all raters first underwent a three-hour training session, whereby they received through instruction on each linguistic category and practiced with a number of non-native speech samples (not included in the current analysis) (for all training materials and onscreen labels, see Saito et al., in press, forthcoming). According to the results, these trained raters showed clear understanding of each linguistic category as well as high inter-rater agreement (Cronbach alpha > .90) for each category. Importantly, their judgement scores were significantly correlated with the actual linguistic properties objectively measured through computer software (e.g., Boersma & Weenik, 2012 for *Praat*; Cobb, 2012 for *Lexical Tutor*) (see Table 5).

TABLE 5 HERE

After the training, the raters proceeded to the rating session for the current study, which lasted for two days: Day 1 for audio ratings (approximately 1.5 hours) and Day 2 for transcript ratings (approximately 0.5 hours).

Results

Comprehensibility Ratings

According to the results of the Cronbach's alpha analysis, high reliability for 9-point comprehensibility ratings were identified among 10 Canadian raters ($\alpha = .95$) as well as 10 Singapore raters ($\alpha = .93$). Therefore, two mean rating scores were calculated for each of the participants by pooling over the Canadian and Singaporean raters, respectively (for descriptive statistics, see Table 6).

To see if the Canadian and Singaporean raters differentially assessed the comprehensibility in Japanese talkers' speech samples, their ratings were then submitted to an independent *t* test. The results showed that Singaporean raters assigned more lenient comprehensibility scores ($M = 4.0$, $SD = 1.2$) than the Canadian raters ($M = 4.7$, $SD = 1.5$), $t(98) = 2.778$, $p = .007$. $d = 0.52$.

TABLE 6 HERE

Pronunciation, Fluency, Vocabulary and Grammar Ratings

Given the high reliability and accuracy in the experienced coders' linguistic analyses reported in the precursor validation study (Saito et al., forthcoming), the linguistic rating scores provided by the coders in the current study were averaged for each participant's speech sample. As summarized in Table 7, each token was assessed based on a 1000-point scale (*0 = non-targetlike, 1000 = targetlike*) according to pronunciation (segmentals, word stress, intonation), fluency (speech rate), vocabulary (appropriateness, richness) and grammar (accuracy, complexity).

TABLE 7 HERE

Linguistic Correlates of Comprehensibility

Next, our analysis focused on the phonological, temporal, lexical and grammatical influences on two kinds of comprehensibility judgments by Canadian and Singaporean raters. First, a set of partial correlation analyses were conducted to see how the four pronunciation and fluency scores (segmentals, word stress, intonation, speech rate) were associated with Canadian and Singaporeans' comprehensibility judgements, respectively, when their lexicogrammar scores were factored out. The results showed that both Canadian and Singaporean raters' comprehensibility judgements were significantly related to the segmental, prosodic and temporal qualities of the L2 speech samples (but not intonation accuracy) at a $p < .01$ level (Bonferroni corrected).

A different set of partial correlation analyses were implemented to investigate the lexicogrammar correlates of comprehensibility when the influence of pronunciation and fluency scores were statistically controlled for. The results also showed that, whereas the Canadian raters' comprehensibility scores were significantly correlated only with grammatical accuracy ($p < .01$), the Singaporean raters' comprehensibility scores were equally interrelated to both lexical appropriateness and grammatical accuracy.

TABLE 8 HERE

Finally, our analysis probed how these phonological, temporal and lexicogrammar variables *interacted* to affect Canadians and Singaporean raters' comprehensibility judgements. To this end, stepwise multiple regression analyses were performed on their comprehensibility ratings as dependent variables, and pronunciation, fluency, vocabulary and grammar ratings as independent variables. To avoid collinearity problems, the decision was made to reduce the number of independent variables by dropping the above-mentioned non-significant predictors (i.e., intonation, lexical richness, grammatical complexity). Although both segmentals and word stress were equally related to comprehensibility, another decision was made to use only the coders' segmental (but not word stress) scores as an independent variable, since these two variables were strongly correlated with each other, $r = .82$. It may be the case that the raters in this study more equally evaluated the quality of segmentals and lexical stress because these two phonological domains tap into a conceptually-overlapping phenomenon: talkers' correct pronunciation of words.

