

**UWL REPOSITORY**  
**repository.uwl.ac.uk**

How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition

Saito, Kazuya, Tran, Mai, Suzukida, Yui, Sun, Hui, Magne, Viktoria ORCID logo ORCID:

<https://orcid.org/0000-0003-1305-9119> and Ilkan, Meltem (2019) How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition*, 41 (5). pp. 1133-1149. ISSN 0272-2631

<http://dx.doi.org/10.1017/S0272263119000226>

This is the Accepted Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/5880/>

**Alternative formats:** If you require this document in an alternative format, please contact: [open.research@uwl.ac.uk](mailto:open.research@uwl.ac.uk)

**Copyright:**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy:** If you believe that this document breaches copyright, please contact us at [open.research@uwl.ac.uk](mailto:open.research@uwl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

**Rights Retention Statement:**

# HOW DO L2 LISTENERS PERCEIVE THE COMPREHENSIBILITY OF FOREIGN-ACCENTED SPEECH?



## *Roles of L1 Profiles, L2 Proficiency, Age, Experience, Familiarity and Metacognition*

Kazuya Saito

Mai Tran

Yui Suzukida

Hui Sun

Viktorija Magne

Meltem Ilkan

---

The current study examines how second language (L2) users differentially assess the comprehensibility (i.e., ease of understanding) of foreign-accented speech according to a range of background variables, including first language (L1) profiles, L2 proficiency, age, experience, familiarity and metacognition. A total of 110 L2 listeners first evaluated the global comprehensibility of 50 spontaneous speech samples produced by low, mid and high-proficiency Japanese speakers of English. The listeners were categorized into two subgroups according to a cluster analysis of their rating scores: lenient and strict. Results showed that while lenient appeared to rely equally on many linguistic areas of speech during their judgements, the strict listeners were strongly attuned to phonological accuracy. Analysis of the background questionnaire data revealed that the more lenient listeners likely had higher levels of awareness of the importance of comprehensibility for communication (metacognition); regularly used L2 English in professional settings (experience); and had L1s more linguistically close to the target speech samples, Japanese-accented English (L1-L2 distance).

---

We are grateful to two anonymous *SSLA* reviewers and the journal senior editor, Luke Plonsky, for their constructive feedback on an earlier version of the manuscript. This study was funded by Birkbeck Additional Research Grant.

## BACKGROUND

To date, there is ample evidence that adult second language (L2) speech is generally marked by some trace of foreign accent. This is very likely due to the inevitable interaction between the L2 and deeply entrenched, strongly developed first language (L1) system in the common linguistic space (Flege, 2016). Given that few late L2 learners can attain nativelike L2 proficiency (e.g., Abrahamsson & Hyltenstam, 2009), a growing number of scholars have emphasized the importance of evaluating the quality of L2 speech based on criteria that relate more to successful L2 communication in real-life settings (e.g., Ortega, 2018). This realistic, fundamental goal for adult L2 speech learning has motivated a great deal of research into the construct of comprehensibility, defined as the amount of effort that listeners need to make in order to understand the content of foreign-accented speech (Derwing & Munro, 2015).

The notion of L2 comprehensibility has been extensively discussed as one key element of L2 oral proficiency scale in acquisition (Saito, 2015; Nagle, 2018), assessment (Isaacs, Trofimovich, & Foote, 2018), and teaching literature (Gordon & Darcy, 2016; Saito & Saito, 2017). For example, in comprehensibility research, native speaking listeners are typically recruited to listen to and rate the comprehensibility of speech samples produced by L2 users with various proficiency levels. Previous research has shown that listeners likely attend to particular linguistic features in order to grasp the overall message of the samples, including segmental contrasts with high functional load and several minimal pairs (e.g., Munro & Derwing, 2006; Suzukida & Saito, forthcoming), prosodic accuracy and fluency (e.g., Kang, Rubin, & Pickering, 2010), and lexical appropriateness and complexity (e.g., Appel, Trofimovich, Saito, Isaacs, & Webb, in press; Saito, Webb, Trofimovich, & Isaacs, 2016). The features which factor most heavily into judgements are known to differ according to a number of variables such as the proficiency level of speakers in the samples (low, mid vs. high comprehensibility) (Saito, Trofimovich, & Isaacs, 2016) and the type of speaking task (interview vs. picture descriptions) (Crowther, Trofimovich, Isaacs, & Saito, 2015).

In addition to these, there is evidence showing that the background profiles of listeners can also affect L2 speech ratings. Listeners tend to be more lenient when they are more familiar with the accent varieties the speech that they are judging (e.g., Bradlow & Bent, 2003; Winke, Gass & Myford, 2013), or have language teaching experience (e.g., Kennedy & Trofimovich, 2008; Saito, Trofimovich, Isaacs & Webb, 2016). For example, a recent study by Saito and Shintani (2016) compared factors affecting the judgements of multilingual and monolingual listeners (Singaporeans vs. Canadians), finding that the former group assigned more lenient scores than the latter. Detailed analysis of the findings showed that the multilingual listeners considered all linguistic information in their judgements while the monolingual listeners were particularly influenced by phonological accuracy.

One gap in our current understanding of L2 comprehensibility judgements is that the majority of studies have relied on the judgments of L1 listeners alone. Accordingly, a growing amount of attention has been given to exploring the generalizability of findings from L1-listener research to other L2 users of English—i.e., exploring how L2 users understand each others' accented speech. This is a particularly pertinent extension of comprehensibility research given the status of English as a lingua franca in many places in the worldwide, and the fact that the majority of English speakers are actually L2 users (Pennycook, 2017). To date, however, there has been

little conclusive research indicating whether there are actual differences between how L1 and L2 listeners perceive the comprehensibility of L2 speech.

