Applied Psycholinguistics, page 1 of 24, 2015
doi:10.1017/S0142716414000502

Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels

KAZUYA SAITO
Birkbeck, University of London

PAVEL TROFIMOVICH
Concordia University

TALIA ISAACS
University of Bristol

Received: February 8, 2014 Accepted for publication: June 6, 2014

ADDRESS FOR CORRESPONDENCE
Kazuya Saito, Department of Applied Linguistics and Communication, Birkbeck, University of London, 26 Russell Square, London WC1B 5DQ, UK. E-mail: kazuyasaitou@hotmail.com

ABSTRACT
The current project aimed to investigate the potentially different linguistic correlates of comprehensibility (i.e., ease of understanding) and accentedness (i.e., linguistic nativelikeness) in adult second language (L2) learners' extemporaneous speech production. Timed picture descriptions from 120 beginner, intermediate, and advanced Japanese learners of English were analyzed using native speaker global judgments based on learners' comprehensibility and accentedness, and then submitted to segmental, prosodic, temporal, lexical, and grammatical analyses. Results showed that comprehensibility was related to all linguistic domains, and accentedness was strongly tied with pronunciation (specifically segmentals) rather than lexical and grammatical domains. In particular, linguistic correlates of L2 comprehensibility and accentedness were found to vary by learners' proficiency levels. In terms of comprehensibility, optimal rate of speech, appropriate and rich vocabulary use, and adequate and varied prosody were important for beginner to intermediate levels, whereas segmental accuracy, good prosody, and correct grammar featured strongly for intermediate to advanced levels. For accentedness, grammatical complexity was a feature of intermediate to high-level performance, whereas segmental and prosodic variables were essential to accentedness across all levels. These findings suggest that syllabi tailored to learners' proficiency level (beginner, intermediate, or advanced) and learning goal (comprehensibility or nativelike accent) would be advantageous for the teaching of L2 speaking.

As many second language (L2) researchers have pointed out, it is crucial to set realistic goals for adult L2 learners, prioritizing understanding over

nativelikeness, in order for learners to be able to communicate successfully in academic and business settings (e.g., Derwing & Munro, 2009; Levis, 2005). Consistent with this agenda, recent research has begun to focus on two listener-derived constructs, namely, comprehensibility (ease of understanding) and accentedness (sounding nativelike), examining how different aspects of language (e.g., phonological, lexical, grammatical, and discourse-level factors) contribute to these constructs (e.g., Kang, Rubin, & Pickering, 2010; Munro & Derwing, 1999; Trofimovich & Isaacs, 2012). Building on this work, the current study aimed to examine linguistic correlates of comprehensibility and accentedness for L2 learners at different ability levels (beginner, intermediate, and advanced). Our overall objective was to clarify the relationship between comprehensibility and accentedness at different levels of L2 oral ability and to identify possible pedagogical implications for learners at different levels, and for their teachers, wishing to pursue comprehensible, but not necessarily unaccented, speech as a learning goal.

BACKGROUND

As languages such as English, Chinese, Arabic, or Spanish become vehicles of international communication, particularly among nonnative speakers, developing adequate L2 oral proficiency is important for many nonnative speakers, especially for achieving their career- and academic-related goals. This holds true not only in L2 contexts (e.g., English in North America) but also in foreign language settings (e.g., English in Asia or most parts of Europe). To assess and promote the development of L2 communicative abilities, much attention has been directed toward establishing performance benchmarks for a given level of learner ability, which typically include tasks that beginner, intermediate, and advanced learners are expected to handle (e.g., the Common European Framework of Reference for Languages and Canadian Language Benchmarks).

However, in terms of learners' L2 pronunciation, which refers here to dimensions associated with linguistic attributes of spoken language (e.g., prosody and segmental accuracy), the linguistic ability of native speakers has long been viewed by teachers and students in many contexts as the *ideal* ultimate learning goal (e.g., Derwing, 2003; Tokumoto & Shibata, 2011). Yet previous research has convincingly shown that few adult learners can attain nativelike L2 pronunciation, even if they begin learning at an early age, and that accent is a common characteristic of L2 speech (e.g., Flege, Munro, & MacKay, 1995). Consequently, what appears to be crucial for L2 pronunciation learning is setting realistic goals in regard to what learners should aim for (Derwing & Munro, 2009; Levis, 2005), with two possible goals being comprehensibility (a broad measure of a speaker's communicative effectiveness, referring to how easily listeners can understand L2 speech) and accent reduction (based on a broad construct of accentedness, encompassing listeners' judgments about how nativelike L2 speech sounds). Briefly, comprehensibility and accentedness are overlapping yet independent constructs, as illustrated by the fact that even some heavily accented L2 speech can be highly

comprehensible (e.g., Derwing & Munro, 2009; Kang et al., 2010; Munro & Derwing, 1999).

From a theoretical perspective, comprehensibility (rather than accentedness) is relevant to L2 development. The interaction hypothesis (e.g., Long, 1996), for instance, posits that language learning takes place precisely when comprehensibility is compromised during conversational interaction involving L2 speakers. When interlocutors encounter communication breakdowns attributable to language, interlocutors often make intuitive or conscious efforts to repair the impaired linguistic detail, relying on clarification requests and comprehension and confirmation checks to facilitate understanding. This conversational behavior, which is termed negotiation for meaning, is hypothesized to be facilitative of adult L2 development (Mackey & Goo, 2007). Given that certain linguistic features in L2 speech might affect comprehensibility and thus trigger negotiation for meaning more than others (e.g., Mackey, Gass, & McDonough, 2000), learners would likely selectively attend to and practice those linguistic domains tied to comprehensibility rather than those that are uniquely linked to accentedness. Therefore, identifying and teaching linguistic features linked to understanding might help learners make the most of the acquisitional value of input and interaction with an interlocutor by helping them notice and repair their nontarget productions.

From a practical, applied perspective, more research is needed to reveal precisely what differentiates accent from comprehensibility, in order to help practitioners decide which pedagogical focus (accent reduction vs. comprehensibility development) they should target through instruction in keeping with students' motivations and ultimate communicative needs. Of course, it may not be possible for teachers to discourage students with a strong desire to sound nativelike from pursuing accent reduction. However, at the same time, it is important to let both teachers and students know that attaining nativelike L2 pronunciation is rare, and that successful L2 communication in various social, academic, and business settings requires comprehensible but not necessarily unaccented, nativelike L2 speech (Derwing & Munro, 2009). Therefore, investigating linguistic correlates of comprehensibility and accentedness emerges as an important research goal in helping teachers select instructional targets consistent with learner needs.