With respect to the Canadian raters, the regression model accounted for 83% for variance in comprehensibility, $F(3, 46) = 81.954$, $p < .001$, with no evidence of strong collinearity ($VIF < 1.787$). According to the results summarized in Table 9, whereas the pronunciation and fluency factors (segmentals, speech rate for 79%) mainly predicted the Canadian raters' comprehensibility judgements, the grammar variable played a minor role (grammatical accuracy for 5%). In terms of the Singaporean raters, the regression model explained 79% of the variance in comprehensibility, $F(3, 46) = 63.026$, $p < .001$; strong collinearity was not found ($VIF < 1.819$). The results illustrated that not only the pronunciation (64%) but also lexical (16%) factors significantly contributed to the Singaporean raters' comprehensibility judgements.

TABLE 9 HERE

Discussion

In conjunction with the increasing importance of English as a language of international communication in many parts of the world, L2 research has extensively examined the linguistic correlates of comprehensible speech from the perspectives of pronunciation, fluency, vocabulary and grammar (e.g., Crowther et al., in press; Isaacs & Trofimovich, 2012; Saito et al., forthcoming, in press; Trofimovich & Isaacs, 2012). Such comprehensibility in previous studies, however, has typically been perceived by monolingual raters of North American English and/or native English speakers with different levels of familiarity with accented speech. To date, some studies have begun to examine in depth the role of rater variability in L2 comprehensibility judgements by carefully controlling the types of raters as an independent variable (e.g., Isaacs & Thomson, 2013). The current investigation is a novel contribution in that it aimed to examine whether and to what degree native speaking raters with multilingual backgrounds differentially perceive L2 comprehensibility, i.e., Singaporean raters who constantly used both various models of English (SE, GA, RP) as their dominant language but also spoke at least one other language (e.g., Mandarin, Malay, Tamil) for daily communication purposes.

First and foremost, both the Canadian and Singaporean raters exhibited high inter-rater agreement (Cronbach alpha = .95 and .93, respectively) in their subjective comprehensibility perceptions of the Japanese learners' L2 speech samples. The results in turn suggest that these naïve raters (who did not have any teaching backgrounds nor much familiarity with Japanese-accented English) did share an intuitive notion of what L2 comprehensibility means, despite not receiving much training based on pre-existing descriptors.

Our findings here are in line with previous L2 speech literature which has extensively shown that even linguistically-trained and naïve raters alike can use simple 7- or 9-point rating scales to reliably judge various linguistic domains of L2 speaking proficiency, such as the quality of vowels and consonants in L2 speech (Piske, MacKay, & Flege, 2001), global aspects of L2 speech, such as comprehensibility and accent (Isaacs & Thomson, 2013), as well as fluency characteristics of L2 speech (Bosker, Pinget, Quené, Sanders, & de Jong, 2013; Derwing, Rossiter, Munro, & Thomson, 2004). Following this line of thought, the current study echoed that the overall quality of L2 speaking proficiency can be measured via native speakers' intuitions: All of our Canadian and Singaporean raters in the study indeed made similar judgements on the overall quality of L2 speech on the continuum of comprehensibility without much disagreement.

At the same time, however, the study also showed that L2 comprehensibility judgements can be influenced by raters' linguistic (e.g., monolingual vs. multilingual) background to some degree. According to the results of the *t*-tests, the Singaporean raters assigned significantly better and more lenient comprehensibility scores to Japanese-accented speech samples than the Canadian raters did. The results of the correlation and regression analyses further revealed that, whereas the Canadian and Singaporean raters generally based their comprehensibility judgements on various linguistic domains of L2 speech spanning pronunciation, fluency, vocabulary and grammar, the relative contribution of the lexicogrammar element, in particular, appeared to be stronger for the latter (16%) than the former (5%) group of raters.