For instance, some studies have found L2 listeners' assessment patterns differ from those of L1 listeners especially when they evaluate familiar accents (produced by L2 speakers with the same and/or similar L1 profiles) (e.g., Foote & Trofimovich, 2018; Ludwig & Mora, 2017). In contrast, others have failed to find any advantage for familiarity at all, pointing out that both L1 and L2 listeners likely arrive at very similar L2 comprehensibility ratings (e.g., Crowther, Isaacs, & Trofimovich, 2016; Munro, Derwing, & Morton, 2006). This inconsistency suggests that L2 users' understanding of foreign-accented speech may be a multifaceted phenomenon that is intricately tied to a range of listener background factors, such as L2 proficiency level (e.g., Eger & Reinisch, 2018; Ludwig & Mora, 2017), quantity/quality of experience with the target language (e.g., Hayes-Harb, Smith, Bent, & Bradlow, 2008; Harding, 2012), and attitude, awareness and metacognition of foreign-accented speech (e.g., Derwing, Munro, & Rossiter, 2002).

In order to shed light on the complex mechanisms underlying L2 users' comprehensibility judgements, the current study took an exploratory approach towards examining how a total of 120 L2 ( $n = 110$ ) and L1 ( $n = 10$ ) users differentially assessed the comprehensibility of 50 beginner, intermediate and advanced-level L2 speech samples. By comparing responses on the listener background questionnaire with comprehensibility ratings, we aimed to reveal which listener factors—L1 profiles, L2 proficiency, age, quantity and quality of experience, and familiarity and metacognition of foreign accented speech—are *relatively* crucial for explaining individual variability in L1 and L2 speech ratings

## METHOD

### L2 Listeners

**Backgrounds.** To capture a wide range of listener backgrounds, we recruited a total of 110 L2 speakers in London, UK. London could be considered as an ideal site for accessing a number of L2 users with highly diverse ethnic, linguistic, experiential and sociopsychological backgrounds given the relatively high number of residents using a language other than English as their main form of communication (20% of residents compared to less than 5% nationwide) (UK Census, 2011). To recruit as many L2 listeners as possible, flyers were circulated at various locations (universities, language schools) and on social media. All the participants were considered as late learners who were first exposed to L2 English after 16 years of age. As described below and summarized in Table 1, these L2 listeners differed substantially in terms of their backgrounds.

1. **L1 Profiles (2 variables):** The listeners' L1 profiles comprised eight groups—German/Romance ( $n = 26$ ), Altaic ( $n = 25$ ), Slavic/Baltic ( $n = 18$ ), Mandarin Chinese ( $n = 15$ ), Austro-Asiatic ( $n = 12$ ), African ( $n = 10$ ), Indo-Iranian ( $n = 3$ ) and Arabic ( $n = 1$ ). Two ordinal scales were devised to measure the impact of L1 profiles on rating score: distance to Japanese-accented English (i.e., accent of the speech samples they were being asked to assess for L2 comprehensibility ;“2” = Japanese, “1” = Korean, Turkish,

Chinese<sup>1</sup>, “0” = other languages); and whether their L1 was from the Indo-European language family (“1” = Indo-European, “0” = Non-Indo European).

2. **L2 Proficiency (2 variables):** Previous research has indicated that L2 users’ speech judgements tend to be influenced by their actual (e.g., Eger & Reinisch, 2018) and self-perceived proficiency levels (e.g., Ludwig & Mora, 2017). To measure this, the listeners first engaged in the same speaking task they were being asked to rate (i.e., timed picture description; see “Material Preparation” below). Thirty seconds of each sample were extracted as a single WAV file and rated by two native speakers of British English on a 9-point scale (*1 = difficult to understand, 9 = easy to understand*). The high inter-rater reliability of the raters (Cronbach  $\alpha = .95$ ) led us to calculate an average score for each participant across the raters (*Range = 1-9*). As an additional measure of L2 proficiency, the participants were asked to rate their own comprehensibility in general using the same 9-point rubric (*Range = 1-8*).
3. **Age (2 variables):** Age-related factors have been found to exert a strong influence on various dimensions of L2 speech learning (e.g., Abrahamsson & Hyltenstam, 2009 for age of arrival vs. nativelike L2 pronunciation attainment; Pichora-Fuller & Souza, 2003 for aging vs. perceptual acuity). Given that L2 listeners’ representation, processing and proficiency is directly relevant to their L2 speech assessments (Eger & Reinisch, 2018), it is reasonable consider age a possibly influential background factor. To this end, the participants were interviewed to uncover their first intensive exposure to L2 English, operationalized as age of arrival in English speaking countries (e.g., UK, USA, Australia) and their chronological age at the time of the testing.
4. **Quantity and Quality of Experience (9 variables):** Following research standards in L2 speech literature (Flege, 2016), the *quantity* of the participants’ L2 experience (defined as sufficient, naturalistic and interactive exposure to the target language) was estimated based on total length of residence in English speaking countries (e.g., UK, USA). To further examine the *quality* of experience, the interview also elicited details about any L2 English teaching experience or linguistics classes they had taken in the past. Using the Language Contact Profile scheme (Freed, Dewey, Segalowitz, & Halter, 2004), the participants were also asked to report on the details of their recent L2 experience by estimating the ratio of their L1 and L2 use over the past few years in three different settings (work, social settings, home)<sup>2</sup>.
5. **Familiarity (4 variables):** Similar to previous literature (e.g., Winke et al., 2013), the participants’ familiarity with different kinds of accented speech was elicited based on self-reports on a 9-point scale (*1 = not familiar at all, 9 = very much familiar*). These items included (a) foreign-accented English in general, (b) Japanese-accented English, (c) Received Pronunciation English, and (d) General American English.

---

<sup>1</sup> While Korean and Turkish belong to the Altaic languages, Chinese is thought to insert much influence on the Japanese phonological systems, suggesting that the linguistic distance between these two languages is close. This is because many Chinese words have been imported to Japanese (see Shibatani, 1990).

<sup>2</sup> Note that the validity of self-reports has been criticized, especially when such methodology is applied to assess the quality of experience over years of immersion (see Ranta & Meckelborg, 2013).

