



Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*, 2(2), 160-182. <https://doi.org/10.1075/jslp.2.2.02cro>

Peer reviewed version

License (if available):  
Unspecified

Link to published version (if available):  
[10.1075/jslp.2.2.02cro](https://doi.org/10.1075/jslp.2.2.02cro)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via John Benjamins Publishing at <http://dx.doi.org/10.1075/jslp.2.2.02cro>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

RUNNING HEAD: Nonnative Listeners' Perception of Accent and Comprehensibility

**Linguistic Dimensions of Second Language Accent and Comprehensibility: Nonnative  
Listeners' Perspectives**

Crowther, D., Trofimovich, P., & Isaacs, T. (accepted). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*.

## **Abstract**

The current study investigated the effect of listener status (native, nonnative) and language background (French, Mandarin) on global ratings of second language speech. Twenty-six nonnative English listeners representing the two language backgrounds ( $n = 13$  each) rated the comprehensibility and accentedness of 40 French speakers of English. These same speakers were previously rated by native listeners and coded for 19 linguistic measures of speech (i.e., segmental errors, word stress errors, grammar accuracy) in Trofimovich and Isaacs (2012). Analyses indicated no difference in global ratings between nonnative and native listeners, or between the two nonnative listener groups. Similarly, no major differences in the linguistic dimensions associated with each group's ratings existed. However, analyses of verbal reports for a subset of nonnative listeners ( $n = 5$  per group) demonstrated that each group attributed their ratings to somewhat different linguistic cues.

*Keywords:* accent, comprehensibility, English language, native, nonnative, pronunciation, raters

# **Linguistic Dimensions of Second Language Accent and Comprehensibility: Nonnative Listeners' Perspectives**

## **1. Introduction**

In keeping with recent calls for a multilingual and multicultural view of second language (L2) learning and teaching (Cook, 2007; Ortega, 2013), L2 interaction research has increasingly begun to address the communicative benefits that L2 knowledge provides for intercultural communication (May, 2014). This research has specifically highlighted L2 users' ability not only to mutually work towards understanding when using a shared L2 (e.g., Jenkins, Cogo, & Dewey, 2011) but also to draw upon all available linguistic and cultural knowledge to facilitate comprehension (e.g., Blommaert, 2010). However, despite this growing research momentum towards a greater focus on the multilingual nature of L2 use, a significant body of pronunciation research has to date been conducted from what is typically referred to as the nativeness view (Levis, 2005). This view encompasses studies which examine L2 speech in reference to native speakers of the target language, who typically act as raters evaluating L2 speech for comprehensibility and accentedness, sometimes in conjunction with other measures (e.g., Derwing & Munro, 2015). Briefly, as a measure of understanding used in a broad sense, comprehensibility targets listeners' perceived ease of understanding, most frequently rated using 9-point scales (Authors, xxxx). Accentedness, which is often used as a measure of nativeness, refers to listener ratings of how well an L2 speaker can approximate the speech patterns of the target language, again most often rated using 9-point scales (Jesney, 2004).

An active line of research carried out within the nativeness tradition has targeted native listeners' perceptions of the linguistic dimensions that uniquely relate to comprehensibility and

accentedness. This research has revealed that accent ratings are primarily linked to such dimensions of phonology as segmental (vowel/consonant) accuracy (Munro & Derwing, 2006; Trofimovich & Isaacs, 2012), word stress, pitch range, pause or syllable length (Kang, 2010; Kang, Rubin, & Pickering, 2010), and speech rate (Munro & Derwing, 2001) and that comprehensibility is additionally related to appropriate grammar and lexical usage and discourse structure (Fayer & Krasinski, 1987; Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, 2015). In essence, this research clarifies which aspects of L2 speech contribute to listener perceptions' of nativeness (making a speaker sound more or less accented) and which aspects relate to communicative success (in terms of making a speaker easier or harder to understand). However, one shortcoming of this research is its exclusive focus on native listeners, with little consideration paid to the dimensions of speech considered important by nonnative listeners. A primary focus on native listeners is incompatible with the growing global trend towards L2 use among nonnative speakers (e.g., Jenkins, 2014; May, 2014). To address this gap, the current study investigates which linguistic dimensions nonnative listeners use in their perception of L2 comprehensibility and accentedness.

### **1.1. A focus on nonnative listeners**

Compared to research targeting native listeners' evaluation of L2 speech (see Derwing & Munro, 2015), there are considerably fewer studies investigating which dimensions underlie nonnative listeners' perception of the speech by fellow L2 users, with existing research focusing on global differences between native and nonnative listeners or between nonnative listeners from different language backgrounds. For example, studies comparing the ratings of accent, intelligibility, or teaching effectiveness by native and nonnative listeners have provided conflicting findings (Winke, 2013), with evidence that nonnative listeners, compared to native listeners, have

demonstrated both more (Fayer & Krasinski, 1987; Kang, 2012) and less severity (Brown, 1995) in rating L2 speech, or have shown no difference in rating (Derwing & Munro, 2013; Flege, 1988; MacKay, Flege, & Imai, 2006). Beyond rating speech, Jun and Li (2010), using a think-aloud protocol, found differences in what the two listener groups consider during the rating of comprehensibility and accentedness. Nonnative listeners focused primarily on segmental errors and such suprasegmental features as intonation and linking, while native listeners relied on global considerations, such as general understanding of the utterance, presence of specific speech characteristics (e.g., lisp), or frustration in deciphering the intended message (see Kormos & Dénes, 2004, for similar evidence from fluency ratings). So even if both sets of listeners converge in their overall ratings in the end, their assessment processes may differ in terms of the linguistic features that contribute to the scores they assign.