Why did the Canadian and Singaporean raters' linguistic backgrounds lead to such different comprehensibility judgement patterns? As discussed earlier, L2 comprehensibility well reflects the amount of effort raters make to collect as much linguistic information as possible from accented L2 speech in order to grasp its overall meaning (Derwing & Munro, 2009). Since the Singaporean raters had more relevant experience with various kinds of English accents on a daily basis (they regularly used various kinds of English models and spoke a few L2s), it seems

reasonable to assume that they easily and quickly adjusted to even completely unfamiliar accented English speech (produced by Japanese learners) with an equal focus on pronunciation and fluency as well as the vocabulary and grammar aspects of L2 speech. Specifically, the Singaporean raters were able to pay sufficient attention to lexicogrammar information—the extent to which the Japanese learners selected conceptually and contextually appropriate vocabulary items, and delivered them using correct grammatical forms. In contrast, the Canadian raters' comprehensibility judgements were mainly based on the phonological accuracy and fluency of the L2 speech.

Another possibility involves the raters' familiarity with some common linguistic characteristics of non-native speakers. Some scholars (e.g., Jenkins, 2002; Setter & Jenkins, 2005) have sought to descriptively identify which linguistic errors are shared by many non-native speakers with various L1 backgrounds without exerting much negative influence on speech intelligibility, such as mispronunciation of certain segmentals (e.g., interdental fricatives), schwa insertion in complex syllables, and monotonous (but not wrong) prosody. Since the Singaporean raters must have accumulated a great deal of experience in decoding and processing such non-nativelike speech signal, they may be able to attend to the universal characteristics of non-native speech in Japanese-accented English and understand it with relative ease. However, more research is warranted that specifically focuses on what linguistic characteristics overlap among non-native speakers of English in Singapore and Japan.

The results presented here concur with previous research evidence that L2 speech assessment can be somewhat subject to raters' individual variability, such as familiarity with L2 accents (e.g., Bradlow & Bent, 2008; Isaacs & Thomson, 2013; Kennedy & Trofimovich, 2008; Winke et al., 2013). These studies generally suggest that more experienced raters (e.g., linguistically-trained raters, ESL teachers, learners of non-native speakers' L1, bilinguals) have more lenient attitudes towards accented speech. Our study added that such raters' leniency per se can be attributed to their ability to sufficiently attend not only to main cues—pronunciation and fluency—but also to secondary cues—vocabulary and grammar—in making L2 comprehensibility judgements. Truly, extracting an overall message from accented speech is a cognitively demanding task, because it requires raters to check every possible piece of linguistic information simultaneously. Yet, multilingual raters, such as native speakers of SE, may not need much cognitive resources to do the task (relative to monolingual raters), thanks to their accumulative experience in L2 accent-decoding process under multilingual conditions (Isaacs & Thomson, 2013).

Implications for L2 Speech Assessment

Whereas the current study revealed the complex relationship between rater variability and the linguistic correlates of L2 comprehensibility judgement, these findings make a considerable contribution to the development of rater training studies examining how native speakers (especially those without much familiarity with accented speech) can better understand and interact with non-native speakers. For example, Derwing, Rossiter and Munro (2002) showed that explicit linguistic explanations on the accent patterns of Vietnamese learners of English enhanced native raters' confidence to successfully interact with these L2 learners in future communicative settings. To further improve the pedagogical value of such rater training, we suggest that raters should not only learn the linguistic characteristics of certain L2s, but also become aware of the fact that the successful understanding of accented L2 speech depends on their ability to attend to all linguistic domains (instead of exclusively relying only on the phonological aspects of L2 speech). This kind of research on rater behaviour and rater training is

crucial, especially in L2 testing settings where raters are expected to make reliable, consistent and adequate assessments of L2 speech while minimizing individual variability among raters (Winke et al., 2013).