To date, several studies have examined linguistic features in L2 production, targeting several dimensions of speech (e.g., pronunciation, lexicon, and grammar). For example, native-speaking listeners tend to extract meaning from L2 speech, drawing on segmental, suprasegmental, and fluency (temporal) detail, such as word stress (Field, 2005), sentence stress (Hahn, 2004), speech rate (Munro & Derwing, 2001), and pitch range, stress, and pause or syllable length (Kang et al., 2010). Corpus studies have also determined the lexical composition of various genres of L2 oral discourse, such as daily conversations (Adolphs & Schmitt, 2003), TV programs (Webb & Rodgers, 2009a) and movies (Webb & Rodgers, 2009b), which might be required for successful comprehension of these genres. With respect to grammar, it has been shown that the nature of grammar in L2 speech depends, among other factors, on the nature of a speaking task (e.g., Foster & Skehan, 1996), the degree to which a speaking task is structured or scaffolded with

supporting information (e.g., Tavakoli & Foster, 2008), and the presence of planning or preparation opportunities available to L2 speakers (e.g., Yuan & Ellis, 2003).

Although this research is overall revealing of the linguistic complexity of L2 production, it does not indicate how multiple linguistic aspects of speech relate to understanding. For instance, it is as yet unclear how phonological, lexical, and grammatical composition of L2 speech in structured monologic speaking tasks, or in unstructured interactive tasks, is linked to comprehensibility and how these linguistic dimensions are tied to accentedness. What emerges as an important research objective, then, is the need to investigate directly how multiple linguistic elements in learner speech *together* determine comprehensibility and how this joint contribution of various linguistic elements differs for accentedness. In the precursor project directly motivating the current research, we first had native-speaking listeners rate the comprehensibility and accentedness of L2 picture narratives produced by 40 French speakers of English. They then assessed the segmental, prosodic, temporal, lexical, grammatical, and discourse-level characteristics of these same narratives using perceptual judgments by experienced listeners (Saito, Trofimovich, & Isaacs, 2014) as well as linguistic coding for 19 categories (e.g., proportion of segmental errors, hesitations/self-corrections, and grammar errors) (Isaacs & Trofimovich, 2012; Trofimovich & Isaacs, 2012). The results showed that listener judgments closely matched the linguistic coding of speech, and that accentedness was strongly linked with phonological aspects of L2 speech (segmental accuracy, in particular), while comprehensibility was associated with variables spanning the dimensions of phonology, lexis, grammar, and discourse structure.

The first noteworthy finding of our precursor study was that listeners with linguistic and pedagogical experience (graduate students in applied linguistics) could accurately and reliably use a 1,000-point continuous sliding scale with clearly identified endpoints to rate several variables spanning the domains of phonology (vowel/consonant accuracy, word stress, intonation, rhythm, and speech rate), lexicon (appropriateness and richness), grammar (accuracy and complexity), and discourse structure (story richness). This result reveals a significant relationship between rater intuition about pronunciation, vocabulary, grammar, and discourse characteristics of L2 speech and the corresponding linguistic properties of speech. This finding is consistent with previous reports showing that linguistically trained and naïve raters alike can use simple 7- or 9-point rating scales to reliably judge the quality of vowels and consonants in L2 speech (Andersson & Engstrand, 1989; 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). Scalar ratings of L2 speech are rare in L2 vocabulary and grammar studies (but see Crossley, Salsbury, & McNamara, 2014; Storch, 2005), where L2 speech is typically examined through lexical profiling and linguistic coding (e.g., Foster, Tonkyn, & Wigglesworth, 2000; Lu, 2012), using such variables as accuracy (e.g., number of error-free clauses), variation (e.g., type frequency), sophistication (e.g., ratio of frequent and infrequent words), and complexity (e.g., ratio of independent and dependent clauses). Thus, as shown

by Saito et al. (2014), rating scales targeting various characteristics of L2 speech represent a reliable and easy to use method of evaluating L2 speech by listeners.

The second relevant finding of our precursor study was that comprehensibility and accentedness were associated with different linguistic dimensions of speech. While comprehensibility was linked to several domains (pronunciation, lexicon, grammar, and discourse structure), accent was associated primarily with segmental and suprasegmental pronunciation detail. This result is compatible with prior research, showing that listener understanding is linked to aspects of pronunciation (Derwing et al., 2004; Kang et al., 2010; Munro & Derwing, 2006; Tajima, Port, & Dalby, 1997), grammar and lexicon (Fayer & Krasinski, 1987; Varonis & Gass, 1982) as well as discourse structure (Isaacs & Trofimovich, 2012) in L2 speech. In contrast, listener judgment of accentedness is mostly based on segmental, suprasegmental, and fluency characteristics of L2 speech, such as vowel and consonant accuracy, syllable duration, stress, and pitch range (Anderson-Hsieh, Johnson, & Koehler, 1992; Winters & O'Brien, 2013). What is unclear, though, is how various linguistic dimensions of speech relate to comprehensibility and accentedness at different levels of learners' L2 oral ability. This is because previous studies that focused on both comprehensibility and accentedness within a single report included a restricted sample of learners in terms of participant numbers and proficiency levels (e.g., Munro & Derwing, 1999; Saito et al., 2014), and studies that included large groups of learners examined only a few linguistic dimensions at a time or targeted only comprehensibility or accentedness (e.g., Anderson-Hsieh et al., 1992; Kang et al., 2010).

The current project therefore investigated this issue in an exploratory study targeting two research objectives. Our first objective was to replicate and test the generalizability of the relationship between comprehensibility and accentedness, as shown by the precursor research (Isaacs & Trofimovich, 2012; Saito et al., 2014; Trofimovich & Isaacs, 2012), for a large sample of L2 learners from another language background, which included 120 adult Japanese speakers of English in Canada with a wide range of L2 oral ability (see below). A large sample of speakers varying in L2 ability allowed us to address our second objective, namely, to investigate linguistic correlates of comprehensibility versus accentedness at different levels of L2 oral ability. Because no previous research has focused on comprehensibility and accentedness at different levels of L2 speaking ability and because the study was conceptualized as exploratory, no specific predictions or hypotheses were proposed.

To address both objectives, we asked inexperienced native-speaking raters to judge comprehensibility and accentedness in short narratives spoken by the 120 learners and then recruited experienced native-speaking raters to evaluate the same narratives for eight linguistic variables spanning the domains of pronunciation, fluency, lexis, and grammar. In sum, we wished to advance our understanding of the comprehensibility and accentedness constructs, by examining how multiple linguistic aspects of speech relate to these constructs at different levels of L2 ability. We also sought to develop pedagogical implications for learners and teachers wishing to target comprehensibility or accent reduction (nativelikeness) as a learning goal at different levels of their L2 oral proficiency development.

