

Is It Because of My Language Background? A Study of Language Background Influence on Comprehensibility Judgments

Jennifer A. Foote and Pavel Trofimovich

Abstract: This study examines the role of listeners' native language (L1) background in judgments of comprehensibility (ease of understanding) for speakers from same and different L1 backgrounds, to determine the extent of a shared second language (L2) comprehensibility benefit. Forty L2 English speakers from Mandarin, French, Hindi, and English backgrounds (10 per group) listened to speech samples from 30 L2 English speakers from Mandarin, French, and Hindi backgrounds (10 per group). Listeners first evaluated each speaker's comprehensibility and provided verbal reports indicating their reasons for each rating. To estimate pronunciation influences on comprehensibility, listeners then rated each speaker for four speech measures (segmental and word stress errors, intonation, speech rate). Results revealed that different speech measures were associated with comprehensibility ratings for different listener–speaker groups, and that a match in L1 background accounted for additional unique variance in comprehensibility ratings, but only for the Mandarin listeners and speakers. Verbal reports indicated that listeners more often considered L1 a benefit when rating speakers from their own L1 and a detriment when evaluating speakers from a different L1. Findings overall point to small effects of shared L1 background on comprehensibility, suggesting alternative priorities for teaching and researching comprehensible L2 speech.

Keywords: comprehensibility, intelligibility, language background, pronunciation, speech perception

Résumé : Les auteurs étudient le rôle des antécédents linguistiques en langue maternelle (L1) des auditeurs dans leurs jugements quant à la compréhensibilité (facilité de compréhension) des locuteurs possédant les mêmes antécédents et des antécédents différents en L1, en vue de déterminer l'importance de l'avantage du partage d'une langue seconde (L2) à cet égard. Quarante locuteurs de l'anglais L2 possédant des antécédents en mandarin, en français, en hindi et en anglais (10 par groupe) écoutent les échantillons parlés de 30 locuteurs de l'anglais L2 possédant des antécédents en mandarin, en français et en hindi (10 par groupe). Les auditeurs évaluent d'abord la compréhensibilité du discours de chaque locuteur et exposent oralement les raisons de leur notation. Pour estimer l'incidence de la prononciation sur la compréhensibilité, les

auditeurs notent ensuite chaque locuteur à l'égard de quatre critères d'évaluation du discours (erreurs de découpage sémantique et d'accentuation des mots, intonation et débit). Les résultats révèlent que des critères d'évaluation du discours différents sont associés aux notes relatives à la compréhensibilité selon les groupes auditeur-locuteur et que la parité des antécédents en L1 explique une portion supplémentaire unique de l'écart dans les notes relatives à la compréhensibilité, mais seulement dans le cas des auditeurs et des locuteurs du mandarin. Les exposés oraux indiquent que les auditeurs considèrent plus souvent la L1 comme un avantage lorsqu'ils évaluent les locuteurs de leur propre L1 et comme un inconvénient lorsqu'ils évaluent les locuteurs d'une L1 différente. Dans l'ensemble, les constatations des auteurs convergent vers l'existence d'une faible incidence d'antécédents partagés en L1 sur la compréhensibilité, ce qui ouvre la voie à d'autres priorités dans l'enseignement des L2 et la recherche sur la compréhensibilité du discours en L2.

Mots clés : antécédents linguistiques, compréhensibilité, intelligibilité, perception du discours, prononciation

The past 20 years have seen increased interest in comprehensibility from pedagogical and research perspectives, as the goal of pronunciation instruction has shifted from attempting to help second language (L2) learners sound like native (L1) speakers to helping learners become easier to understand (Levis, 2005). However, while comprehensibility is an appropriate goal for L2 learning, it is also complex. There are many variables that contribute to how well one person understands the speech of another. At different times, L2 speakers of a given language may be using their L2 to communicate with L1 speakers of the target language or with other L2 speakers. While some research has explored the role of different L1 backgrounds in the perception of L2 speech, much remains unknown. For instance, little is known about the role that L1 background may play in the relative contributions of specific linguistic dimensions of speech, such as segmentals or suprasegmentals, to comprehensibility judgments. Further, there are unanswered questions about how the L1 backgrounds of both listeners and speakers influence how difficult listeners judge speech to be and the extent to which listeners attribute ease or difficulty in understanding to the language background of the speaker. Given that far more people communicating in English are using it as an L2 rather than an L1, it is important that comprehensibility be examined with L2 listeners, not just L1 listeners.