Importantly, mutual intelligibility is by definition a bidirectional phenomenon between a speaker and a listener, both of whom are expected to make efforts to attain successful L2 communication (Jenkins, 2002). Although much research attention has been given towards investigating how non-native speakers should improve the comprehensibility of their utterances (the speaker → listener intelligibility) (Derwing & Munro, 2009), few empirical studies have ever illustrated how native speakers adjust their listening strategies to accented speech (the listener → speaker intelligibility) (cf. Saito & van Poeteren, 2012). To improve native listeners' abilities to decode lexicogrammar information in L2 speech in an optimal manner, the current study suggests that they should be frequently exposed to a wide range of phonological patterns of accented L2 speech, which will in turn increase the amount of their relevant experience in L2 word recognition and processing (Bradlow & Bent, 2008).

Implications for L2 Oral Teaching

The results of the study also provided several implications for teaching L2 oral skills. To improve comprehensibility in L2 speech, it would be a very crucial initial step for students and teachers alike to know how native speakers actually perceive L2 comprehensibility. For example, although pronunciation instruction has long been ignored in many ESL (Foote, Holtby, & Derwing, 2012) and EFL (Saito & van Poeteren, 2012) classrooms (for a review, see Isaacs, 2008), the study actually showed that native speakers mainly pay attention to phonological and temporal information in L2 speech as the primary cues for L2 comprehensibility. Thus, these findings suggest that students should receive explicit instruction especially on crucial segmentals and suprasegmentals for comprehensibility (e.g., Kissling, 2012; Thomson & Derwing, 2015) as well as receive a well-balanced mixture of controlled and communicative practice activities (e.g., Baker, 2014; Saito, 2012).

In addition, it is important to reiterate our finding here that the approximate use of lexicogrammar in L2 speech provided additional information for, in particular, Singaporean raters' comprehensibility judgements. This in turn indicates that teachers need to prioritize the lexical and grammatical aspects of learner speech as another important factor for improved speaking proficiency, especially when their students may potentially interact with various kinds of interlocutors (including bilingual and monolingual native speakers) in a globalized society. Previous research has found that L2 learners need to increase their vocabulary size beyond the first 3,000 word families in order to understand everyday spoken discourse (e.g., Van Zeeland & Schmitt, 2013) and other speech genres (e.g., Webb & Rodgers, 2009). Research in the field of psycholinguistics has also introduced a range of vocabulary indices highly relevant for L2 speaking proficiency, such as lexical diversity, richness, sophistication, abstractness and sense relations (e.g., Crossley, Salsbury, McNamara, & Jarvis, 2011). Most importantly, our study suggests that teachers should be encouraged to enhance students' accurate and fluent use of language via an integrative teaching approach which highlights a communicative focus on pronunciation, vocabulary, and grammar form, especially in the context of meaning-oriented and content-based classrooms.

Conclusion and Future Directions

The findings of the current study led to two overall conclusions. First, the results confirmed those of the precursor research (e.g., Saito et al., forthcoming), which showed that native speaking raters' comprehensibility judgements are influenced by phonological and

temporal qualities as primary cues, and, to a lesser degree, by lexical and grammatical domains of L2 speech as secondary cues. Second, the study also demonstrated that the raters' linguistic backgrounds did make some impact on L2 speech assessment patterns. The Singaporean raters, who not only used various models of English but also spoke a few L2s on a daily basis in a multilingual environment, tended to assign more lenient comprehensibility scores because of their relatively high sensitivity to, in particular, lexicogrammatical information. On the other hand, the comprehensibility judgements of the Canadian raters, who used only North American English in a monolingual environment, were mainly determined by the phonological accuracy and fluency of the L2 speech.

The study took an exploratory approach towards investigating the intricate relationship between raters' backgrounds, their comprehensibility judgements, and relevant linguistic factors relevant to the perceived comprehensibility. Therefore, it is important to address several limitations for the purpose of future replication studies of this kind. First, it needs to be emphasized that the findings were exclusively based on Japanese learners' picture descriptions. Since this type of task has been reported to be a variable which can affect L2 learners' language performance and quality in both the phonological (Derwing et al., 2004) and lexicogrammatical (Skehan, 2009) domains of L2 speech, it is crucial to see the generalizability of the findings using other task modalities, including more argumentative, formal and complex tasks whereby L2 learners are induced to demonstrate a more varied and sophisticated use of language (Hulstijn, Schoonen, de Jong, Steinel, & Florijn, 2012).