Table 1. *Length of residence and age of arrival profiles for 120 Japanese speakers (frequency counts)*

Length of Residence	*n*	Age of Arrival	*n*
<1 year	26	16–20 years	11
1–5 years	14	21–25 years	44
6–10 years	19	26–30 years	39
11–20 years	34	31–35 years	18
21–41 years	27	36–40 years	8
Total	120	Total	120

METHOD

Participants

Speakers. The participants were 120 adult Japanese speakers of English (M_{age} = 40.3 years, range = 20–70; 17 males, 103 females) from the Canadian cities of Montreal (n = 43) and Vancouver (n = 77). As summarized in Table 1, the speakers represented a wide range of age of arrival in Canada (AOA) and length of residence (LOR) profiles, with a mean AOA of 26.6 years (range = 18–40) and a mean LOR of 12.4 years (range = 0.01–41.0). A broad range of AOA and LOR was important because adult L2 speakers are believed to attain greater pronunciation proficiency with an earlier timing of first exposure to the L2 (Flege et al., 1995) and with an increasing amount of experience (usually operationalized as LOR in the target country), especially when they use their L2 on a daily basis (Flege & Liu, 2001) and demonstrate high willingness to communicate (Derwing & Munro, 2013). All Japanese participants expressed a high level of motivation toward improving their L2 oral ability to successfully achieve various tasks by virtue of the fact that they were studying or working in English-speaking environments where they regularly interacted with native and nonnative speakers of English in a predominantly English-medium context. For a native speaker baseline, 10 native English undergraduate students (M_{age} = 25.1 years) were recruited from an English-speaking university in Montreal (5 males, 5 females) to complete the three oral tasks (see below). The baseline data served as a native speaker benchmark for raters to use in evaluating Japanese speakers.

Inexperienced raters. To judge the comprehensibility and accentedness of the extemporaneous speech samples produced by 120 Japanese and 10 native English speakers, 5 native English undergraduate students (M_{age} = 27.6 years; 2 males, 3 females) were recruited as inexperienced raters from an English-speaking university in Vancouver, Canada. Following a common definition of inexperienced raters (e.g., Isaacs & Thomson, 2013) and previous research on comprehensibility and accentedness (e.g., Derwing & Munro, 2009), the raters had no linguistic and pedagogical training. Using a 6-point scale (1 = *not at all*, 6 = *very much*), the

raters judged their familiarity with Japanese-accented English at a mean of 1.3 (range = 1–2) and reported minimal contact with Japanese speakers of English.

Experienced raters. To conduct linguistic analyses of phonological, lexical, and grammatical characteristics of the recorded speech samples, 5 native English speakers (M_{age} = 29.4 years; 2 males, 3 females) were recruited as raters from the pool of graduate students in applied linguistics at an English-speaking university in Montreal. The raters had between 1 and 10 years of teaching experience in various settings (M = 4.0 years) and had all taken a graduate-level semester-long course on applied phonetics and pronunciation teaching. Using the same scale, these raters judged their familiarity with Japanese-accented English at a mean of 3.4 (range = 1–5).

Procedure

Speaking task. Following previous L2 pronunciation studies (e.g., Derwing & Munro, 2009; Hopp & Schmid, 2013; Munro & Mann, 2005), extemporaneous speech was elicited via a timed picture description task. Given the demanding nature of this task (Derwing et al., 2004), especially for beginner-level speakers (e.g., LOR < 1 year), the task was modified as follows: (a) instead of using a series of thematically linked images, speakers described seven separate pictures, with three keywords printed as hints; (b) to control for speakers' lack of familiarity with the task, the first four pictures were used for practice and the last three were targeted for analyses; and (c) to minimize the amount of conscious speech monitoring (see Ellis, 2005), speakers were given a very small amount of planning time (i.e., only 5 s) before describing each picture. These measures helped ensure that all speakers, regardless of their L2 oral ability levels, could successfully complete the task, providing sufficient spontaneous speech data without excessive hesitations and dysfluencies.

The three target pictures (Pictures A, B, and C) depicted a table left out in a driveway in heavy rain (keywords: rain, table, and driveway), three men playing rock music with one singing a song and the other two playing guitars (keywords: three guys, guitar, and rock music), and a long stretch of road under a cloudy blue sky (keywords: blue sky, road, and cloud). The keywords were carefully chosen to elicit problematic segmental and syllable structure features for Japanese speakers of English (Saito, 2014), on the assumption that the speakers would reveal their pronunciation ability through the use of these difficult features in speech. For instance, Japanese speakers have been reported to neutralize the English /r/–/l/ contrast ("rain, rock, brew, and crowd" vs. "lane, lock, blue, and cloud") and to insert epenthetic vowels between consecutive consonants (/dəraɪvə/ for "drive," /θəri/ for "three," and /səkaɪ/ for "sky") and after word-final consonants (/teɪbələ/ for "table" and /myuzɪkə/ for "music") in borrowed words (i.e., Katakana).

All speech recording was carried out individually in quiet rooms in university labs, community centers, or participants' homes in Montreal or Vancouver, using a digital Roland-05 audio recorder (44.1-kHz sampling rate with 16-bit quantization). The project was advertised on regional community websites and in local newspapers with the goal of investigating general L2 speaking skills of

Japanese immigrants to Canada. All instructions were delivered in Japanese by the researcher (a native speaker of Japanese) to ensure that all speakers understood the procedures. To minimize possible "language mode" effects from using Japanese, the first four pictures described by the speakers were treated as practice to allow the speakers to become comfortable using English as part of the task. The remaining three pictures (A, B, C, in that order) described by the speakers were used for the main analysis. In total, the speakers generated 390 picture descriptions (3 pictures by 120 Japanese and 10 English speakers). On average, about 5–10 s from the beginning of each description was extracted for each speaker, for a total mean length of 25 s for the three picture descriptions combined (range = 14.5–32.4 s). The total duration of these samples was deemed sufficient, compared to 15–30 s samples used for rating in similar pronunciation studies (e.g., Derwing & Munro, 1997), to elicit listeners' impressionistic ratings of speech.

Speech rating. The experimental procedure consisted of two sets of analyses. The target speech materials, which were elicited from 120 Japanese speakers of English, were first rated by five inexperienced raters for comprehensibility and accentedness. The same audio recordings were then evaluated by five linguistically trained (experienced) raters for eight linguistic measures spanning the domains of phonology, lexis, and grammar.

Comprehensibility and accentedness rating. For comprehensibility and accentedness rating, the 390 picture descriptions produced by the 120 Japanese and 10 English speakers were arranged in separate blocks, organized by picture, with 130 audio samples in each block. To reduce fatigue, the raters assessed each block on separate days in individual rating sessions, which all together lasted about 3 hr, with the order of blocks counterbalanced across raters (e.g., ABC, BCA, and ACB). In each listening session, the samples were presented using Praat speech editing software (Boersma & Weenink, 2012). After familiarizing themselves with each picture prompt, the listeners randomly heard each audio sample once before making a scalar judgment for comprehensibility and accentedness, in that order. Based on prior research, comprehensibility was defined as the degree of ease or difficulty in raters' understanding of L2 speech (Derwing & Munro, 2009). Accentedness was defined as raters' perception of the degree to which L2 speech is influenced by his/her native language and/or colored by other nonnative features (Trofimovich & Isaacs, 2012). Both constructs were rated using separate 9-point scales (1 = *very easy to understand, no accent*; 9 = *very hard to understand, heavily accented*). Before proceeding to the 130 target samples, the raters assessed five preliminary files for practice. They were told that the data set represented a range of ability levels, from nativelike speakers to complete beginners, and were asked to use the entire scale.