With respect to how listeners' language background impacts their perception of L2 speech, research has generally revealed an advantage for listeners familiar with speakers' language background, although these findings are also inconsistent (e.g., Bradlow & Bent, 2008; Winke & Gass, 2013). The benefits of a shared language background between nonnative listeners and speakers were shown by Bent and Bradlow (2003), who found, for instance, that low proficiency Korean speakers were as intelligible to Korean listeners as high proficiency Chinese speakers. Harding (2012) similarly found a shared language advantage for both Japanese and Chinese listeners' comprehension of Japanese and Chinese speakers, though this advantage was more pronounced for the Chinese group. In contrast, Major, Fitzmaurice, Bunta, and Balasubramanian (2002) showed that while Spanish listeners performed better in a listening comprehension exam when hearing a Spanish speaker of English, Japanese and Chinese listeners did not benefit from a shared listener-speaker background (see also Munro, Derwing, & Morton,

2006). It appears, then, that nonnative listeners might understand speech better when spoken by speakers from their own background at least some of the time, but the strength of this effect likely depends on factors such as speaker and listener proficiency, context, and learner background characteristics (Smith & Hayes-Harb, 2011; Xie & Fowler, 2013). In sum, while the above findings are interesting, they reveal little information beyond documenting quantitative differences in the perception of L2 speech by different groups of listeners and virtually no qualitative insight into the dimensions used by nonnative listeners in evaluating L2 speech for various constructs, including comprehensibility and accentedness.

## **1.2. A lingua franca perspective**

Further research relevant to understanding the linguistic dimensions underlying the perception of L2 speech by nonnative listeners comes from studies of intercultural communication among nonnative speakers using a shared L2 (Deterding & Kirkpatrick, 2006; Jenkins, 2000; Seidlhofer, 2011). This research, primarily looking at L2 English, has targeted sources of communicative breakdowns during interactions involving nonnative speakers, focusing on aspects of phonology, such as segmental accuracy (e.g., Jenkins, 2000), and lexis and grammar, such as the omission of articles, pluralisation of uncountable nouns, and the nonuse of the third person present tense –s (e.g., Seidlhofer, 2004). This approach is best exemplified in Jenkins' (2000) proposed Lingua Franca Core (LFC), or a set of linguistic dimensions most relevant to mutual understanding among nonnative users of English. The LFC distinguishes between core pronunciation features necessary for mutual understanding and noncore features that mark regional accent (see also Walker, 2010). The core features of the LFC largely include segmental aspects of L2 production, such as the articulation of individual vowels and consonants (Park & Wee, 2015). However, suprasegmental features, including word stress, pitch range, and rhythm—speech dimensions

shown to be relevant for native listeners in subsequent research (e.g., Field, 2005; Kang, 2010; Kang, Rubin, & Pickering, 2010; Isaacs & Trofimovich, 2012; Saito et al., 2015)—are mostly relegated to noncore status in this framework. More recent ELF research, though, has argued for the role of intonation as a means to negotiate and maintain successful interaction (Pickering, 2009; Pickering & Litzenberg, 2011). Despite the controversy surrounding the contents of the LFC (e.g., Jenkins, 2000, vs. Scheuer, 2005), this research has yielded evidence of the potential importance of various linguistic dimensions linked to nonnative listeners' understanding of L2 speech, motivating further research eliciting nonnative listeners' ratings.

## **2. The current study**

The overarching goal of this study was therefore to extend current L2 speech research to include nonnative listeners and to broaden the knowledge base about nonnative listener perception of L2 speech beyond studies of misunderstanding in intercultural communication. The specific objectives were to determine which linguistic dimensions of L2 speech are associated with nonnative listeners' ratings of comprehensibility (ease of understanding) and accentedness (nativeness), and to directly compare nonnative listeners who do and do not share speakers' language background. Our focus on comprehensibility and accentedness was motivated from both theoretical and applied perspectives. In terms of theory, a measure of ease or difficulty of understanding may reflect L2 speakers' conversational experience better than a measure of nativeness, such that the linguistic dimensions tied to comprehensibility, rather than those linked to accent, would be more likely to create opportunities for learners to engage in interaction-driven learning (e.g., Mackey, Gass, & McDonough, 2000). And from a practical viewpoint, distinguishing comprehensibility from accentedness can help language teachers isolate and prioritize those linguistic targets with the greatest potential to influence listeners' understanding.<sup>1</sup>



This study's objectives were addressed through a partial replication of a study by Trofimovich and Isaacs (2012) focusing on 60 native English listeners' ratings of comprehensibility and accentedness in the speech of 40 French speakers of English. Trofimovich and Isaacs conducted correlation and regression analyses between the listeners' ratings and 19 coded linguistic measures in the speakers' output, showing that accentedness was linked primarily to measures of phonology (e.g., segmental accuracy, word stress) while comprehensibility was related to measures of both phonology (e.g., word stress) and lexis/grammar (e.g., grammatical accuracy, lexical richness). The current study advances this research by considering the perceptions of nonnative listeners from two language backgrounds (French, Mandarin) using the same speech stimuli and speech measures, thus allowing for direct comparisons between the two nonnative groups as well as across the nonnative and the original native listener data. A further methodological feature of this study is that it involved not only correlational data (as in most L2 speech research) but also listeners' introspective reports about the dimensions they attended to while making comprehensibility and accentedness judgments (see Jun & Li, 2010). The current study was guided by the following research questions:

1. Do native and nonnative listeners differ in their perception of L2 comprehensibility and accentedness? Does nonnative perception differ as a function of listener language background?
2. Which linguistic dimensions of L2 speech are linked to listeners' perception of L2 comprehensibility and accentedness? Do these dimensions vary as a function of listener status (nonnative vs. native) or language background (French, Mandarin)?

Based on previous research focusing on listener background effects, we predicted that there might be differences in how native and nonnative listeners approach the rating of L2

speech, with nonnative listeners being more likely than native listeners to attend to fine-grained, segmental and suprasegmental dimensions of L2 speech (e.g., Jun & Li, 2010). We also predicted a potential shared-L1 advantage—in line with the previously reported interlanguage speech intelligibility benefit for speakers and listeners sharing a common language background (e.g., Bent & Bradlow, 2003). Thus, in terms of comprehensibility, we anticipated that French listeners' familiarity with French-accented speech may lead them to perceive it to be easier to understand. And based on prior research on understanding in intercultural communication, we anticipated that nonnative listeners would rely on segmental aspects of L2 production (e.g., Jenkins, 2000) and might also invoke aspects of L2 lexis and grammar (e.g., Seidlhofer, 2004) in their rating. Because no previous research has compared linguistic dimensions of L2 comprehensibility and accentedness for native versus nonnative listeners, no specific predictions were made regarding potential differences in the linguistic dimensions associated with these ratings.