Another crucial concept that has not been featured in the study but has been extensively discussed in the previous L2 assessment literature is intelligibility. Different from comprehensibility (generally measured via raters' scalar ratings), intelligibility has been defined and measured in substantially different ways between individual studies, such as the accuracy of listeners' L2 speech transcriptions (e.g., Derwing & Munro, 1997) or comprehension questions related to the content of L2 speech (e.g., Hahn, 2004). In conjunction with the lack of any standard methodology for evaluating intelligibility (for a review, see Isaacs, 2008), future studies need to first test the validity of each method as a way to measure what raters actually understand after listening to accented L2 speech, and then explore which linguistic errors would be relatively detrimental to intelligibility.

Third, the underlying assumptions in the study regarding the relationships between raters' L2 exposure/use and the relative contribution of elements of the speech signal is broad, because our relevant discussion simply drew on a set of statistical differences. To remedy this, speech samples need to be scrutinized in more detail via objective measures such as *Praat* (Boersma, & Weenink, 2012) for segmental, prosodic, and temporal qualities, and *Coh-Metrix* (Graesser, McNamara, Louwerse, & Cai, 2004) for lexical and grammatical qualities of L2 speech. Then, it may be important to take not only a quantitative but also a qualitative approach (e.g., interviews) to examining in depth the way these Singaporean raters, in particular, interact with other non-native speakers of English under highly multilingual environments. It would also be intriguing to ask how their own perception of their multilingual backgrounds actually impacted their ratings of Japanese-accented English during their listening experiments (see Isaacs & Trofimovich, 2012 for a reading-aloud protocol during L2 speech assessment).

Finally, future studies also need to investigate listener facets in perceived L2 speaking proficiency, focusing especially on non-native speaking raters' comprehensibility judgements. Whereas non-native and native raters generally agree upon which L2 speech samples are the most difficult and easiest to understand (e.g., Munro, Derwing, & Morton., 2006), there is some

evidence that certain non-native raters have a better understanding of accented speech than native raters (e.g., Jenkins, 2002). This result sheds light on the existence of potentially different L2 speech assessment patterns among various types of non-native raters (e.g., beginner-, intermediate- and advanced-level non-native raters in ESL vs. EFL contexts).

The Authors

Kazuya Saito is a Junior Lecturer in the Department of Applied Linguistics and Communication at Birkbeck, University of London, UK. His program of research investigates how a combination of form-focused and meaning-oriented instruction can help adult learners develop their second language oral ability, especially in the domains of pronunciation, fluency, vocabulary, and grammar.

Natsuko Shintani is a Senior Lecturer at Faculty of Education, University of Auckland. She obtained her PhD in Language Teaching and Learning from the University of Auckland in 2011. Her research interests encompass roles of interaction in second language acquisition and task-based language teaching.

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. doi: 10.1111/j.1467-9922.2009.00507.x

Baker, A. (2014). Exploring Teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*, 136-163. doi 10.1002/tesq.99

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer*. 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. doi: 10.1177/0265532212455394

Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707-729. doi: 10.1016/j.cognition.2007.04.005

Brown, A. (1999). *Singapore English in a nutshell: an alphabetical description of its features*. Federal Publications.

Cobb, T. (2012). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca/vp/

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly, 45* (1), 182-193. doi: 10.5054/tq.2010.244019

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (in press). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*. doi: 10.1002/tesq.203

Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review, 59,* 545-564. doi: 10.3138/cmlr.59.4.547

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 12*, 1–16. doi:.10.1111/j.1467-1770.1995.tb00963.x

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490. doi: 10.1017/s026144480800551x

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. doi: 10.1080/01434630208666468

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.