Understanding L2 comprehensibility

When talking about comprehensibility, it is important to distinguish it from intelligibility. Typically in L2 research, comprehensibility denotes

subjective judgments about how easy or difficult speech is to understand, while intelligibility measures actual understanding (Munro & Derwing, 1995). Measurements of comprehensibility are typically made by having listeners judge samples using rating scales (e.g., Munro & Derwing, 1995; Trofimovich & Isaacs, 2012), while intelligibility often involves participants transcribing utterances (e.g., Munro & Derwing, 1995; Xie & Fowler, 2013). One of the chief differences between the constructs is that comprehensibility captures difficulty in processing speech, since speech that requires a lot of effort to understand will receive lower scores even if the speech is ultimately understood. While the constructs are closely related, then, they are not perfectly aligned, as utterances that are fully intelligible may not be judged as perfectly comprehensible (e.g., Munro & Derwing, 1995). When considering the role of understanding in successful communication between interlocutors, comprehensibility may be the more important measure, as listeners can get frustrated talking with someone whose speech is difficult to understand, even if communication is ultimately successful.

The role of language background in L2 comprehensibility

There is an assumption that L1 background, particularly when L2 speakers share an L1, affects the understanding of L2 speech. Studies investigating whether test-takers are advantaged when sharing the language background of the speaker in listening tests have revealed mixed results, with some language groups showing an advantage, but with little overall effect in some cases and even a disadvantage in other cases (e.g., Harding, 2012; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002; Smith & Bisazza, 1982). Trained test-raters may also rate speakers from their own language background more leniently (e.g., Winke & Gass, 2013). Outside of language testing, most research investigating the role of L1 background in the understanding of L2 speech has employed intelligibility measures, focusing on what Bent and Bradlow (2003) called the interlanguage speech intelligibility benefit (ISIB).

There is a wealth of evidence that L1 speakers generally find L2 less intelligible than L1 speech (e.g., Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Smith & Hayes-Harb, 2009). However, for L2 speakers listening to L1 and L2 speech, the picture is complex. Several studies have found evidence of an ISIB of some type (e.g., Bent & Bradlow; Hayes-Harb et al.; Smith, Bradlow, & Bent, 2003; Xie & Fowler, 2013). However, other studies have revealed weak or negative results (e.g., Algethami, Ingram, & Nguyen, 2011; Stibbard & Lee, 2006), and several investigations have shown mixed results, often with the proficiency of the listeners and speakers affecting the

findings. ISIB studies are difficult to summarize due to the variety of findings and approaches; however, it is fair to say that while there can be an ISIB for L2 speakers, it is uncertain how strong the effect is, and it depends on many factors, such as proficiency, context, and learner background characteristics (Smith & Hayes-Harb, 2011; Xie & Fowler).

The majority of the studies investigating the role of L1 background in understanding L2 speech have been interested primarily in intelligibility rather than comprehensibility, and most used narrow measures of intelligibility, such as identifying individual words or transcribing short sentences. There are good reasons for using narrow measures. Most research on shared intelligibility is focused on the idea that language users with a shared L1 have similar phonological representations, and narrow measures allow for greater control over specific acoustic properties of speech. However, for real-world communication, it is important to understand mutual *comprehensibility* of speech for L1 and L2 speakers, in addition to mutual intelligibility of speech in tightly controlled conditions.

Unfortunately, there has been little research investigating L2 listeners' ratings of L2 comprehensibility (but see Matsuura, 2007; Matsuura, Chiba, Mahoney, & Rilling, 2014). A rare exception is a study targeting both L2 comprehensibility and intelligibility for listeners and speakers from several L2 backgrounds (Munro, Derwing, & Morton, 2006). In that study, listeners with L1 Cantonese, Mandarin, Japanese, or English backgrounds listened to speech samples from Cantonese, Japanese, Polish, and Spanish speakers of English. While the English listeners found all groups equally comprehensible, there were group differences for other listeners. For example, the Cantonese listeners found the speech from their own language background more comprehensible than the speech of Japanese, Polish, and Spanish speakers; Japanese listeners found their L1 group's speech easier to understand than the speech of Cantonese speakers but not by the other speaker groups. However, the effect sizes were small and there was fairly strong agreement across all listeners in their judgments of speech, leading Munro et al. to conclude that the role of the L1 of the listener was less important overall than the linguistic characteristics of the speech itself. This finding was echoed in Crowther, Trofimovich, and Isaacs (2016), who found no differences between Mandarin and French listeners' ratings of French speakers of English.