6. **Metacognition of L2 Speech (5 variables):** Metacognition refers to the conscious awareness of one's own mental processes (Flavell, 1979). In this study, metacognition was operationalized as listeners' explicit understanding and awareness of the linguistic characteristics of foreign-accented speech. Metacognitive awareness of accented speech properties has been previously found to enhance listeners' understanding, confidence and attitudes in L2 speech judgements (e.g., Derwing et al., 2002 for L1 listeners' intelligibility assessment of Vietnamese-accented English). To tap into the participants' metacognitive awareness, they were asked to use a 9-point scale (*1 = not important, 9 = very important*) to give their opinion on how important they thought five different aspects of language were for successful communication: (a) speaking English without any accent like a native speaker; (b) speaking comprehensible English regardless of accentedness; (c) good pronunciation; (d) appropriate vocabulary/grammar; and (e) idiomatic and sophisticated expression.

The entirety of the data collection (interview, speech assessment/recording) took place in a quiet room at the researchers' residences, universities, and community centers (see Supporting Information-A and IRIS for the questionnaire that we used during the interview).

Table 1 *Summary of L2 Listener Background Variables*

<u>A. L1 profiles</u>				
• Distance to English ( $n = 64$ for L1 Indo European; $n = 46$ for L1 non-Indo European)				
• Distance to Japanese ( $n = 6$ for L1 Japanese; $n = 33$ for L1 Korean, Turkish & Chinese $n = 71$ for others)				
<u>B. L2 proficiency</u>				
	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
• Other perception of comprehensibility	5.3	2.6	1	9
• Self-perception of comprehensibility	3.0	1.5	1	8
<u>C. Age</u>				
• Chronological age	30.4	7.4	20	59
• Age of arrival	24.4	6.0	16	55
<u>D. Experience</u>				
• Length of residence	4.9	6.2	0.4	39
• Prior experience in TESOL	1.2	3.5	0	29
• Prior experience in linguistics training	0.3	0.4	0	1
• Recent L2 experience: L1 use at work	15.1	22.1	0	80
• Recent L2 experience: L2 use at work	83.1	23.4	10	100
• Recent L2 experience: L1 use in social settings	40.2	28.7	0	100
• Recent L2 experience: L2 use in social settings	56.9	29.7	0	100
• Recent L2 experience: L1 use at home	60.4	39.3	0	100
• Recent L2 experience: L2 use at home	35.7	38.1	0	100
<u>E. Familiarity</u>				
• Foreign-accented L2 English speech	6.5	2.0	2	9
• Japanese-accented L2 English speech	4.5	2.6	1	9
• Received Pronunciation	7.2	1.8	2	9
• General American	7.0	1.7	1	9
<u>D. Metacognition</u>				
• Nativelikeness	4.0	2.1	1	9
• Comprehensibility	7.7	1.7	1	9
• Pronunciation	6.4	1.8	1	9
• Lexicogrammar	6.0	2.0	1	9
• Complexity/sophistication	3.3	2.0	1	9

**Reduced Categories.** Whereas the raw categories covered a total of 24 listener background variables, it is possible that some may overlapped considerably (especially within the same thematic categories), resulting in multicollinearity problems. Because of this, we conducted a factor analysis to capture the latent categories underlying individual variability in the background characteristics of the 110 listeners. All raw background scores were submitted to a factor analysis with Varimax rotation. Following Loewen and Gonulal's (2015) field-specific guidelines, the factorability of the dataset was considered relatively high as shown by Bartlett's test of sphericity ( $\chi^2 = 1486.07, p < .001$ ) and the Kaiser-Meyer-Olkin measure of sampling adequacy (.509). Using the maximum likelihood method, the model identified nine factors with eigenvalues beyond 1, accounting for 74.76% of variance. A value of 0.6 was used as the threshold coefficient for practically significant factor loadings. Each factor is summarized in Table 2 and was interpreted as follows:

Factor 1 was "L1-L2 Distance," as items with the highest loadings pertained to participants' L1 profiles relative to English and Japanese. Factor 2 was "L2 proficiency," as the relevant items reflected the degree to which participants spoke comprehensible L2 English and demonstrated familiarity with various kinds of L2 English. Items related to experience were factored into 5 separate groups: Factor 3 for "Quantity of Experience," Factor 4 for "Social Use," Factor 5 for "Professional Experience," Factor 6 for "Use at Work," and Factor 7 for "Use at Home." Interestingly, the items related to metacognition appeared to tap into two different dimensions of listeners' orientations towards L2 speech assessment. Factor 8 was called "Preference for Nativelikeness," since the items pertained to the degree to which listeners believed in the importance of nativelike L2 oral proficiency with little familiarity with Japanese-accented English (the accent used in the rating materials). Factor 9 was designated as "Metacognition of Comprehensibility," because the items corresponded to listeners' unified notion of the importance of "comprehensible" pronunciation and lexicogrammar.