Phonological, lexical, and grammatical analysis. The 130 target audio samples were also evaluated by linguistically trained raters for eight audio- and transcript-based measures developed and validated in a previous project (Saito et al., 2014). These sessions took place on three different days, with the first 2 days devoted to

audio-based judgments (about 2 hr) and the last day spent evaluating transcripts (about 1 hr).

Audio-based measures. Three picture descriptions (Pictures A, B, and C) for each speaker were combined and stored in a single audio file, in order to provide the raters with sufficient content in duration to make judgments. The raters listened to and evaluated each sample using four segmental, prosodic, and temporal categories: (a) segmental errors (substitution, omission, or insertion of individual consonants or vowels), (b) word stress errors (misplaced or missing primary stress), (c) intonation (appropriate, varied versus incorrect and monotonous use of pitch), and (d) speech rate (speed of utterance delivery). During the first session, the raters received a thorough explanation of the four rated categories (see Appendix A) and the rating procedure and then evaluated 3 practice samples not included in subsequent analyses. For each practice sample, they were asked why they made their decisions and then received feedback to ensure that the rated categories were understood and applied appropriately. The raters then proceeded to rate a selection of 50 samples, presented to each rater in a unique random order. In the second session, the raters reviewed the four categories and then followed the same procedure to rate the remaining 80 samples, again presented in a unique random order.

The rating was carried out using the MATLAB software, and the raters used a free-moving slider on a computer screen to assess each of the four categories. If the slider was placed at the leftmost (negative) end of the continuum, labeled with a frowning face, the rating was recorded as 0; if it was placed at the rightmost (positive) end of the continuum, labeled with a smiley face, it was recorded as 1000. The slider was initially placed in the middle of each scale, and the raters were told that even a small movement of the slider may represent a fairly large difference in the rating. Except for the frowning and smiley faces and accompanying brief verbal descriptions for the endpoints of each category, the scale included no numerical labels or marked intervals (for onscreen labels, see Appendix A). A 1000-point sliding scale thus allowed raters to make fine-grained judgments for each linguistic category without being tied to discrete-point labels typical of Likert scales. To ensure the quality of the raters' analysis, they also had the option to listen to the same speech sample again until they felt satisfied with their judgment.

Transcript-based measures. To remove pronunciation and fluency as possible confounds in raters' judgments of lexis and grammar, the raters were presented with written transcripts of the audio samples in the final rating session, consistent with the procedure used earlier by Crossley et al. (2014). Following verification of the orthographically transcribed audio samples, the transcripts were cleaned by removing spelling clues signaling pronunciation-specific errors (e.g., *lock music* was transcribed as "rock music"), obvious mispronunciations based on contextual information available in the pictures (e.g., *ought side* was transcribed as "outside" and *lawn Lee* was transcribed as "lonely"), and orthographic markings of pausing (e.g., *uh*, *um*, *oh*, and *ehh*). The raters assessed the lexical and grammatical aspects of the transcripts using the following four categories: (a) lexical appropriateness (accuracy of vocabulary), (b) lexical richness (varied and sophisticated use of

vocabulary), (c) grammatical accuracy (errors in word order, grammar endings, and agreement), and (d) grammatical complexity (amount of subordination). At the beginning of the session, the raters first received an explanation of the four categories (see Appendix A) and practiced the procedure by rating 3 additional written samples. During practice, the raters were asked to explain their decisions and received feedback to ensure their full understanding of the categories. Subsequently, the raters evaluated all 130 written transcripts presented via the MATLAB software in a unique random order. The three transcripts for Picture A, B, and C descriptions were displayed on screen all at once, always in the same order, and the raters assessed their lexical and grammatical content with similar free-moving sliders (see Appendix A).

Posttask questionnaire. After completing the audio- and transcript-based sessions, the raters used 9-point scales to assess the extent to which (a) they understood the rated categories (1 = *I did not understand at all*, 9 = *I understand this concept well*) and (b) they could comfortably and easily use them (1 = *very difficult*, 9 = *very easy and comfortable*). For all categories, the raters demonstrated high levels of understanding for all of the linguistic categories ($M = 8.7$), ranging from a mean of 7.8 (grammatical complexity) to 9.0 (segmentals/speech rate), and rated them as easy to use ($M = 8.2$), ranging from a mean of 7.7 (grammatical complexity) to 9.0 (intonation). Thus, the raters appeared confident in their ability to assess the phonological, lexical, and grammatical dimensions of L2 speech.

RESULTS

Rater consistency

The five inexperienced raters were overall consistent in their rating of the 390 speech samples, demonstrating high reliability indexes (Cronbach alpha) for both comprehensibility ($\alpha = 0.95$) and accentedness ($\alpha = 0.98$). Therefore, mean comprehensibility and accentedness scores were computed for each speaker by averaging across all listeners' ratings, with resulting comprehensibility and accentedness scores correlated at $r\ (118) = .89$ ($p < .0001$). Because linguistic judgments by the experienced raters involved the use of categories that were presumably more complex and less intuitive than comprehensibility and accentedness, the five raters' scores showed less agreement. The reliability indexes were nevertheless acceptable, exceeding the benchmark value of 0.70–0.80 (Larson-Hall, 2010) for pronunciation ($\alpha_{segmentals} = 0.91$; $\alpha_{word\ stress} = 0.88$; $\alpha_{intonation} = 0.84$; $\alpha_{speech\ rate} = 0.89$), vocabulary ($\alpha_{appropriateness} = 0.85$; $\alpha_{richness} = 0.86$), and grammar ($\alpha_{accuracy} = 0.83$; $\alpha_{complexity} = 0.79$). The raters' scores were therefore considered sufficiently consistent and were averaged across the five raters to derive a single score per rated category for each speaker.

Linguistic correlates of comprehensibility and accentedness

Our first objective was to determine how 120 Japanese speakers' performance across several linguistic domains related to their comprehensibility and

Table 2. *Summary of a three-factor solution based on a principal component analysis of the eight rated linguistic variables*

Factor 1 (pronunciation)	Segmental errors (0.84), word stress (0.87), intonation (0.85), speech rate (0.73)
Factor 2 (lexicogrammar sophistication)	Lexical richness (0.87), grammatical complexity (0.85)
Factor 3 (lexicogrammar accuracy)	Lexical appropriateness (0.87), grammatical accuracy (0.84)

Note: All eigenvalues are >0.7.