### **3. Method**

#### **3.1. Listeners**

The participants included 26 nonnative English-speaking, international graduate students from four large universities in the American Midwest. These participants were assigned to two listening groups ( $n = 13$ ) based on their native language (L1) background: French (12 French, 1 Belgian) and Mandarin Chinese (12 Chinese, 1 Taiwanese). The inclusion of French listeners allowed for examining a shared background effect, since the L1 of the speakers was French. The Mandarin listeners represented a typologically distant language from French, specifically in terms of its segmental inventory and rhythm, with Mandarin (unlike French) also being a tonal language (Duanmu, 2007; Jun, 2005; Swan & Smith, 2001). Graduate students were targeted due

to university requirements that they reach a required benchmark of English proficiency: a minimum score of 80 on the TOEFL iBT and 19/30 in both the speaking and listening subtests. The Mandarin listeners (nine female) all attended the same university. Due to a limited number of French-speaking graduate students at one university, seven of the 15 French listeners (nine female) were recruited from three additional universities. To facilitate direct comparisons between native and nonnative listeners, data from the native English group who evaluated the same speech materials in Trofimovich and Isaacs (2012) were used for comparison. The native listeners consisted of 60 undergraduate students (34 females), native speakers of North American English, enrolled in a Canadian English-medium university. All listeners represented various academic disciplines. Table 1 summarizes background characteristics of the three listener groups; Kruskal-Wallis tests (conducted because the data were nonnormally distributed) revealed no differences between the two nonnative groups ( $p > .18$ ).

TABLE 1

### **3.2. Speech materials**

To allow for maximum comparability of findings, the target audio recordings were the same as those used by Trofimovich and Isaacs (2012) with native listeners. The recordings came from 40 native French speakers of L2 English from Quebec, Canada (27 females;  $M_{age} = 35.6$  years, 28–61). Their speech was recorded as part of an eight-frame picture narrative, which depicted two travelers bumping into each other on a busy street corner, accidentally exchanging their bags, and realizing their mistake upon returning home (Derwing, Munro, Thomson, & Rossiter, 2009). All stories were edited down to between 23 and 36 seconds with initial dysfluencies (i.e., uh, um) removed. Each recording was coded for 19 linguistic measures, which spanned four broad

categories of phonology, fluency, lexis/grammar, and discourse (for examples and further details on the measures and coding, including reliability measurements, see Trofimovich & Isaacs).

For phonology, there were six coded categories at segmental and suprasegmental levels: (1) *segmental errors* (total number of vowel and consonant substitutions divided by total number of segments articulated); (b) *syllable structure errors* (total number of vowel and consonant insertion and deletion errors over total number of syllables articulated); (3) *word stress errors* (total number of instances where primary stress was misplaced or missing over total number of polysyllabic words produced); (4) *rhythm as vowel reduction ratio* (number of correctly reduced syllables in both polysyllabic words and function words divided by total number of obligatory vowel reduction contexts); (5) *intonation errors* (number of correct pitch patterns at the end of phrases over total number of instances where pitch patterns were expected); and (6) *pitch range* (difference between the highest and lowest fundamental frequency F0 values).

For fluency, there were six categories designed to describe various dysfluencies in L2 speech: (7) *filled pausing* (total number of nonlexical pauses, such as uh and um, longer than 400 milliseconds); (8) *unfilled pausing* (number of unfilled pauses longer than 400 milliseconds); (9) *pause errors* (number of filled and unfilled pauses produced within clauses or phrases, where no such pauses would normally be expected, divided by number of total pauses produced); (10) *repetitions/self-corrections* (sum of all immediately repeated and self-corrected words over total number of words produced); (11) *articulation rate* (total number of syllables produced divided by total duration of speech sample in seconds, excluding dysfluencies); and (12) *mean length of run* (mean number of syllables produced between two adjacent filled or unfilled pauses greater than 400 milliseconds).

There were four categories for lexis/grammar: (13) *grammar errors* (total number of words with at least one error divided by total word count); (14) *lexical errors* (total number of incorrectly used lexical expressions over total number of words uttered); (15) *token frequency* (total number of words produced, normalized for sample length); and (16) *type frequency* (total number of unique words produced, normalized for sample length). Under discourse, there were three additional coded categories: (17) *story cohesion* (number of adverbials used as cohesive devices between storytelling elements, such as *therefore* or *as a result of*); (18) *story breadth* (number of distinct propositions or storytelling elements, that is, predicate followed by another element, used within speech sample); and (19) *story depth* (number of different proposition categories in speech sample), with all discourse measures adjusted for sample length.

### **3.3. Procedure**

The audio recordings were rated by nonnative listeners for comprehensibility and accent using 9-point Likert scales following Trofimovich and Isaacs (2012). Comprehensibility was defined as the degree of ease or difficulty in listeners' understanding of L2 speech (1 = *hard to understand*, 9 = *easy to understand*), and accent referred to how different a speaker sounded from a native speaker of North American English (1 = *heavily accented*, 9 = *not accented at all*). The nonnative listeners completed ratings in two groups. The first group consisted of eight listeners from each L1 who completed ratings through an online interface using SurveyMonkey. All listeners first received the definition of the two rated constructs, and were then given the opportunity to familiarize themselves with the narrative task (Derwing & Munro, 2013). They then completed three practice ratings using non-L1 French speakers completing the same task and proceeded to rate the 40 target recordings for comprehensibility and accent (in that order),

with the presentation order randomized for each listener. Although the listeners had the option to listen to each recording multiple times, it was rare that a second listening was used.