Table 2 *Factor Analysis of the Listener Background Variables*

	Factor 1 (L1-L2 Distance)	Factor 2 (L2 Proficiency)	Factor 3 (Quantity of Experience)	Factor 4 (Social Use)	Factor 5 (Professional Experience)	Factor 6 (Use at Work)	Factor 7 (Use at Home)	Factor 8 (Preference for Nativelikeness)	Factor 9 (Metacognition of Comprehensibility)
<u>A. L1 profiles</u>									
Distance to English	<b>-.853</b>	-.030	-.048	-.041	-.051	-.027	.162	.027	.015
Distance to Japanese	<b>-.725</b>	-.068	-.099	.115	.322	.176	-.173	-.067	-.197
<u>B. L2 proficiency</u>									
Other perception	-.164	<b>.648</b>	.086	-.132	-.289	-.060	.241	-.171	.323
Self-perception	.124	<b>.761</b>	-.200	-.252	.064	.123	.119	-.054	-.089
<u>C. Age</u>									
Chronological age	.110	-.072	<b>.907</b>	-.224	.163	.112	.008	-.026	-.053
Age of arrival	.150	-.262	.319	-.501	.353	.240	.234	.040	.082
<u>D. Experience</u>									
Length of residence	-.045	.182	<b>.748</b>	.238	-.152	-.132	-.109	-.087	-.159
TESOL experience	.036	.069	.153	-.191	<b>.772</b>	-.072	.191	-.068	.060
Linguistics training	-.151	.072	-.120	.121	<b>.803</b>	-.068	-.209	.009	.110
L1 use at work	-.064	-.026	.004	.114	-.030	<b>.912</b>	-.018	.013	.049
L2 use at work	.026	-.143	-.010	-.076	.076	<b>-.924</b>	-.011	.019	-.043
L1 use in social settings	.000	-.146	.087	<b>.755</b>	-.050	.292	.293	.019	-.006
L2 use in social settings	-.122	.141	-.020	<b>-.773</b>	.044	-.256	-.261	.036	-.011
L1 use at home	-.080	.043	-.002	.138	.014	.022	<b>.921</b>	.089	.024
L2 use at home	-.023	-.049	.055	-.125	.013	.004	<b>-.928</b>	-.031	-.005
<u>E. Familiarity</u>									
Foreign accents	.089	.522	.036	-.166	-.304	.131	-.116	-.457	.172
Japanese English	-.479	.105	.071	-.009	.087	-.090	-.076	<b>-.613</b>	.139
Received Pronunciation	.046	<b>.667</b>	.203	.087	-.110	-.331	.093	.123	.113
General American	-.178	<b>.682</b>	.202	.017	.078	-.035	-.041	-.025	.222
<u>D. Metacognition</u>									
Nativelikeness	-.112	.187	-.068	-.059	.000	-.078	.102	<b>.672</b>	.233
Comprehensibility	.028	.049	-.218	-.214	-.013	.132	.004	-.094	<b>.722</b>
Pronunciation	.015	-.082	-.043	.057	.046	-.097	.111	.100	<b>.794</b>
Lexicogrammar	.067	.084	.081	.125	.123	.088	-.105	.061	<b>.732</b>
Complexity/sophistication	.144	-.066	.263	.153	-.036	.111	-.355	.397	.395

Note. All loadings > .6 were highlighted in bold.

We are grateful to two anonymous SSLA reviewers and the journal senior editor, Luke Plonsky, for their constructive feedback on an earlier version of the manuscript. This study was funded by Birkbeck Additional Research Grant.

## Comprehensibility Judgements

**Material Preparation.** The stimuli for the comprehensibility judgements were 50 spontaneous speech samples representing a range of L2 speaking proficiency (Saito & Shintani, 2016). The dataset comprised the timed picture descriptions of L1 Japanese speakers which were carefully selected from the researchers' unpublished corpus of audio recordings of 200+ Japanese learners of English in Calgary, Montreal and Vancouver (deposited in IRIS; Marsden, Mackey, & Plonsky, 2016). The talkers' L2 experience profiles were equally distributed in terms of age of arrival ( $M = 27.8$ ,  $Range = 19-40$ ) and length of residence in Canada ( $M = 2.7$  years,  $Range = 1$  month to 11 years). In the task, the talkers described seven pictures depicting various scenes with three seconds of planning time each. The first four picture descriptions served as practice, while 10 seconds from each of the remaining three descriptions were extracted and used to generate the final speech samples. The length of each sample (30 seconds) is comparable to similar studies on L2 comprehensibility (e.g., Isaacs et al., 2018).

**Procedure.** First, each listener received a briefing on the purpose of the research and an explanation of L2 comprehensibility. Next, the listeners familiarized themselves with the three picture prompts and practiced the rating procedure using three samples not included in the main dataset (for training scripts and onscreen labels, see Supporting Information-B). All speech samples were then randomly ordered, presented to the listeners via *Praat* (Boersma & Weenik, 2018), and rated for comprehensibility using a 9-point scale ( $1 = difficult to understand$ ,  $9 = easy to understand$ ). Each session lasted approximately 60 minutes.

## Pronunciation, Fluency and Lexicogrammar Analyses

**Measures.** To illustrate how L1 and L2 listeners' comprehensibility ratings related to the linguistic qualities of the speech samples, we recruited five L1 English graduate students in applied linguistics to carry out a set of accuracy and fluency analyses (for a summary of the results, see Table 3). After receiving training developed and validated in Saito, Trofimovich and Issacs (2017), these raters listened carefully to each audio file via a custom MATLAB-based computer program, and provided holistic judgments of three specific aspects of L2 pronunciation proficiency (the correct use of segmentals, word stress, intonation). For each rating, a moving slider was used to record judgements on a 1000-point scale ( $0 = non-targetlike$ ,  $1000 = targetlike$ ). Using the same procedure, they then read written transcripts of the samples and rated them for two specific aspects of L2 lexicogrammar proficiency (appropriate use of vocabulary and morphology). Given the raters' high inter-rater reliability ( $\alpha > .80$ ), the scores were combined to generate a single average score for each sample and category.

For the fluency analysis, two researchers used *Praat* (Boersma & Weenik, 2018) to separately analyze each sample for three key elements of L2 fluency—i.e., breakdown, speed and repair (Kormos, 2012). For breakdown fluency, we counted the number of filled (e.g., ah, oh, eh) and unfilled ( $> 250$ ms of silence) pauses, and divided them by the total number of words.

Following Kormos's (2012) suggestion that the location of pauses should mirror the different cognitive processes involved in speech, we calculated pauses at the end of clauses (representing speakers' conceptualization of what to convey) and in the middle clauses (representing speakers' ongoing linguistic encoding of how to say it). For speed fluency, we divided the total number of syllables by the total phonation time (without any filled pauses) (i.e., articulation rate). For repair fluency, we divided the number of repetitions and self-corrections by the total number of words.