Table 3. *Results of multiple regression analyses using the factors of pronunciation and lexicogrammar as predictors of comprehensibility and accentedness*

Predicted Variable	Predictor Variables	Adj. R^2	ΔR^2	F	p
Comprehensibility	Pronunciation	.49	.49	110.96	<.0001
	Lexicogrammar accuracy	.71	.22	143.26	<.0001
	Lexicogrammar sophistication	.79	.08	148.93	<.0001
Accentedness	Pronunciation	.60	.60	134.82	<.0001
	Lexicogrammar accuracy	.71	.11	143.79	<.0001
	Lexicogrammar sophistication	.77	.06	178.86	<.0001

Note: The variables entered into the regression equation were the three factors obtained in the principal component analysis reported in Table 2.

accentedness ratings. The linguistic scores for all speakers were first submitted to a principal component analysis with Varimax rotation and Kaiser criterion eigenvalue set at 0.70 (Stevens, 2002), to examine whether the eight rated categories showed any underlying patterns based on their clustering. As summarized in Table 2, the principal component analysis revealed three factors accounting for 87.03% of the total variance. Factor 1, "Pronunciation," consisted of all pronunciation scores. Factor 2, "Lexicogrammar Sophistication," included lexical richness and grammatical complexity. Factor 3, "Lexicogrammar Accuracy," comprised lexical appropriateness and grammatical accuracy.

The resulting three factors were then used as predictor variables in two separate stepwise multiple regression analyses to examine their contribution to comprehensibility and accentedness as criterion variables. Although the two regression models accounted for roughly the same amount of total variance (79% for comprehensibility and 77% for accentedness), the ratio of variance explained by the three factors differed (see Table 3). The pronunciation factor alone accounted for most variance in accentedness (60%), whereas both pronunciation (49%) and lexicogrammar (30%) factors contributed sizably to comprehensibility.

The next analyses focused on the pronunciation and lexicogrammar domains separately, targeting their possible influences on comprehensibility and accentedness. For pronunciation, partial correlation analyses were computed to examine

Table 4. *Partial correlations between the pronunciation and lexicogrammar variables and comprehensibility and accentedness*

Pronunciation Variable	Comprehensibility	Accentedness
Segmental errors[a]	.73*	.81*
Word stress[a]	.64*	.70*
Intonation[a]	.52*	.59*
Speech rate[a]	.58*	.50*
Lexical appropriateness[b]	.31*	.06
Lexical richness[b]	.01	.03
Grammatical accuracy[b]	.51*	.17
Grammatical complexity[b]	.15	.11

[a]Variables partialed out from each correlation include lexical appropriateness and richness, as well as grammatical accuracy and complexity.
[b]Variables partialed out from each correlation include vowel/consonant errors, word stress, intonation, and speech rate.
*$\alpha < 0.01$ (Bonferroni corrected).

links between segmental, prosodic, and temporal characteristics of L2 speech and comprehensibility and accentedness, while controlling for lexicogrammar. As shown in Table 4, all pronunciation categories were significantly correlated with comprehensibility and accentedness. Fisher r to z transformations (Bonferroni adjusted), conducted to explore statistical differences in correlation coefficient strength, revealed that none of the pronunciation categories differed in the strength of their association with comprehensibility, but that accentedness was more strongly associated with segmentals than with intonation ($p = .0008$) and speech rate ($p < .0001$). For lexicogrammar, similar partial correlation analyses examined associations of lexical and grammatical categories with comprehensibility and accentedness, while controlling for pronunciation. As illustrated in Table 4, lexical appropriateness and grammatical accuracy were linked with comprehensibility, but none of the lexicogrammar categories were significantly associated with accentedness. According to Fisher r to z transformations, comprehensibility showed a stronger association with grammatical accuracy than with lexical appropriateness ($p = .002$).

Comprehensibility at different ability levels

Our second objective was to focus on how phonological, lexical, and grammatical characteristics of L2 speech relate to beginner, intermediate, and advanced levels of L2 comprehensibility and accentedness. To address this objective, for comprehensibility, 120 Japanese speakers were divided into four equal L2-speaking proficiency groups with nonoverlapping ranges of comprehensibility ratings (shown in Table 5). The speakers' scores for the four pronunciation categories were then

Table 5. *Means and standard deviations for speaker groups based on rank-ordered comprehensibility ratings*

	M	*SD*	Range
Low beginner ($n = 30$)	6.03	0.61	5.33–7.33
High beginner ($n = 30$)	4.80	0.20	4.53–5.20
Intermediate ($n = 30$)	4.06	0.27	3.60–4.47
Advanced ($n = 30$)	2.79	0.69	1.40–3.53
Native baseline ($n = 10$)	1.04	0.06	1.00–1.13

Note: Comprehensibility ranges from 1 = *very easy to understand* to 9 = *very hard to understand*.

submitted to a between-group comparison using a two-way analysis of variance (ANOVA), with repeated measurements on the pronunciation category. This analysis revealed a significant effect of group, F (3, 116) = 67.52, $p < .001$, and category, F (3, 348) = 72.60, $p < .001$, as well as a significant Group × Category interaction, F (9, 348) = 6.25, $p < .001$. Tests of interaction effects (Bonferroni adjusted) further showed that (a) word stress and intonation significantly distinguished the four comprehensibility groups from each other ($p < .001$), with medium to large effect sizes (Cohen $d = 0.78$–1.04); (b) segmentals distinguished between low and high beginner groups ($p < .001$), with a medium effect size ($d = 0.68$), and between intermediate and advanced groups ($p < .001$), with a large effect size ($d = 1.77$); and (c) speech rate significantly distinguished between low and high beginner groups ($p < .001$), with a large effect size ($d = 1.86$).

A similar two-way ANOVA comparing the four lexicogrammar scores for the four comprehensibility groups yielded a significant effect of group, F (3, 116) = 35.47, $p < .001$, and category, F (3, 348) = 260.17, $p < .001$, as well as a significant Group × Category interaction, F (9, 348) = 2.99, $p = .002$. Tests of interaction effects (Bonferroni adjusted) showed that (a) lexical appropriateness distinguished between low and high beginner groups ($p = .001$), with a large effect size ($d = 0.95$); (b) grammar accuracy distinguished between low and high beginner groups ($p < .001$) and between intermediate and advanced groups ($p = .003$), with medium to large effect sizes ($d = 0.79$ and 0.90); and (c) both lexical richness ($p = .048$) and grammatical complexity ($p = .012$) distinguished between high beginner and intermediate groups, with medium effect sizes ($d = 0.74$ and 0.84). Summary statistics and overall level distinctions for comprehensibility based on these comparisons are shown in Table 6.