Following a practice rating and task familiarization session, the second group of nonnative listeners ( $n = 5$  per L1 background) rated 30 of 40 recordings in randomized order. For the final 10 recordings, they first rated each speaker and then explained their scoring decisions (Authors, xxxx). Five of these recordings differed across the listeners. The other five were identical, purposefully selected to represent a range of comprehensibility and accent levels, based on mean ratings from Trofimovich and Isaacs (2012). Before engaging in verbal reports for the target recordings, the listeners practiced rating followed by verbal comments using two non L1-French audio recordings. All verbal reports were audio recorded for analysis.

### **3.4. Analysis**

Cronbach's alpha was calculated for ratings of comprehensibility ( $\alpha_{\text{Mandarin}} = .94$ ,  $\alpha_{\text{French}} = .90$ ) and accentedness ( $\alpha_{\text{Mandarin}} = .94$ ,  $\alpha_{\text{French}} = .97$ ), with both nonnative listener groups showing high internal consistency comparable to that obtained by Trofimovich and Isaacs (2012) for native listener ratings of comprehensibility ( $\alpha_{\text{English}} = .99$ ) and accentedness ( $\alpha_{\text{English}} = .99$ ). Mean comprehensibility and accentedness scores were then calculated for each speaker, averaging across all listeners, separately for each listener group. A priori power analyses carried out using statistical software R revealed less than a 30% chance of finding significant group differences or associations between listeners' ratings and the 19 coded measures of speech, assuming medium-sized effects. However, even smaller sample sizes ( $n = 10$ ) have previously proven sufficient for detecting associations in similar research (e.g., Authors, xxxx). For all statistical tests, alpha was set at .05. Because the nonnative group ratings were not normally distributed, nonparametric tests were run (Larson-Hall, 2010).

The listeners' verbal reports were transcribed and subsequently coded by the first author for references to any of the four linguistic categories: phonology, fluency, lexis/grammar, and discourse. Broad linguistic dimensions were used instead of the 19 specific variables (e.g., intonation errors, token frequency, articulation rate) because the comments often did not contain sufficient detail to categorize them narrowly. For instance, a comment coded for phonology indicated that a listener attributed his or her rating to an aspect of segmental or suprasegmental phonology, as in *She has a weird intonation and about the accent so the pronunciation sounds American but the intonation not so that's why I just put five because I don't know how to think because of this* (accent rating by French listener). A comment coded for fluency, such as *So it was kinda easy to understand the beginning but towards the ends there were some like confused like she stopped talking and changing her words so that made it a little bit harder to understand* (comprehensibility rating by a Mandarin listener), would include references to different elements of this construct, including repetitions/self-corrections (*changing her words*), and pausing and articulation rate (*she stopped talking*). Because listeners often mentioned multiple reasons for their decisions, a comment could receive multiple codes. For instance, the comment *So like there are some error in grammar or pronunciation but still I could understand the story* (comprehensibility rating by a Mandarin listener) was coded as referring to both lexis/grammar and phonology, as neither reason was given priority over the other or greater depth in explanation. After the initial coding was completed, the second author rechecked the transcripts and discussed any disagreements until consensus was reached. Listener comments for comprehensibility and accentedness were tallied separately for each listener group (French, Mandarin) across the four broad categories (phonology, fluency, lexis/grammar, discourse).

## 5. Results

### 5.1. Comprehensibility and accentedness

The first analysis focused on listeners' comprehensibility and accentedness ratings. Table 2 summarizes descriptive statistics for both nonnative groups, along with those reported for native listeners in Trofimovich and Isaacs (2012). Spearman correlations ( $r_s$ ) revealed strong associations between comprehensibility and accentedness for each group ( $r_{\text{Mandarin}} = .92$ ;  $r_{\text{French}} = .91$ ;  $r_{\text{English}} = .91$ ), and Wilcoxon signed-ranks tests targeting within-group comparisons showed that all listeners rated the speakers higher for comprehensibility than accent,  $z_{\text{Mandarin}} = -4.94$ ,  $p < .001$ ,  $z_{\text{French}} = -5.49$ ,  $p < .001$ ,  $z_{\text{English}} = -5.36$ ,  $p < .001$ . However, Kruskal-Wallis tests, conducted to compare the three listener groups' comprehensibility and accentedness ratings, revealed no significant differences between the groups for either comprehensibility,  $\chi^2(2) = 1.62$ ,  $p = .45$ , or accentedness,  $\chi^2(2) = 1.96$ ,  $p = .38$ , suggesting that the listeners—regardless of their background—assigned similar ratings to the speakers. Furthermore, the ratings for each construct were strongly correlated across the three groups ( $r_s > .90$ ), indicating that the recordings rated high or low in comprehensibility or accentedness by one listener group were perceived the same way by the other two.

TABLE 2

### 5.3. Linguistic measures of speech

The next analysis investigated associations between listeners' comprehensibility and accentedness ratings and the 19 coded measures based on the speakers' language output. We set .70 as a benchmark of a large effect size, which reflected about 50% of shared variance in correlation measures. These associations (Spearman  $r_s$ ) are summarized in Table 3. Six coded measures featured strong associations with comprehensibility for all groups ( $r_s > .70$ ): word



stress errors, rhythm, mean length of run, token frequency, type frequency, and story breadth. However, only word stress errors and rhythm were strongly associated with accentedness across groups. In terms of group differences, story breadth was tied to the Mandarin and English listeners' perception of comprehensibility, but not to the French listeners' comprehensibility ratings. Furthermore, type frequency was linked to the French listeners' accentedness ratings but was less relevant to the Mandarin and English listeners' judgments. In sum, the three listener groups showed close similarity in how linguistic dimensions of L2 speaker output patterned with listeners' judgments of comprehensibility and accentedness.

TABLE 3

#### **5.4. Verbal reports**

The final analysis focused on the verbal reports provided by the two groups of L2 listeners ( $n = 5$  per group) immediately after assigning comprehensibility and accentedness ratings. In total, the Mandarin listeners made 49 comments and the French 63, but there was variability across listeners in both the frequency and detail of comments (see Table 4).