Table 3 *Descriptive Results of Pronunciation, Lexicogrammar & Fluency Measures*

	<i>M</i>	<i>SD</i>	<i>Range</i>	
			<i>Min</i>	<i>Max</i>
<b>A. Accuracy</b>				
• Segmentals	354	147	70	840
• Word stress	429	119	240	810
• Intonation	326	134	120	770
• Lexical appropriateness	714	125	413	926
• Morphosyntactic accuracy	482	200	78	892
<b>B. Fluency</b>				
• Mid-clause pause ratio	.21	.17	.00	.73
• Final-clause pause ratio	.13	.11	.00	.47
• Articulation rate	2.8	0.5	1.1	4.1
• Repair ratio	.05	.06	.00	.30

**Reduced Categories.** To control for the potential overlap between the nine accuracy and fluency measures, all linguistic scores were submitted to a factor analysis with varimax rotation. The factorability of the entire dataset was confirmed via two tests: Bartlett's test of sphericity ( $\chi^2 = 262.17, p < .001$ ) and the Kaiser-Meyer-Olkin measure of sampling adequacy (.781). Adopting the maximum likelihood method, the analysis identified four factors with eigenvalues beyond 1 (summarized in Table 4) which accounted for 84.6% of the variance in the linguistic measures. Factor 1 was labeled as "Pronunciation", because it encompassed all the pronunciation scores (segmentals, word stress, intonation). Factor 2 was labeled as "Meaning Delivery," because it captured how fluently the speech carried the meaning of the utterances (regardless of phonological accentedness). Final-clause pause ratio and repair ratio loaded separately onto Factors 3 and 4, respectively. In conjunction with Kormos's (2012) fluency framework, these two factors were labelled according to the type of cognitive processing they were presumed to represent—Factor 3 as "Conceptualization" and Factor 4 as "Monitoring".

Table 4 Factor Analysis of the Pronunciation, Lexicogrammar and Fluency Measures

	Factor 1 (Pronunciation)	Factor 2 (Meaning Delivery)	Factor 3 (Conceptualization)	Factor 4 (Monitoring)
<u>A. Accuracy</u>				
• Segmentals	<b>.847</b>	.201	.223	.043
• Word stress	<b>.888</b>	.373	.118	-.034
• Intonation	<b>.822</b>	.336	.052	.005
• Lexical appropriateness	.361	<b>.810</b>	.102	.014
• Morphosyntactic accuracy	.504	<b>.708</b>	.085	.028
<u>B. Fluency</u>				
• Mid-clause pause ratio	-.299	<b>-.804</b>	.089	.234
• Final-clause pause ratio	-.194	-.064	<b>-.928</b>	.161
• Articulation rate	.137	<b>.723</b>	.514	.034
• Repair ratio	.031	-.071	-.137	<b>.937</b>

Note. All loadings > .6 were highlighted in bold.

### Baseline Data (L1 Listeners' Comprehensibility Ratings)

To provide a point of comparison for L2 listeners' assessment patterns, we adopted the same L1 listeners and L2 comprehensibility ratings reported previously in Saito and Shintani (2016) with the same dataset (i.e., 50 picture description narratives). In the original study, the listeners were carefully selected to represent relatively homogeneous listener backgrounds. All of them were undergraduate students majoring non-linguistics subjects (e.g., business, psychology) at a university in Vancouver, Canada without any prior ESL/EFL teaching experience. They self-reported as monolinguals who had English as a main language of communication throughout their lives (their parents were L1 English speakers), and showed a minimal degree of familiarity with Japanese-accented English speech ( $M = 1.3$  on a 6-point scale). Given the high inter-rater agreement of these baseline raters ( $\alpha = .95$ ), their scores were combined to generate a single average rating for each sample and category.

## RESULTS

The first objective of the statistical analysis was to explore whether it was possible to categorize the 110 L2 listeners into smaller groups based on their comprehensibility ratings. To this end, the L2 users' ratings were submitted to hierarchical cluster analysis using Ward's method of minimum variance with Euclidian square distance intervals. A visual inspection of the dendrogram (see Figure 1) points to two large groups of relatively homogeneous listeners: Group A ( $n = 64$ ) and Group B ( $n = 46$ ). Since both of the groups demonstrated relatively high inter-rater agreement ( $\alpha = .90, .89$ , respectively), their comprehensibility scores were averaged for each speech sample, and compared to those of the L1 baseline group. According to a one-way repeated-measures ANOVA, comprehensibility ratings significantly differed across the three groups,  $F(2, 98) = 108.648, p < .01$ , Partial  $\eta_p^2 = .699$ . A post-hoc multiple comparison analysis demonstrated that Group A ( $M = 5.9, SD = 1.6$ ) assigned significantly higher and more lenient scores to the 50 samples than the L1 Baseline ( $M = 5.2, SD = 1.4$ ) with small-to-medium effects ( $d = 0.46$ ); while the L1 Baseline assigned significantly more lenient scores than Group B ( $M = 4.6, SD = 1.6$ ) small-to-medium effects ( $d = .039$ ). Based on this analysis, Group A was designated as L2 Lenient, and Group B as L2 Strict.

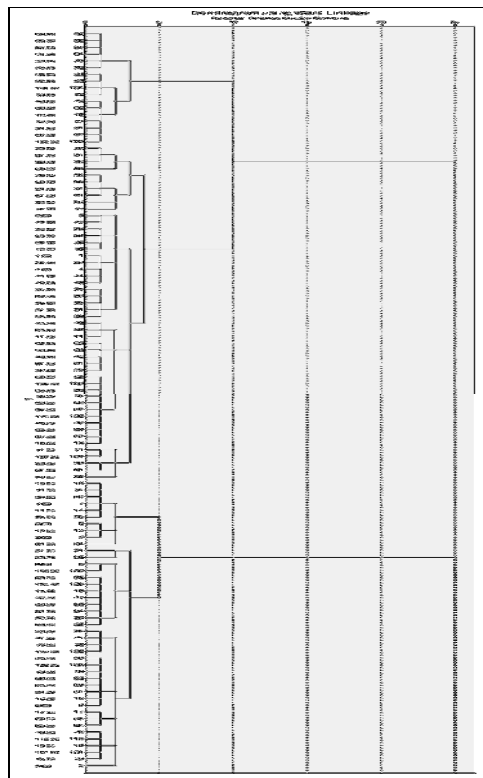


Figure 1. Dendrogram Tree of Hierarchical Clusters based on Listeners' Comprehensibility Scores

The next objective of the analysis was to examine the extent to which the three listener groups (L2 Lenient, L2 Strict, L1 Baseline) differentially weighted pronunciation, fluency and lexicogrammar information in their comprehensibility judgements. Following the same methodology and justifications presented in our original research (Saito & Shintani, 2016), the decision was made to run three different sets of multiple regression analyses with each group's comprehensibility ratings as the dependent variable, and the four linguistic factor scores as predictor variables (i.e., Pronunciation, Meaning Delivery, Conceptualization, and Monitoring). Given that our approach here was exploratory, we chose the "stepwise" method.