Deterding, D., & Poedjosoedarmo, G. (2000). To what extent can the ethnic group of young Singaporeans be identified from their speech. *The English language in Singapore: Research on pronunciation*, 1-9. doi: 10.1111/j.1467-9922.2004.00282.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a Second Language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141-172. doi: 10.1017/s0272263105050096

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313–326. doi: 10.1111/j.1467-1770.1987.tb00573.x

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399-423. doi: 10.2307/3588487

Foote, J., Holtby, A., & Derwing, T. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*, 1-22.

Gupta, A. F. (2006). The situation of English in Singapore. *World Englishes: Critical Concepts in Linguistics, 2*, 369.

Gupta, A. F., & Yeok, S. P. (1995). Language shift in a Singapore family. *Journal of Multilingual and Multicultural Development, 16*, 301-314. doi: 10.1080/01434632.1995.9994609

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 193-202. doi: 10.3758/bf03195564

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201-223. doi: 10.2307/3588378

Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics, 34*, 361-394. doi: 10.1017/s0142716411000737

Hulstijn, J.H., Schoonen, R., De Jong, N.H., Steinel, M.P, & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing, 29*, 203 - 221. doi: 10.1177/0265532211419826

Isaacs, T., (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review 64*, 555–580. doi: 10.3138/cmlr.64.4.555

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. doi 10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34,* 475-505. doi: 10.1017/s0272263112000150

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*, 83–103. doi: 10.1093/applin/23.1.83

Kang, O., Rubin, D., & Pickering, L. (2010). Supresegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal, 94*, 554-566. doi: 10.1111/j.1540-4781.2010.01091.x

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review, 64,* 459–489. doi: 10.3138/cmlr.64.3.459

Lalwani, A. K., Lwin, M., & Li, K. L. (2005). Consumer Responses to English Accent Variations in Advertising. *Journal of Global Marketing, 18*(3-4), 143-165. doi: 10.1300/J042v18n03_07

Leimgruber, J. R. (2014). Singlish as defined by young educated Chinese Singaporeans. *International Journal of the Sociology of Language, 2014*(230), 45-63.

Lim, L. (2000). Ethnic group differences aligned? Intonation patterns of Chinese, Indian and Malay Singaporean English. . In A. Brown, D. Deterding & L. E. Ling (Eds.), *The English language in Singapore* (pp. 20-21). Singapore Singapore Association of Applied Linguistics.

Munro, M., Derwing, T., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111–131. doi: 10.1017/s0272263106060049

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468. doi: 10.1111/lang.12082

Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*, 520–531. doi: 10.1016/j.system.2006.09.004

Munro, M. J., Derwing, T. M., & Burgess, C. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52*, 626-637. doi: 10.1016/j.specom.2010.02.013

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: a review. *Journal of Phonetics*, *29*, 191–215. doi: 10.1006/jpho.2001.0134

Munro, M., & Mann, V. (2005). Age of Immersion as a predictor of foreign accent. *Applied Psycholinguistics, 26,* 311-341. doi: 10.1017/s0142716405050198

Saito, K. (2011). *Beginner, intermediate and advanced Japanese learners of English in Canada.* Unpublished corpus of second language speech.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 842-854. doi: 10.1002/tesq.67

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. doi: 10.1080/09658416.2011.643891

Saito, K., Trofimovich, P., & Isaacs, T. (in press). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*. doi: 10.1017/s0142716414000502

Saito, K., Trofimovich, P., & Isaacs, T. (forthcoming). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*, 510–532. doi 10.1093/applin/amp047

Tan, C. H., & Gupta, A. F. (1992). Post-vocalic /r/ in Singapore English. *York Papers in Linguistics, 16*, 139-152.

Tan, Y. Y. (2002). *The acoustic and perceptual properties of stress in the ethnic subvarieties of Singapore English.* (PhD Unpublished Doctoral Thesis), National University of Singapore, Singapore.