TABLE 4

##### **5.4.1. Comprehensibility**

When discussing features of speech that influenced their perception of comprehensibility, the two nonnative listener groups showed distinct response patterns. The French listeners not only made more total comments (39), compared to the Mandarin listeners (28), their comments were also more diverse, referencing measures related to phonology (12), fluency (16), and lexis/grammar (11). The Mandarin listeners, however, focused primarily on fluency (17), with limited references to lexis/grammar (6), discourse (3), and phonology (2). Thus, the French listeners seemed to make direct connections between their ease of understanding a speaker and

various features of his or her output, as demonstrated in the following examples, all from French (F) raters listening to a target recording (R).

- I would say that's kinda easy to understand 8 but then sometimes it confusing with the pronounce like him her (pronounced with silent /h/) each other etcetera (F3: R38; phonology category).
- And it was also really easy to understand because of the pronunciations of words but also the tone employed the stress and stuff like that (F3: R40; phonology category).
- I'm... right now because I understood what she was saying but it was so interrupted but I needed really to focus on what she was saying instead of the pauses and stuff like that so I would say 6 (F3: R22; fluency category).
- Ok so this time it's I put just 7 because I understand what she she says but like she she hesitates a lot it's I need more concentration (F5: R30; fluency category).
- What was hard to understand was grammar yeah I don't know maybe that's a bad answer but... Ok grammar mixing up the subjects and the verbs and so I have to remember what was the subject because he didn't use the right one and again I can hear an accent but that's not the problem (F2: R3; lexis/grammar category).
- So I put 8 for comprehensibility I could understand everything without making any effort but it was not perfect especially in the choice of vocabulary (F4: R16; lexis/grammar category).

In contrast, the chief difficulty in comprehensibility cited by the Mandarin (M) listeners was in how fluently the lexical content was delivered, as illustrated in the following comments.

- So for comprehensibility the story was quite clear he was not doing like uh um I thought he was telling the story quite smoothly (M2: R16; fluency category).

- In a given time the story was maybe just a start the sentences were broken but kinda going slow so kinda understand it (M2: R19; fluency category).
- It's like he can pronounce every word almost right but I still don't understand what maybe if I can connect every word I know what he's saying but I don't really understand every sentence he say (M5: R3; fluency category).
- It's hard to understand because it's not the words are not connected together and there's lots of like I don't know in the sentence (M4: R37; fluency category).
- This person similar to previous one but she has a limited vocabulary but the thing makes me understand her is that she can catch the key words although now you know the complete sentence (M1: R36; lexis/grammar category).

With respect to comprehensibility, then, the difference between the French and the Mandarin listeners was that the former group associated comprehensibility ratings with a wider range of dimensions (phonology, fluency, lexis/grammar) while the latter mainly focused on fluent delivery of speech, with limited reference to lexis/grammar.

#### **5.4.2. Accentedness**

When discussing their accentedness ratings, both nonnative listener groups referred primarily to aspects of phonology, specifically at the segmental or word levels. For the Mandarin listeners, 19 out of the 21 comments referred to segmental or word-level errors. Of the two non-phonology comments, one referred to fluency and one to the lexicon. It appears that the Mandarin listeners focused on the word level in rating accentedness, as shown in the following examples.

- He has better accent and comprehensibility but still in some words like straight tower you can tell the difference from a native speaker (M1: R35; phonology category).

- And the way she pronounce women is a little bit different from the accent I was used to (M4: R38; phonology category).
- It's like I will say she's accented is like she's any words like a building and street or streeting like this way but not a nature way to speak English (M5: R18; phonology category).

As illustrated in the following comments, the French listeners' comments about accentedness also centered on phonology issues (20), with the remaining four comments related to fluency.

- Her accent I gave her a 3 because the few words I heard she was saying weren't that badly pronounced so that's why I gave her a 3 for the few words she said she said them right (F1: R36; phonology category).
- ...and I would say 2 for the accent he have a strong pronunciation sometimes we can see it with the R and also the O (F3: R3; phonology category).

However, unlike the Mandarin listeners who were unfamiliar with French, the French listeners relied on their French knowledge, as they consistently made both implied and overt comparisons between English and French in evaluating accentedness, as shown in these excerpts.

- The under the rhythm sounds really American to me so the only accent I hear is just on letters or particular words (F2: R26; phonology category).
- It sounds like he's speaking French but he's changing the words into English yeah that's exactly there's no effort to work on the accent at all (F4: R3; phonology category).
- So it's easy to understand because it sounds French sometime she uses some words in French or she use English word but in a very French way (F5: R29; phonology category).

With respect to accentedness rating, then, the primary difference between the French and the Mandarin listeners was that the former had access to a shared L1 with the speakers, which

enabled them to draw comparisons between L2 speech and the target variety. In other respects, the two listener groups were similar, in that both were oriented towards segment and word-level phonology in evaluating accentedness.

## **6. Discussion**

This study addressed two main questions: (1) what differences (if any) exist between native and nonnative listeners in their ratings of L2 comprehensibility and accentedness, and (2) which linguistic measures are linked to listeners' perception of L2 speech and whether these measures differ as a function of listener status (nonnative vs. native) and their language background (Mandarin vs. French). Analyses comparing ratings between nonnative and native listeners revealed similarities in global ratings and showed no major differences in linguistic dimensions associated with native and nonnative listener perception of L2 speech. However, analyses of verbal reports for a subset of nonnative listeners demonstrated that the Mandarin and French listeners attributed their ratings to somewhat different linguistic cues.