To determine the suitability of conducting multiple regression analyses, several necessary conditions were carefully checked. First, as explained in the manuscript, the 9 predictors originally included in the linguistic analyses (pronunciation, fluency & lexicogrammar) were reduced to 4 predictors through Factor Analyses. Second, the normality of each dependent variable (three groups' averaged comprehensibility ratings) was confirmed by Kolmogorov-Smirnov tests ( $p > .05$ ).

As summarized in Table 5, all the regression models appeared to account for roughly similar amounts of variance in L2 comprehensibility judgements across the three listener groups—80.7% for L2 Lenient, 80.6% for L2 Strict, 81.1% for L1 Baseline.<sup>3</sup> However, the relative weights of pronunciation, meaning delivery, conceptualization and monitoring factors differed to some degree. Specifically, we noted that the L1 listeners' comprehensibility ratings were more strongly influenced by phonological accuracy (54.7%) compared to ratings of the L2 listeners (38.8% for L2 Lenient, 41.7% for L2 Strict). Among the L2 listeners themselves, the lenient L2 listeners appeared to be more capable of attending to meaning delivery (the appropriate and fluent use of lexicogrammar) than the strict L2 listeners (32.1% vs. 27.8%).

---

<sup>3</sup> Compared to typical  $R^2$  values in regression models from L2 research, where the median is around .38 (see Plonsky & Ghanbar, 2018), the amount of variance reported here appears to be large ( $> .80$ ). In conjunction with what L2 speech research has thus far reported, the findings are not surprising; there is ample empirical evidence showing that raters greatly rely on pronunciation, fluency and lexicogrammar information to understand the content of L2 speech ( $R^2 = 60-80$ ) (for a comprehensive summary, see Saito & Plonsky, in press).

Table 5 Results of Multiple Regression Analyses Using the Linguistic Factors as Predictors of Comprehensibility

Predicted variable	Predictor variables	Adjusted $R^2$	$R^2$ change	$F$
Comprehensibility (L2 Lenient)	Pronunciation	.388	.388	30.384 <sup>a</sup>
	Meaning Delivery	.709	.321	57.212 <sup>a</sup>
	Conceptualization	.777	.068	53.372 <sup>a</sup>
	Monitoring	.807	.030	46.951 <sup>a</sup>
Comprehensibility (L2 Strict)	Pronunciation	.417	.417	34.278 <sup>a</sup>
	Meaning Delivery	.694	.278	53.412 <sup>a</sup>
	Conceptualization	.774	.080	52.508 <sup>a</sup>
	Monitoring	.806	.032	46.693 <sup>a</sup>
Comprehensibility (L1 Baseline)	Pronunciation	.547	.547	57.956 <sup>a</sup>
	Meaning Delivery	.742	.195	67.682 <sup>a</sup>
	Conceptualization	.786	.043	56.154 <sup>a</sup>
	Monitoring	.811	.025	48.125 <sup>a</sup>

Note. <sup>a</sup> $p$  value was below .001. The variables entered into the regression equations included Pronunciation, Meaning Delivery, Conceptualization, and Monitoring.

The final objective of the analysis was to explore which listener background factors could distinguish between the two L2 listener groups. To this end, a binary logistic linear regression was performed with listener group as the dichotomous dependent variable (Lenient, Strict), and the nine listener background factors (L1-L2 Distance, L2 Proficiency, Quantity of Experience, Social Use, Professional Experience, Use at Work, Use at Home, Preference for Nativelikeness, Metacognition of Comprehensibility) as independent variables. Using the forward stepwise method, the model was shown to be significant ( $\chi^2 = 4.285$ ,  $p = .038$ ) with three predictors—(a) Metacognition,  $B = -.584$ ,  $Exp(B) = .589$ ,  $p = .014$ ; (b) Use at Work,  $B = .523$ ,  $Exp(B) = 1.687$ ,  $p = .013$ ; and (c) L1-L2 Distance,  $B = -.431$ ,  $Exp(B) = .650$ ,  $p = .042$ —and explained a total of 19.8% of the variance in rating scores. The results of Nagelkerke r-squared measure demonstrated that the three listener background factors differentially contributed to the group distinction between the lenient and strict groups—Metacognition (8.6%), Use at Work (6.7%) and L1-L2 Distance (4.5%).

## DISCUSSION AND CONCLUSION

The current study set out to examine the sources of individual differences in L1 and L2 listeners' comprehensibility ratings of foreign-accented speech. The L2 listeners widely differed in terms of their L1 profiles, L2 proficiency, prior teaching/linguistics experience, amount of L2 use at work, socially, and at home, and the degree of their awareness, attitude and metacognition of foreign accented speech. In terms of *product* (what comprehensibility scores listeners eventually provided), compared to a baseline group ( $n = 10$  monolingual L1 listeners), these listeners were readily divided into two groups: (a) L2 listeners who assigned higher comprehensibility scores than the L1 Baseline (L2 Lenient); and (b) L2 listeners who assigned lower comprehensibility scores than the L1 Baseline (L2 Strict).