Accentedness at different ability levels

The final analyses targeted the relationship between linguistic categories and accentedness for 120 Japanese speakers of different ability levels. As with comprehensibility, the speakers were divided into four groups based on their accentedness ratings, with nonoverlapping distribution of scores (shown in Table 7). The

Table 6. *Means (standard deviations) for rated pronunciation and lexicogrammar categories at four levels of second language comprehensibility and accentedness*

	Pronunciation				Lexicon		Grammar	
Group	Segmentals	Word Stress	Intonation	Speech Rate	Appropriateness	Richness	Accuracy	Complexity
				Comprehensibility				
Low beginner	294 (100)	372 (86)	271 (106)	325 (152)	654 (110)	327 (160)	369 (158)	235 (119)
High beginner	389 (87)	471 (74)	376 (96)	560 (119)	750 (91)	450 (171)	477 (120)	299 (113)
Intermediate	454 (106)	555 (99)	476 (146)	614 (128)	768 (102)	571 (169)	559 (151)	414 (156)
Advanced	652 (116)	669 (107)	586 (149)	709 (123)	829 (58)	616 (184)	691 (142)	472 (164)
				Accentedness				
Low beginner	282 (94)	373 (78)	373 (78)	370 (169)	661 (115)	348 (187)	375 (151)	245 (138)
High beginner	375 (55)	459 (70)	459 (70)	499 (170)	744 (95)	415 (164)	509 (147)	289 (80)
Intermediate	477 (98)	554 (85)	554 (85)	620 (101)	786 (92)	555 (169)	569 (170)	385 (149)
Advanced	655 (113)	681 (106)	609 (139)	719 (117)	810 (82)	645 (154)	639 (169)	500 (163)

Note: Dashed lines separate second language comprehensibility and accentedness levels that are distinguished by a given linguistic category according to Bonferroni-corrected comparison.

Table 7. *Means and standard deviations for speaker groups based on rank-ordered accentedness ratings*

	M	*SD*	Range
Low beginner ($n = 30$)	7.49	0.52	6.87–8.47
High beginner ($n = 30$)	6.51	0.19	6.27–6.80
Intermediate ($n = 30$)	5.60	0.38	4.87–6.20
Advanced ($n = 30$)	3.90	0.84	1.80–4.80
Native baseline ($n = 10$)	1.04	0.08	1.00–1.27

Note: Accentedness ranged from 1 = *no accent* to 9 = *heavily accented.*

speakers' scores for the four pronunciation categories were then submitted to a similar between-group comparison using a two-way ANOVA, which yielded a significant main effect of group, F (3, 116) = 82.21, $p < .001$, and category, F (3, 348) = 65.12, $p < .001$, but no significant two-way interaction, F (9, 348) = 1.62, $p = .11$. According to tests of simple main effects (Bonferroni adjusted), all pronunciation categories significantly distinguished the four comprehensibility groups from each other ($p < .001$), with large effect sizes ($d = 0.84$–1.14). With respect to lexicogrammar, a similar ANOVA revealed a significant main effect of group, F (3, 116) = 28.69, $p < .001$, and category, F (3, 348) = 262.11, $p < .001$, as well as a significant Group × Category interaction, F (9, 348) = 3.30, $p = .001$. Tests of interaction effects (Bonferroni adjusted) showed that (a) both lexical appropriateness ($p = .008$) and grammar accuracy ($p < .01$) distinguished between low and high beginner groups, with medium to large effect sizes ($d =$ 0.78 and 0.89); (b) lexical richness distinguished between high beginner and intermediate groups ($p = .011$), with a small effect size ($d = 0.37$); and (c) grammatical complexity distinguished between high beginner and intermediate ($p = .046$) and between intermediate and advanced groups ($p = .009$), with medium to large effect sizes ($d = 0.78$ and 0.85). Summary statistics and overall level distinctions for accentedness based on these comparisons appear in Table 6.

DISCUSSION

Motivated by prior research on comprehensibility and accentedness (e.g., Derwing & Munro, 2009), the current project aimed to examine contributions of several linguistic factors (i.e., pronunciation, vocabulary, and grammar) to these rated constructs. First and foremost, the results of this study, which targeted 120 Japanese speakers of English, closely replicated the findings of our previous research based on 40 French speakers of English (Isaacs & Trofimovich, 2012; Saito et al., 2014; Trofimovich & Isaacs, 2012). That is, comprehensibility appears to be related to segmental, prosodic, temporal, lexical, and grammatical aspects of L2 speech, while accentedness is mainly associated with pronunciation factors, particularly with segmental accuracy. These differences in listener ratings reveal a complex nature of linguistic influences on listener perception of L2 comprehensibility and

accentedness. When asked to rate comprehensibility, native-speaking listeners seem to give priority to the quality of all available linguistic resources in L2 speech in order to arrive at overall meaning in a timely and efficient way. However, in terms of accentedness, listeners likely prioritize segmental accuracy ahead of prosodic, temporal, lexical, and grammatical characteristics of L2 speech, arguably owing to the saliency of segmental substitutions to the listener and the relative learning difficulty of certain segmental contrasts for the L2 speaker (Munro & Derwing, 2006). Although adult L2 speakers can perform at nativelike levels in terms of L2 vocabulary and grammar (e.g., Birdsong & Molis, 2001), they often fail to master nativelike pronunciation (e.g., Flege, Yeni-Komshian, & Liu, 1999), with such learning difficulties being most pronounced for segmentals (Abrahamsson, 2012) compared to suprasegmentals (Trofimovich & Baker, 2006).

The results presented here provide empirical evidence for the widely accepted view that a speaker who reaches a certain threshold of phonological, lexical, and grammatical ability can be highly comprehensible while still being fairly accented due to segmental inaccuracies (Derwing & Munro, 2009). Listener-based differences in judgments of comprehensibility and accentedness also imply that listeners likely engage in different types of behaviors when rating each construct. Because understanding associative content involves simultaneous processing of all available linguistic information (i.e., form and meaning), comprehensibility judgments tend to be highly resource sensitive. In essence, the more comprehensible L2 speech is, the less effortful it is for listeners to understand what the speaker wants to convey (Munro & Derwing, 1995). Conversely, due to a strong link between accentedness judgments and segmental detail of L2 speech (i.e., more attention to form and less to meaning), accent rating appears to be invariably fast, effortless, and intuitive. For example, Munro, Derwing, and Burgess (2010) demonstrated that listeners can detect foreign accents even within a single word played backward, that is, with minimal linguistic and content information available. Similarly, native-speaking listeners can rapidly adapt to foreign-accented speech when exposed to it, suggesting that the acoustic/phonetic detail that feeds into listener perception of accent can be detected rapidly and then used to aid subsequent speech processing (e.g., Bradlow & Bent, 2008).