### **6.1 Global ratings**

At the global level, the three listener groups behaved similarly, rating comprehensibility higher than accentedness, which aligns with prior research showing that the two ratings reflect overlapping yet distinct constructs (Derwing & Munro, 2015). Of more importance, though, was the lack of significant difference between groups, supporting research which has previously identified little to no difference between native and nonnative listeners (Derwing & Munro, 2013; Flege, 1988; MacKay et al., 2006). A similarity between these studies and the current investigation is a focus on global ratings of speech, such as judgments of overall comprehensibility and accentedness. When differences have been found to occur, the rating focus has either typically been descriptive, targeting perceived levels of annoyance, distraction,

and instructional competence (Fayer & Krasinski, 1987; Kang, 2012), or occupationally specific, including tour guides' ability to describe itineraries and deal with customer concerns (Brown, 1995).

The current data from global ratings also showed little benefit of shared language background for French listeners, compared to Mandarin listeners, in rating the L2 speakers from the French-language background, for both comprehensibility and accentedness. This is consistent with studies showing small or inconsistent familiarity effects based on a shared listener-speaker background (Harding, 2012; Major et al., 2002; Munro et al., 2006). Based on the results of Hayes-Harb, Smith, Bent, and Bradlow (2008), the shared background benefit might manifest itself only for low-proficiency nonnative listeners evaluating low-proficiency speakers. In this study, all listeners were university students who had surpassed a minimum threshold of speaking proficiency deemed necessary for academic study, and the 40 speakers represented a range of ability levels, with low-proficiency speakers comprising only about a quarter of the sample.

## **6.2. Linguistic dimensions of accentedness**

In line with previous research, listeners appeared to attend primarily to measures of phonology when rating accentedness (Trofimovich & Isaacs, 2012; Saito et al., 2015). In correlation analyses, these dimensions involved suprasegmental features of L2 speech, namely, word stress errors and rhythm. These findings for nonnative listeners were identical to those reported by Trofimovich and Isaacs for native raters, where word stress errors and rhythm together accounted for a total of 76% of variance in accentedness. Based on the high correlation between the measures of word stress and rhythm ( $r_s = -.81$ ), it appears that errors in word stress placement co-occurred with insufficiently reduced vowels in unstressed positions, revealing an expected link between vowel quality and stress. In contrast to correlation analyses, nonnative listeners'

verbal reports about accentedness centered on segmental aspects of phonology, accounting for 88% of all comments for the Mandarin listeners (21/24) and 83% of the comments for the French listeners (20/24). These findings are again consistent with those reported by Trofimovich and Isaacs, where segmental influences on accent ratings were most clearly observed in teachers' verbal reports, not in correlation analyses.

In order to rate L2 accentedness, listeners likely pay attention only to those dimensions of speech that help them determine how nonnative speech differs from the target variety (Authors, xxxx; Munro, 2008). For French speakers of L2 English, such dimensions most likely include stress placement and vowel quality (reduction), captured through the measures of word stress errors and rhythm.<sup>2</sup> Indeed, a key difference between English and French speech is in rhythmic timing, which entails that the alterations between stress and unstressed syllables in English are not present in French (Ramus, Nespors, & Mehler, 1999). It is noteworthy, however, that both French and Mandarin listeners were cognizant of the importance of word stress and rhythm for French speakers to sound nativelike in English. While the French listeners may have relied on their background knowledge of French, which was evident in their explicit reference to French in verbal reports for accentedness, the Mandarin listeners, whose native rhythmic pattern falls somewhere between that of English and French (Swan & Smith, 2001), had limited potential for a crosslinguistic advantage.

It is likely that the relatively advanced level of L2 proficiency for Mandarin listeners may have offset their lack of familiarity with French, helping them perform similarly to the French listeners (and to the native listeners) in evaluating French speakers. This conclusion is consistent with diminished benefits of a shared listener-speaker background for listeners beyond an initial level of L2 proficiency (Hayes-Harb et al., 2008). And the overt emphasis on segmental errors,

evident in verbal comments of both the French and Mandarin listeners, is compatible with earlier research on rater reports showing that segmental aspects of phonology are generally salient for both experienced and inexperienced raters and that suprasegmental aspects of phonology are difficult to describe without reference to specialized terminology (Authors, xxxx).

### **6.3. Linguistic dimensions of comprehensibility**

This dataset provides one of the first demonstrations that nonnative and native listeners are similar in associating L2 comprehensibility with a range of linguistic dimensions, spanning the domains of phonology, fluency, lexis, grammar, and discourse. The Mandarin and French listeners performed similarly to each other (and to the group of native listeners) in that their comprehensibility ratings were linked to measures of word stress and rhythm (suprasegmental phonology), mean length of run (fluency), type/token frequency (breadth and depth of vocabulary), and story breadth (discourse), supporting earlier research with native listeners (Isaacs & Trofimovich, 2012; Fayer & Krasinski, 1987; Saito et al., 2015). Unlike accentedness, which is mostly associated with aspects of segmental and suprasegmental phonology, comprehensibility seems to require listeners to process a broad range of linguistic information, such that ease of understanding is shaped at multiple linguistic levels.

The links between comprehensibility ratings and linguistic content of L2 speech were quantitatively similar for the nonnative listeners who do and do not share the speaker's language background and for the native and nonnative listeners, with the only exception that the French listeners showed no association with the measure of story breadth (see Table 3). It is plausible that, at least for advanced nonnative listeners from different language backgrounds, L2 comprehensibility is not associated with quantitative differences in estimating the effort involved in understanding. Put differently, advanced L2 proficiency (which likely reflects awareness of L2



pronunciation, along with more experience using the L2 and more generally being an expert language user) might help eliminate major differences not only between nonnative listeners with and without familiarity with the speaker's linguistic background, but also between native and nonnative raters.