The role of linguistic variables in L2 comprehensibility

Researchers have increasingly been interested in understanding how different linguistic variables in L2 speech contribute to intelligibility

and comprehensibility, targeting for the most part L1 listeners as raters. For example, some studies have investigated the impact of specific speech characteristics on comprehensibility and intelligibility, including primary stress (Hahn, 2004), speech rate (Munro & Derwing, 2001), and functional load (Munro & Derwing, 2006). Investigating how multiple speech characteristics contribute to comprehensibility ratings, Kang, Rubin, and Pickering (2010) found that prosodic aspects of L2 speech accounted for around 50% of the variance in the comprehensibility ratings.

Recent studies have moved beyond prosody to investigate a wider range of linguistic variables and their relationship to L2 comprehensibility. Isaacs and Trofimovich (2012) had 60 L1 English listeners rate the comprehensibility of 40 L1 French speakers of English. These ratings were then correlated with 19 speech measures related to pronunciation, fluency, grammar, lexis, and discourse, revealing that comprehensibility was linked to all targeted linguistic categories (see also Trofimovich & Isaacs, 2012). Other recent research has looked at how linguistic measures relate to comprehensibility as a function of the speakers' L1 background (Crowther, Trofimovich, Saito, & Isaacs, 2015), task type (Crowther, Trofimovich, Isaacs, & Saito, 2015), and speaker proficiency level (Saito, Trofimovich, & Isaacs, 2017). Although all these variables appear to influence which L2 speech characteristics contribute to comprehensibility ratings, a common finding is that comprehensible L2 speech is linked to several linguistic dimensions of L2 output, including its segmental and suprasegmental content and fluency characteristics (e.g., Kang, 2010; Kang et al., 2010; Saito et al.).

Compared to research focusing on L1 listeners, there has been less work examining which speech characteristics contribute to comprehensibility for L2 listeners (e.g., Field, 2005; Matsuura et al., 2014). Given the number of English listeners who are not L1 speakers, it is important to understand if there are differences in how speech is perceived based on the language background of both speakers and listeners. O'Brien (2014) showed that L2 German listeners' ratings of comprehensibility were predicted by phonetic, lexical, morphological, and fluency measures and by stress placement accuracy. And researchers working within a lingua franca perspective have targeted communication between L2 interlocutors for features of pronunciation that affect intelligibility (e.g., Deterding & Kirkpatrick, 2006; Jenkins, 2002). By way of summary, it appears that L2 listeners may differ from L1 listeners in which aspects of speech they consider most important for comprehensibility; however, some features (e.g., lexical stress) may have a similar impact on intelligibility for both L1 and L2 listeners (e.g., Field). Despite increasing research in this area, it is still

unknown how the linguistic variables that contribute to L2 comprehensibility judgments may differ for L2 listeners and speakers from different combinations of L1 backgrounds. In addition, the studies investigating the role of linguistic variables in speech judgments have taken different approaches to measuring such variables. For example, Kang et al. (2010) and Isaacs and Trofimovich (2012) largely used acoustic measurements and other coded measures, while some of the more recent studies (e.g., Crowther, Trofimovich, Isaacs, & Saito, 2015; Saito et al., 2017) have employed multiple raters to assess speech variables. This approach (also used here and discussed in detail below) has been validated against acoustic measures (Saito et al.) yet may make it difficult to compare findings across prior studies.