The second outcome of this study was a description of linguistic variables characterizing different levels of L2 comprehensibility and accentedness. As was argued in the Introduction, this information is crucial for establishing learning benchmarks and developing instructional materials for adult L2 learners with different learning goals. As summarized in Table 6, the results again indicate that comprehensibility and accentedness consist of distinct linguistic components contributing differently to various levels of each construct. For comprehensibility, word stress and intonation are equally important at all levels (beginner → intermediate → advanced); attaining a minimum level of segmental accuracy, fluency, lexical appropriateness, and grammatical accuracy is relatively important at the initial stage (low beginner → high beginner); and segmental precision and grammatical accuracy characterize the highest skill level (intermediate → advanced). For accentedness, several pronunciation variables (segmentals, word stress, intonation, and speech rate) are equally important at all levels (beginner → intermediate → advanced); a fundamental level of lexicogrammar (lexical appropriateness and

richness, plus grammatical accuracy) is important initially (low beginner → high beginner); and it is mainly grammatical complexity (along with pronunciation variables) that determines nonaccented, nativelike L2 speech at the highest skill level (intermediate → advanced).

The multifaceted relationship between the listener-based constructs of comprehensibility and accentedness and linguistic properties of L2 speech may contribute to a clearer understanding of several current issues in L2 speech research. One such issue, for example, is the question of which linguistic dimensions of pronunciation (described broadly as segmentals vs. suprasegmentals) directly impact on L2 comprehensibility development, which has been a source of debate (e.g., Hahn, 2004). While some researchers have claimed that targeting prosody and fluency (as opposed to individual vowels and consonants) has a stronger impact on comprehensibility (e.g., Derwing, Munro, & Wiebe, 1998), others have argued that learners must attend to crucial segmental features of L2 speech, especially if they wish to communicate successfully with other nonnative speakers (e.g., Jenkins, 2000). The current findings suggest that the relative weight of instructional focus on segmentals versus suprasegmentals, particularly with the view of improved comprehensibility, may vary as a function of learner ability level. While consistent attention should be given to word stress and intonation throughout L2 oral development (Field, 2005; Hahn, 2004), students might need to be encouraged to shift their focus from improving fluency (Derwing et al., 2004) to refining segmental accuracy (Saito, 2013) as their L2 comprehensibility develops.

A third outcome of the current findings pertains to the relationship between accuracy and complexity of L2 oral production. For example, it has been argued that complexity relates to L2 learners' desire to use advanced language, which might in turn exhaust most of the available cognitive resources that would otherwise be used to avoid grammatical errors. As a result, an increase in linguistic complexity tends to co-occur with an increased error rate, revealing a trade-off between complexity and accuracy (e.g., Skehan, 2009). The current findings showed that grammatical accuracy and complexity are dissociated at the advanced levels of comprehensibility and accentedness, such that there was a strong link between comprehensibility and accuracy and between accentedness and complexity. This implies that a complex trade-off between grammar complexity and accuracy (cf. Skehan, 2009; Robinson, 2011) might be associated with different learning goals. Whereas learners aspiring to attain unaccented, nativelike L2 speech may focus on the use of complex language, those wishing to improve their overall comprehensibility may prioritize accuracy over complexity.

The fourth outcome of the current findings is that they can be used to can inform strategic criteria and steps for enhancing adult L2 learners' phonological, lexical, and grammatical performance from the perspective of comprehensibility and accentedness across the ability spectrum. Achieving unaccented, nativelike speech would exclusively require most adult learners to focus on pronunciation (and especially on segmental accuracy). Thus, if learners express an interest in sounding nativelike, despite the inherent difficulty of attaining this goal (e.g., Flege et al., 1995), an instructional focus on accent minimization or reduction should not be rejected. However, what is important is to inform learners that linguistic nativelikeness is rarely attested in adult L2 learners (e.g., Abrahamsson, 2012) and

that an exclusive focus on the segmental detail of speech (with a view of reducing accent) does not appear to be the most efficient choice if the learning goal is the development of L2 comprehensibility (Derwing & Munro, 2009).

Improving comprehensibility would most likely involve an integrative approach targeting crucial pronunciation, vocabulary, and grammar features that affect successful L2 communication. For instance, teaching Japanese learners to achieve beginner-level comprehensibility would include the development of optimal fluency, good prosody, and precise vocabulary use. Thus, it would be effective to provide learners with explicit vocabulary instruction, particularly targeting frequent words in L2 oral discourse (e.g., Schmitt, 2008), while simultaneously helping them pronounce these words with appropriate prosody (Field, 2005; Hahn, 2004) and at an optimal speaking rate (Munro & Derwing, 2001). At the later stages of L2 comprehensibility development, teachers might also wish to encourage learners to produce different types of words (instead of using the same lexical items repetitively) through various kinds of meaning-focused input and output tasks (Schmitt, 2008), while at the same time drawing their attention to segmental and grammatical errors during such tasks via a range of interactive feedback techniques (Saito, 2013).

Conclusion

Two broad conclusions can be drawn from the findings of the current study. First, native-speaking listeners evaluate L2 speech differently when they judge ease of understanding versus linguistic nativelikeness. Comprehensibility captures the extent to which L2 speakers have reached a certain threshold of phonological, lexical, and grammatical ability needed for their conversational partners to successfully understand them. All together these linguistic characteristics of L2 speech determine how much effort and time are required for listeners to extract meaning (see Munro & Derwing, 1995). In contrast, accentedness can be used as an index of listeners' effortless, intuitive, and likely implicit judgments of the extent to which L2 speakers have mastered, in particular, segmental accuracy with respect to production. Second, linguistic correlates of comprehensibility and accentedness vary according to speakers' L2 skill. While an emphasis on segmental accuracy and grammatical complexity plays an important role in accentedness (especially at high ability levels), a tailored approach is needed for L2 comprehensibility, with prosody, temporal variables, and lexical accuracy ideally targeted for beginner to intermediate learners, and segments, prosody, and grammatical accuracy for intermediate to advanced learners.

While these findings offer insights into the relationship among linguistic properties of L2 speech, listener judgments, and learner ability levels, they also bring to light several methodological limitations. First, it needs to be acknowledged that the current data set may not have sufficiently captured the speakers' vocabulary and grammar ability due to the limited nature of the task (timed picture description with three key words provided) and sample length (about 30 s per speaker). Thus, longer speech samples may be needed (cf. 3 min in Lu, 2012; and 5 min in Foster & Skehan, 1996; and Yuan & Ellis, 2003) in order to obtain a more refined picture of lexical, grammatical, and temporal correlates of

comprehensibility and accentedness. Second, it is crucial to test the generalizability of the current findings to other populations of learners and other contexts, especially with respect to different task conditions, including monologue, interview, and two-way interaction tasks (e.g., Derwing et al., 2004) and various kinds of raters, such as native versus nonnative listeners (e.g., Munro, Derwing, & Morton, 2006). Third, pedagogical suggestions for improving L2 comprehensibility and accentedness in the current study must be tested in future classroom-based quasi-experimental research, ideally with both a speech perception and a production component. The ultimate outcome of this research will be a tailored syllabus targeting segmental, prosodic, temporal, lexical, and grammatical aspects of L2 speech, with the view of helping learners become primarily comprehensible but also more nativelike L2 users.