When it comes to *process* (how listeners arrived at such ratings), results showed that the L1 listeners' judgements were strongly tied to pronunciation accuracy. L2 listeners, on the other hand, seemed to equally weigh different areas of language (pronunciation, fluency, lexicogrammar) in their judgements. This finding is in line with previous research indicating differences in how inexperienced (monolingual) and experienced (multilingual) L1 listeners react to L2 speech (e.g., Saito & Shintani, 2016). Among the L2 listeners, however, the more lenient group appeared to give more weight to lexical appropriateness and fluency in their judgements. In addition, three background factors were found to distinguish between the ratings of the two groups: (a) metacognition (how much aware they were of the importance of "comprehensible" pronunciation and lexicogrammar); (b) experience (how much they were regularly using L2 English especially in professional [business, school] settings) and (c) L1-L2 distance (L1 distance from Japanese and English).

The findings for these three background factors can be interpreted with reference to psycholinguistic views on the role of experience in sound and word recognition. According to the well-known notion in psycholinguistics, perceptual adaptation (see Witteman, Weber, & McQueen, 2013 for a comprehensive review), *first language* listeners have the capacity to adjust, revise and develop their existing representations when exposed to systematic and novel deviations from familiar linguistic regularities. After a sufficient amount of experience, listeners become not only better at recognizing a range of unfamiliar sounds (e.g., Norris, McQueen, & Cutler, 2003), but also at understanding foreign accents (e.g., Bradlow & Bent, 2008). The findings of the current study can be interpreted as providing some of the first evidence for the occurrence of this phenomenon among *second language* listeners themselves.

For example, we found that more lenient L2 listeners were more conscious of the important roles that comprehensible pronunciation and lexicogrammar (instead of nativelikeness) play in successful communication. These comprehensibility-oriented L2 listeners may in turn have been more willing to accommodate particular varieties of accented-speech (e.g., Japanese speakers' spontaneous speech in English) (e.g., Derwing et al., 2002), and strive to robustly develop their L2 comprehension skills by making the most of opportunities to interact with L1 and L2 interlocutors alike (see Leow, 2000 for the roles of awareness in general L2 learning). This latter interpretation stems from the finding that lenient listeners reported using L2 English very frequently in business and school settings with talkers from multiple language backgrounds—a key condition for successful perceptual adaptation (Sidas, Alexander, & Nygaard, 2009).

In addition, we found that L2 listeners assigned higher comprehensibility scores when they had certain L1 profiles (L1 proximity to English and Japanese) which were similar to the target speech samples—Japanese-accented English. This proximate L1-L2 distance may have helped promote the listeners' perceptual adaptation—a phenomenon referred to as "the interlanguage speech intelligibility benefit" (Hayes-Harb et al., 2008). All of the factors listed here led L2 listeners to accumulate more experience with various types of English speech and by extension modify their phonological representations to better comprehend the foreign-accented samples.

Two findings that run contrary to previous literature is the lack of relationship between L2 proficiency, familiarity, and the lenient vs. strict distinction. First, it is important to remember that we here operationalized L2 proficiency according to the groups' performance on the same listening task (picture narrative judged by L1 listeners). One possible explanation for the lack of any effect could be that L2 users' production abilities may not necessarily index the way they

process foreign-accented speech, which concurs with theoretical and empirical evidence in L2 speech research. Whereas many major theories agree that perception and production are essentially interlinked (e.g., Flege, 2016), the former may precede and not necessarily chronologically align with the latter (Baker & Trofimovich, 2006). The relationship between L2 users' proficiency and their L2 speech assessment thus needs to be demonstrated in replication studies, especially those which include more outcome measures that reflect various dimensions of L2 users' proficiency on both the perception and production levels (cf. Ludwig & Mora, 2017).

Second, the current study did not identify L2 users' familiarity with Japanese accented English as a significant predictor of their individual differences in comprehensibility judgements. These results are discordant with existing literature which has evidenced the role of familiarity in distinguishing between L1 and L2 speech assessment (e.g., Winke et al., 2013). The lack of familiarity effects in the current study may be attributable to the greater importance of metacognition, experience and/or L1-L2 distance in comprehensibility judgement; that is, the L2 listeners' familiarity with Japanese accented English may have been indirect at best, if their L1 profiles were far from English/Japanese, they did not use English on a daily basis, or/and they did not have high-level were not highly awareness of comprehensibility as a crucial goal of L2 communication.

### 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.
- Appel, R., Trofimovich, P., Saito, K., Isaacs, T., & Webb, S. (in press). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking English raters. *International Journal of Applied Linguistics*.
- Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, 44, 231–250.
- Boersma, D., & Weenink, P. (2018). *Praat: Doing phonetics by computer version 6.0.29*. Retrieved from <http://www.praat.org>
- Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106, 707–729.
- UK Census (2011) Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/articles/languageinenglandandwales/2013-03-04>
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, 99, 80-95.
- Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*, 2, 160-182
- Derwing, T. M. & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.
- 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.
- Eger, N. A., & Reinisch, E. (2018). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition*. doi: 10.1017/S0272263117000377
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34, 906.
- Flege, J. (2016, June). *The role of phonetic category formation in second language speech acquisition*. Plenary address delivered at New Sounds, Aarhus, Denmark.
- Foote, J. A., & Trofimovich, P. (2018). Is It Because of My Language Background? A Study of Language Background Influence on Comprehensibility Judgments. *Canadian Modern Language Review*, 74, 253-278.
- Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation*, 2, 56-92.
- Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, 29, 163-180.
- Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of phonetics*, 36, 664-679.
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26, 349-356.

- Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, 35, 193-216.
- Kang, O., Rubin, D., Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English, *Modern Language Journal*, 94, 554-566.
- Kormos, J. (2012). *Speech production and second language acquisition*. Routledge.
- Leow, R. P. (2000). A study of the role of awareness in foreign language behavior. *Studies in Second Language Acquisition*, 22, 557-584.
- Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In Plonsky, L. (Ed), *Advancing quantitative methods in second language research*. New York: Routledge.
- Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, 3, 167-198.
- Marsden, E., Mackey, A., & Plonsky, L. (2016). *Breadth and depth: The IRIS repository*. Routledge.
- Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520-531.
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Nagle, C. (2018). Motivation, Comprehensibility, and Accentedness in L2 Spanish: Investigating Motivation as a Time-Varying Predictor of Pronunciation Development. *The Modern Language Journal*, 102, 199-217.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47, 204-238.
- Ortega, L. (2018). SLA in uncertain times: Disciplinary constraints, transdisciplinary hopes. *Working Papers in Educational Linguistics*, 33, 1.
- Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge.
- Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, 42, 11-16.
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, 102(4), 713-731.
- Ranta, L. & Meckelborg, A. (2013). How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *The Canadian Modern Language Review*, 69, 1-33.
- Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, 65, 563-595.
- Saito, K., & Plonsky, L. (in press). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*.
- Saito, K., & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive second language comprehensibility? *TESOL Quarterly*, 50, 421-446.
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37, 217-240.

- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 38, 439-462.
- Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2016). Re-examining phonological and lexical correlates of second Language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.). *Interfaces in second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141-156). Bristol, UK: Multilingual Matters.
- Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness and sense relations. *Studies in Second Language Acquisition*, 37, 677-701.
- Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21, 589-608.
- Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- Sidas, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125, 3306-3316.
- Suzukida, Y., & Saito, K. (forthcoming). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the Functional Load principle.
- 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.
- Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75, 537-556.

## SUPPORTING INFORMATION-A: RATER QUESTIONNAIRE

### Basic info

- (1) Age: \_\_\_\_\_ years old
- (2) Age of Arrival: \_\_\_\_\_ years old
- (3) Where?      ① UK              ② North America      ③ Australia/NZ      ④ Others
- (4) Why? ① Study abroad      ② Work abroad      ③ Immigration      ④ Others
- (5) Have you ever taught English before? \_\_\_\_\_ years (e.g., 0-10 years)
- (6) Have you ever had music training before? \_\_\_\_\_ years (e.g., 0-10 years)
- (7) Have you taken any linguistics classes/training before?      (0 = no; 1 = yes)

### Length of residence

- (8) UK: \_\_\_\_\_ years
- (9) North America: \_\_\_\_\_ years
- (10) Australia/NZ: \_\_\_\_\_ years
- (11) Others: \_\_\_\_\_ years (0 = NO) (which countries? \_\_\_\_\_ )

### L2 English Learning in Classroom Settings

- (12) Age of learning in classroom settings: \_\_\_\_\_ years old
- (13) Length of learning in classroom settings: \_\_\_\_\_ years

### Languages in Use other than L2 English

- (14) First language (from birth/most dominant): Which language? ( \_\_\_\_\_ ) (LANG ID)
- (15) L3: Other dominant language A (other than L2 English): Which language? ( \_\_\_\_\_ ) (0 = NO or LANG ID)
- (16) L3: Other dominant language B (other than L2 English): Which language? ( \_\_\_\_\_ ) (0 = NO or LANG ID)

### Use Profile: Average over the past 1-2 years

#### Frequency at work/school (professional settings)

- (17) L1/most dominant: \_\_\_\_\_ % (0-100%)
- (18) English (with fluent speakers): \_\_\_\_\_ % (0-100%)
- (19) English (with non-fluent speakers): \_\_\_\_\_ % (0-100%)
- (20) Dominant Lang A: \_\_\_\_\_ % (0-100%)
- (21) Dominant Lang B: \_\_\_\_\_ % (0-100%)

Frequency with friends (social settings)

- (22) L1/most dominant: % (0-100%)
- (23) English (with fluent speakers): % (0-100%)
- (24) English (with non-fluent speakers): % (0-100%)
- (25) Dominant Lang A: % (0-100%)
- (26) Dominant Lang B: % (0-100%)

Frequency at home

- (27) L1/most dominant % (0-100%)
- (28) English (with fluent speakers): % (0-100%)
- (29) English (with non-fluent speakers): % (0-100%)
- (30) Dominant Lang A: % (0-100%)
- (31) Dominant Lang B: % (0-100%)

**Experience with Japanese**

- (32) I am in contact with Japanese people (1 = very infrequently, 9 = very frequently)
- (33) I have studied Japanese (0 = no, 1 = yes)
- (34) Length of learning: years (e.g., 0-30 years)
- (35) Length of residence: years (e.g., 0-3 years)

**Familiarity (9-point scale: 1 = I am not familiar at all, 9 = I am very much)**

With English

- (36) Received Pronunciation (1-9)
- (37) General American (1-9)
- (38) Australian/NZ English (1-9)

With "foreign accented" English

- (39) I am familiar with different kinds of foreign accented English (1-9)
- (40) I am familiar with Japanese accented English (1-9)

**Self-Assessment of L2 English**

***Rate your own oral proficiency (L2 English) in terms of nativelikeness and comprehensibility***

Comprehensibility

- (41) How easy is it to understand your L2 English? (1 = difficult to understand, 9 = easy to understand)

**Metacognition: Judging others' L2 English oral proficiency**

*While judging others' L2 English oral proficiency, which aspects of language do you think are relatively crucial for successful communication? Please rate the following statements on a 9-point scale (1 = not important, 9 = very important)?*

- (42) Speaking English without any accent like a native speaker
- (43) Speaking comprehensible English regardless of accentedness
- (44) Good pronunciation
- (45) Appropriate vocabulary/grammar
- (46) Idiomatic & sophisticated expression

**SUPPORTING INFORMATION-B: Training Scripts and Onscreen Labels of L2**

**Comprehensibility Judgements**

**A. Training scripts for comprehensibility judgement**

---

Comprehensibility This term refers to how much effort it takes to understand what someone is saying. If you can understand with ease, then a speaker is highly comprehensible. However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker has low comprehensibility.

---

**B. Onscreen labels**

How comprehensible? (easy to understand - difficult to understand)

9. Easy to understand

8.

7.

6.

5.

4.

3.

2.

1. Difficult to understand