Despite similarity in global ratings, the two nonnative listener groups showed differences in the quality of their assessment. The French listeners' comments revealed an emphasis on a more diverse set of linguistic measures linked to comprehensibility (phonology, fluency, lexis/grammar), compared to the Mandarin listeners who mainly reported on fluency. Put simply, the two groups' ratings were quantitatively similar, but likely for different reasons. This result is similar to recent findings by Hayes-Harb and Hacking (2015) who showed that native English listeners, while being consistent in their numerical rating of accentedness, attributed their ratings to different reasons, often citing nonlinguistic information, such as speakers' possible social and cultural background. While there was no evidence of such listener behavior in this dataset, the Mandarin and French listeners nonetheless differed in their verbal reports. This difference likely reflected the French listeners' familiarity with French, which may have reduced their processing burden, allowing them to focus more on phonological, lexical, and grammatical accuracy in L2 output. In contrast, the Mandarin listeners, for whom French-accented English was less familiar, likely found the processing of the overall message more effortful, concentrating on fluency barriers to comprehensibility. Although possible links between comprehensibility and speaker accuracy versus fluency for nonnative listeners varying in familiarity with the speaker's L1 is an intriguing finding (which awaits further research), this result should be viewed against the backdrop of no numerical differences in comprehensibility ratings. Even if the listeners weighed their ratings differently, the outcome was highly similar.

## **7. Directions for future research and conclusion**

Future research would need to address several limitations of the current study. One limitation is that the current study addressed comprehensibility, or perceived ease of understanding. Just because L2 speech is perceived to be easy to understand does not guarantee a message is understood accurately, though comprehensibility is highly tied to intelligibility, which reflects listeners' accuracy of understanding (Derwing & Munro, 2015). In addition, it would be important to include a larger sample of both listeners and speakers to increase statistical power. It would also be essential to consider a wider range of L1 backgrounds, both from the speaker and listener perspective, to allow for greater generalizability of research findings.<sup>3</sup>

In this study, both nonnative listener groups were composed of university graduate students of relatively high L2 proficiency; therefore, it remains to be seen if nonnative listeners' assessments of L2 speech might differ as a function of listener L2 proficiency. Further, the nonnative listeners were asked to reflect on their rating process in their L2 rather than their L1, which may have limited their ability to fully express themselves. Using listeners' L1 for verbal reports (e.g., Kahng, 2014) may thus allow listeners to provide more elaborate explanations for their rating behavior, or potentially access vocabulary and concepts not available to them in their L2. Last but not least, it would be crucial to target nonnative listeners' perception of L2 speech collected from various tasks, including unstructured interactive tasks, in order to increase the applicability of listener ratings to real-life interaction (e.g., Authors, xxxx).

The chief aim of the current study was to investigate how nonnative and native listeners perceived the comprehensibility and accentedness of L2 speech recorded by L2 speakers of English as part of a picture narrative task. At a global level, our findings showed no essential differences between the listener groups, reinforcing the idea that—at least for the materials,

tasks, and participants involved in this study—what is easy to understand and what sounds nativelike for one set of listeners might evoke similar reactions in another listener group, despite possible differences in listener status, such as being a native versus nonnative language user, or in listener L1 background. These findings must be refined in future research, ideally with learners in classroom settings engaged in pedagogical tasks, with the view of enhancing instructional approaches for effective teaching of L2 pronunciation targeting various populations of L2 speakers and listeners.

### **Note**

1. We acknowledge that a measure of listeners' actual accuracy, such as the use of orthographic transcription, comprehension questions, or true/false judgments (Munro & Derwing, 2015), compared to listeners' scalar ratings of ease or difficulty of understanding, would provide a more objective (and, hence, less impressionistic) measure of listeners' understanding of the speech. Nevertheless, comprehensibility remains an influential (and useful) construct, often considered to be a user-friendly, reliable measure of understanding in both research settings and in high-stakes assessment contexts (see Isaacs & Trofimovich, 2012). Targeting comprehensibility, rather than intelligibility, also allowed us to make direct comparisons with the findings reported by Trofimovich and Isaacs (2012), who focused only on comprehensibility and accentedness.
2. The association between L2 accentedness and type frequency for the French listeners might also be interpreted as indicative of these listeners' familiarity with French. The speakers who produced more content (in terms of distinct word types) received a higher accentedness rating, suggesting that only for the French listeners (for whom presumably understanding the

segmental and suprasegmental detail of individual words was less effortful) considered a non-phonology dimension in rating accentedness.

3. An additional concern raised by an anonymous reviewer was that general attitudes towards French-accented English may have impacted listeners' evaluation of comprehensibility and accentedness. Though this is a valid concern, as attitudinal and affective factors have previously been shown to influence rater behavior (e.g., Kang & Rubin, 2009), none of the Mandarin listeners identified the speech samples as being French-accented, which minimizes the potential for overt affective and attitudinal influences on speech rating.

## References

Authors (xxxx).

Bent, T., & Bradlow, A. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*, 1600-1610.

Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

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

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, *12*, 1-15.

Cook, V. (2007). The goals of ELT: Reproducing native-speakers or promoting multicompetence among second language users? In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching*. New York: Springer.

- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, 63, 163-185.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Philadelphia: John Benjamins.
- Derwing, T.M., Munro, M. J., Thomson, R.I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533-557.
- Deterding, D., & Kirkpatrick, A. (2006). Emerging south-east Asian English and intelligibility. *World Englishes*, 25 (3/4), 391-409.
- Duanmu, S. (2007). *The phonology of standard Chinese: The phonology of the world's languages*. Oxford: Oxford University Press.
- Duff, P. (2015). Transnationalism, multilingualism, and identity. *Annual Review of Applied Linguistics*, 35, 57-80.
- 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. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, 84, 70-79.
- 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., & Hacking, J. F. (2015). Beyond rating data: What do listeners believe underlies their accentedness judgments? *Journal of Second Language Pronunciation*, 1, 43-64.

- 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.
- Issacs, 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.
- Jenkins, J. (2014). *English as a lingua franca: The politics of academic English language policy*. New York: Routledge.
- Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching*, 44, 281- 315.
- Jesney, K. (2004). The use of global foreign accent rating in studies of L2 acquisition. Calgary, AB, Canada: University of Calgary Language Research Centre Reports.
- Jun, H. G. & Li, J. (2010). Factors in raters' perceptions of comprehensibility and accentedness. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 53-66), Ames, IA: Iowa State University.
- Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford: Oxford University Press.
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64, 809-854.

- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301-315.
- Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9(3), 249-269.
- Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28, 441-456.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, 94, 554-566.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145-164.
- 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.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471-497.
- MacKay, I. R., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, 27, 157-183.

- Major, R. C., Fitzmaurice, S. F., Bunta, F., Balasubramanian, C. (2001). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36, 173-190.
- May, S. (Ed.). (2014). *The multilingual turn*. New York: Routledge.
- Munro, M. J. (2008). Foreign accent and speech intelligibility. *Phonology and second language acquisition*, 193-218.
- 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., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Ortega, L. (2013). SLA for the 21<sup>st</sup> century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, 63(S1), 1-24.
- Park, J. S-Y. & Wee, L. (2015). English as a lingua franca: Lessons for language and mobility. In C. Stroud & M. Prinsloo (Eds.). *Language, literacy and diversity: Moving words* (pp. 55-71). New York: Routledge.
- Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics* 6, 235-255.
- Pickering, L. & J. Litzenberg. (2011). Intonation as a pragmatic resource in ELF interaction, revisited. In A. Archibald, A. Cogo & J. Jenkins (Eds.), *Latest trends in ELF research* (pp. 77-92). Newcastle, UK: Cambridge Scholars Publishing.



- Ramus, F., Nespov, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Saito, K., Trofimovich, P., & Isaacs, T. (2015). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*. Published online 02 February 2015.
- Scheuer, S. (2005). Why native speakers are (still) relevant). In K. Dziubalska-Kolaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 111-130). Frankfurt a/M: Peter Lang
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209-239.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Smith, B. L., & Hayes-Harb, R. (2011). Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of Phonetics*, 39, 115-120.
- Swan, M., & Smith, B. (Eds.). (2001). *Learner English: A teacher's guide to interference* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15, 905-916.
- Walker, R. *Teaching the pronunciation of English as a lingua franca*. Oxford: Oxford University Press.
- Winke, P. (2013). Rating oral language. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-7). Blackwell Publishing, Ltd.

Winke, P., & Gass, S. (2013). The influence of L2 experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, *47*, 762-789.

Xie, X., & Fowler, C. A. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, *41*, 369-378.

Table 1

*Means (Standard Deviations) for Listener Background Characteristics*

Variable	Mandarin	French	English
Chronological age (years)	25.31 (2.18)	25.31 (2.66)	20.7 (1.75)
Residence in USA (years)	2.04 (1.93)	1.93 (2.53)	–
Age of L2 learning (years)	10.38 (2.03)	11.00 (1.96)	–
TOEFL iBT score (0-120)	99.23 (9.02)	95.58 (12.20)	–
English speaking (1-9) <sup>a</sup>	6.31 (1.80)	6.77 (1.79)	8.88 (0.37)
English listening (1-9) <sup>a</sup>	6.85 (1.57)	7.54 (1.33)	8.92 (0.28)
English use (0-100%) <sup>b</sup>	48.46 (23.04)	60.77 (27.83)	93.04 (7.15)
English use with NSs (0-100%) <sup>b</sup>	36.92 (24.63)	35.00 (26.80)	–

*Note.* <sup>a</sup>Self-rating (1 = *low*, 9 = *high*). <sup>b</sup>Self-rating (0% = *none*, 100% = *all the time*).

Table 2

*Descriptive Statistics for Comprehensibility and Accentedness Ratings*

Listeners	Comprehensibility			Accentedness		
	<i>M (SD)</i>	<i>Mdn</i>	<i>Range</i>	<i>M (SD)</i>	<i>Mdn</i>	<i>Range</i>
Mandarin	6.48 (1.74)	6.65	2.23–8.85	5.40 (1.74)	4.73	2.46–8.77
French	6.65 (1.23)	6.69	3.08–8.69	4.95 (1.75)	4.50	2.77–8.38
English	6.24 (1.55)	6.15	2.43–8.92	5.10 (1.50)	4.64	3.13–8.50

Table 3

*Spearman Correlations Between 19 Speech Measures and Speech Ratings*

Linguistic measure	Comprehensibility			Accentedness		
	Mandarin	French	English	Mandarin	French	English
Segmental errors	-.58	-.55	-.59	-.67	-.64	-.61
Syllable structure errors	-.46	-.44	-.47	-.46	-.48	-.48
Word stress errors	-.80*	-.78*	-.77*	-.73*	-.79*	-.82*
Rhythm	.79*	.80*	.80*	.73*	.82*	.79*
Intonation errors	.50	.50	.56	.46	.50	.49
Pitch range	-.08	-.03	-.01	-.10	-.10	-.09
Filled pausing	-.47	-.40	-.40	-.37	-.33	-.32
Unfilled pausing	-.23	-.28	-.22	-.16	-.17	-.09
Pause errors	-.59	-.58	-.59	-.39	-.47	-.45
Repetitions/corrections	-.58	-.58	-.57	-.57	-.59	-.58
Articulation rate	.61	.56	.60	.51	.54	.45
Mean length of run	.72*	.71*	.72*	.60	.69	.59
Grammar errors	-.47	-.52	-.53	-.40	-.55	-.51
Lexical errors	-.46	-.41	-.50	-.37	-.35	-.39
Token frequency	.75*	.71*	.75*	.62	.69	.61
Type frequency	.76*	.72*	.75*	.63	.71*	.62
Story cohesion	.43	.40	.49	.35	.40	.34
Story breadth	.74*	.67	.71*	.66	.65	.65
Story depth	.49	.42	.41	.48	.45	.45

Note. \* $r_s > .70$ .

Table 4

*Number of Comments Reported by Mandarin and French Listeners*

Listener	Comprehensibility			Accentedness		
	Mandarin	French	Sum	Mandarin	French	Sum
Listener 1	5	6	11	8	3	11
Listener 2	8	6	14	6	5	11
Listener 3	3	11	14	2	4	6
Listener 4	6	6	12	3	7	10
Listener 5	6	10	16	2	5	7
Total comments	28	39	67	21	24	46