APPENDIX A

Training materials and onscreen labels for pronunciation and lexicogrammar judgment

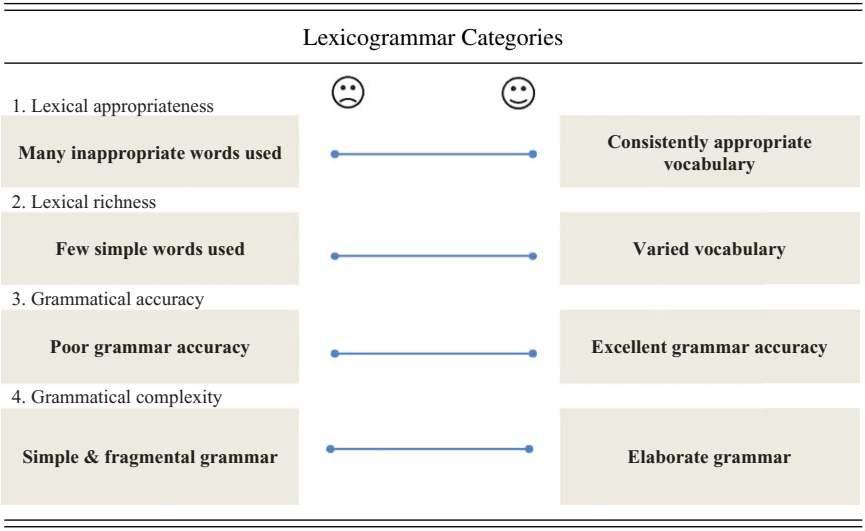
Pronunciation Categories	
Segmental errors	This refers to errors in individual sounds. For example, perhaps somebody says "*road" "rain"* but you hear an "l" sound instead of an "r" sound. This would be a consonant error. If you hear someone say "*fan*" "*boat*" but you hear "*fun*" "*bought*," that is a vowel error. You may also hear sounds missing from words or extra sounds added to words. These are also consonant and vowel errors.
Word stress	When an English word has more than one syllable, one of the syllables will be a little bit louder and longer than the others. For example, if you say the word "computer," you may notice that the second syllable has more stress (comPUter). If you hear stress being placed on the wrong syllable, or you hear equal stress on all of the syllables in a word, then there are word stress errors.
Intonation	Intonation can be thought of as the melody of English. It is the natural pitch changes that occur when we speak. For example, you may notice that when you ask a question with a yes/no answer, your pitch goes up at the end of the question. If sxomeone sounds "flat" when they speak, it is likely because their intonation is not following English intonation patterns.
Speech rate	Speech rate is simply how quickly or slowly someone speaks. Speaking very quickly can make speech harder to follow, but speaking too slowly can as well. A good speech rate should sound natural and be comfortable to listen to.

APPENDIX A (*cont.*)

Pronunciation Categories		
1. Vowel and/or consonant errors		
Frequent		**Infrequent or absent**
2. Word stress errors affecting stressed and unstressed syllables		
Frequent		**Infrequent or absent**
3. Intonation (i.e., pitch variation)		
Too varied or not varied enough		**Appropriate across stretches of speech**
4. Speech rate		
Too slow or too fast		**Optimal**

Lexicogrammar Categories	
Lexical appropriateness	This dimension refers to the appropriateness of the vocabulary words used by the speaker. If the speaker uses incorrect or inappropriate words, including words from the speaker's native language, lexical accuracy is low. Conversely, lexical accuracy is high if the speaker has all the lexical items required to accomplish the speaking task and does so using frequently-used and/or precise lexical expressions.
Lexical richness	This dimension also refers to the vocabulary used by the speaker. However, what is important here is how sophisticated this vocabulary is, taking into account the demands of the speaking task. If the speaker uses a few simple, unnuanced words, the speech lacks lexical richness. However, if the speaker's language is characterized by varied and sophisticated uses of English vocabulary, the speech is lexically rich.
Grammatical accuracy	This refers to the number of grammar errors that the speaker makes, including errors in word order and morphological ending.
Grammatical complexity	This dimension is about the complexity and sophistication of the speaker's grammar. If the speaker uses basic, simple, or fragmented structures or sentences, grammatical complexity is low. Grammatical complexity is high if the speaker uses elaborate and sophisticated grammar structures.

APPENDIX A (*cont.*)



ACKNOWLEDGMENTS

This study was partially funded by Grant-in-Aid for Scientific Research 26770202 from the Ministry of Education, Culture, Sports, Science, and Technology in Japan (to K.S.). Support was also provided by a Social Sciences and the Humanities Research Council of Canada grant (to P.T. and T.I.) and an EU Marie Curie Career Integration grant (to T.I.). We are grateful to the reviewers for providing constructive comments on an earlier version of this paper. We gratefully acknowledge George Smith and Ze Shan Yao for help with data collection and all of the volunteer participants for the project.

REFERENCES

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition, 34*, 187–214.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*, 425–438.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529–555.

Andersson, U., & Engstrand, O. (1989). Perceived strength and identity of foreign accent in Swedish. *Phonetica, 46*, 138–154.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language, 44*, 235–249.

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.

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

Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review, 59*, 545–564.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 12*, 1–16.

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. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning, 63*, 163–185.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393–410.

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

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313–326.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399–423.

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527–552.

Flege, J., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125–3134.

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory & Language, 41*, 78–104.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language. *Applied Linguistics, 21*, 354–375.

Foster, R., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299–323.

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

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.

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). Supresegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal, 94*, 554–566.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 369–377.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition: Second language acquisition* (pp. 413–468). New York: Academic Press.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Review, 96*, 190–208.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*, 471–497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–452). Oxford: Oxford University Press.

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

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285–310.

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

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

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

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

Munro, M. J., & Mann, V. A. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics, 26*, 311–341.

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

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning, 63*, 499–529.

Saito, K. (in press). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics, 24*.

Saito, K., Trofimovich, P., & Isaacs, T. (2014). *Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study*. Unpublished manuscript.

Schmitt, N. (2008). State of the art: Instructed second language vocabulary acquisition. *Language Teaching Research, 12*, 329–363.

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

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*, 153–173.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25*, 1–24.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*, 439–473.

Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes, 30*, 392–408.

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

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*, 905–916.

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

Webb, S., & Rodgers, M. P. H. (2009a). The vocabulary demands of television programs. *Language Learning, 59*, 335–366.

Webb, S., & Rodgers, M. P. H. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*, 407–427.

Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*, 486–507.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1–27.