Tan, Y. Y. (2012). Age as a factor in ethnic accent identification in Singapore. *Journal of Multilingual and Multicultural Development, 33*, 569-587. doi: 10.1080/01434632.2012.682730

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36.* doi: 10.1093/applin/amu076

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition, 4*, 114–136. doi: 10.1017/s027226310000437x

Webb, S., & Rodgers, M. P. H. (2009). The vocabulary demands of television programs. *Language Learning, 59*, 335–366. doi: 10.1111/j.1467-9922.2009.00509.x

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. doi: 10.1177/0265532212456968

Table 1.
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 2.
Singaporean Raters' Language Backgrounds

	Most dominant language	Father's L1	Mother's L1	Proficiency in other languages (*1 = Beginner, 2 = Intermediate, 3 = Advanced*)
Rater 1	English	English	English	Mandarin (3), Malay (1), Tamil (1)
Rater 2	English	English	Mandarin	Mandarin (2)
Rater 3	English	English	English	Mandarin (2)
Rater 4	English	English	Mandarin	Mandarin (3)
Rater 5	English	Mandarin	Mandarin	Mandarin (3)
Rater 6	English	English	English	Tamil (3), Mandarin (1), Malay (1)
Rater 7	English	English	Mandarin	Mandarin (3)
Rater 8	English	English	English	Mandarin (2)
Rater 9	English	English	English	Tamil (3)
Rater 10	English	Malay	Malay	Malay (2), Mandarin (2)

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' Use of English Models

	M (%)	*SD* (%)	Contents
General American	17.0	10.3	TV (10 raters), Internet (9 raters), movies (5 raters)
Received Pronunciation	12.5	14.2	TV (6 raters), Internet (5 raters), school (2 raters), friends (1 rater), family (1 rater)
Singapore English	70.5	22.3	Friends (10 raters), family (6 raters), school (6 raters), TV (2 raters), Internet (1 rater)

Table 5.
Summary of Linguistic Predictors for Trained Raters' Pronunciation, Fluency, Vocabulary and Grammar Judgment in Saito et al. (forthcoming)

Rater judgment measures	Linguistic predictors
A. Phonology and fluency	
Segmentals	No. of vowel and consonant errors
Word stress	No. of word stress errors
Intonation	No. of intonation errors
Speech rate	Mean length of run; no. of unfilled pauses; articulation rate
B. Lexicogrammar	
Lexical appropriateness	No. of lexical errors
Lexical richness	Type frequency, token frequency
Grammatical accuracy	No. of grammatical errors
Grammatical complexity	Subordinate clause ratio

Table 6.
Summary of Comprehensibility Scores of 50 Japanese Learners

Comprehensibility	*M*	*SD*	*Range*
Canadian raters	4.7	1.5	1.3-7.7
Singaporean raters	4.0	1.2	1.7-6.5

Note. 1 = easy to understand, 9 = difficult to understand

Table 7.
Summary of Linguistic 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 8
Partial Correlations Between the Pronunciation and Lexicogrammar Variables and Comprehensibility Judged by Canadian and Singaporean Raters

Linguistic variables	Comprehensibility (Canadian raters)	Comprehensibility (Singaporean raters)
Segmental errors[a]	.72*	.62*
Word stress[a]	.65*	.54*
Intonation[a]	.37	.31
Speech rate[a]	.60*	.58*
Lexical appropriateness[b]	.36	.52*
Lexical richness[b]	.02	.04
Grammatical accuracy[b]	.47*	.47*
Grammatical complexity[b]	.06	.17

Note. *$\alpha < .01$ (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 9
Results of Multiple Regression Analyses Using the Linguistic Factors as Predictors of Comprehensibility

Predicted variable	Predictor variables	*Adjusted* R^2	R^2 *change*	*F*	*p*
Comprehensibility (Canadian raters)	Segmentals	.67	.67	98.530	$p < .0001$
	Speech rate	.78	.12	88.048	$p < .0001$
	Grammatical accuracy	.83	.05	81.954	$p < .0001$
Comprehensibility (Singaporean raters)	Segmentals	.57	.58	66.819	$p < .0001$
	Lexical appropriateness	.73	.16	67.707	$p < .0001$
	Speech rate	.79	.06	63.026	$p < .0001$

Note. The variables entered into the regression equations included segmentals, speech rate, lexical appropriateness and grammatical accuracy.