Most of the studies investigating the role of L2 background on comprehensibility and intelligibility have been quantitative in nature and have not investigated the actual beliefs that listeners have about how they perceive L2 speech. However, a few studies have asked raters to comment on what they believe underlies their judgments. A study by Jun and Li (2010) investigated how verbal reports by L1 speakers and L2 speakers (L1 = Korean, Russian/Ukrainian, and Chinese) of English differed when completing a task rating L2 speech samples from a variety of L1 backgrounds. They found that L2 and L1 speakers differed in their qualitative responses to L2 speech. L2 speakers were more likely to mention specific linguistic features, and L1 speakers were more likely to mention comprehensibility when rating for accentedness. A recent study by Crowther, Trofimovich, and Isaacs (2016) found that Mandarin and French listeners differed when making qualitative assessments of the comprehensibility of speech samples by French speakers of English, with French listeners commenting on a broader range of linguistic features than the Mandarin listeners. Harding (2008) used qualitative data from L2 listeners to understand their views of having different accents on a listening test. He reported that listeners viewed a shared L1 as a distraction, not an advantage. However, the use of verbal report data for investigating how L2 listeners believe language background impacts comprehensibility is thus far limited.

The current study

In their research targeting mutual understanding of L2 speech for different groups of L2 speakers and listeners, Munro et al. (2006) suggested, but did not directly test, the possibility that variations in comprehensibility ratings might largely reflect the properties of the speech itself. An interesting question, then, is how the language background of L2 listeners may affect the relative importance of various

linguistic dimensions of L2 speech to comprehensibility. Therefore, the chief goal of the current study was to examine whether L2 listeners' language background determines which speech characteristics contribute to their judgments of comprehensibility for L2 speakers from different language backgrounds and whether this benefit is limited to specific linguistic features used by listeners in rating L2 comprehensibility.

Another interesting question is how much listeners themselves attribute L1 background to their understanding of L2 speech when making comprehensibility judgments. There is currently a lack of research investigating L2 speakers' beliefs around the role of language background in understanding L2 speech. Verbal reports allow for a window into what is salient in the mind of the listener making decisions about speech samples and offer an understanding of rating processes that quantitative ratings alone cannot (Gass & Mackey, 2000).

The following research questions were asked:

1. How do pronunciation, prosody, and fluency characteristics of L2 speech (segmental errors, word stress, intonation, speech rate) contribute to L2 listeners' judgments of comprehensibility for L2 speakers from different L1 backgrounds?
2. Is there a comprehensibility benefit beyond what can be attributed to pronunciation, prosody, and fluency characteristics of L2 speech for L2 listeners whose L1 background matches that of L2 speakers?
3. How much do L2 listeners overtly attribute their L2 comprehensibility judgments to the language background of L2 speakers?

The first two questions were addressed through quantitative analyses of speech measures and comprehensibility ratings; the third question was examined through descriptive analyses of listeners' verbal reports. In this study we chose to focus on language backgrounds that are highly dissimilar in terms of both phonology and their roles in a World English context: French, Mandarin, and Hindi.

Method

Participants

Listeners

The listeners were 30 L2 English users recruited from an English-medium university in Montreal, Canada. They came from three L1 backgrounds, with 10 listeners per group: Mandarin, French, and Hindi (seven women, three men per group), referred to as Mandarin-L, French-L, and Hindi-L groups (where "L" stands for "listeners").

Table 1: Background characteristics of listeners

Background variables	Mandarin-L <i>M (SD)</i>	French-L <i>M (SD)</i>	Hindi-L <i>M (SD)</i>	English-L <i>M (SD)</i>
Age (years)	19.8 (1.3)	24.2 (5.6)	23.9 (2.2)	31.5 (11.8)
Percent English use (0–100%)	54.5 (23.2)	56.0 (17.9)	74.5 (7.3)	81.0 (11.0)
Self-rated comprehensibility (1–9)	4.3 (1.4)	6.8 (1.3)	8.0 (1.6)	
Familiarity with Mandarin (1–9)	8.7 (0.5)	3.2 (2.3)	1.0 (0.0)	4.4 (2.8)
Familiarity with French (1–9)	2.1 (1.8)	9.0 (0.0)	3.1 (1.3)	7.6 (1.4)
Familiarity with Hindi (1–9)	1.5 (1.6)	1.6 (0.0)	9.0 (0.0)	3.4 (3.2)

Note: English use (0% = none, 100% = all the time); comprehensibility (1 = very difficult to understand, 9 = very easy to understand); familiarity (1 = not familiar, 9 = very familiar).

Another 10 native English listeners (six women, four men) were recruited for comparison purposes (English-L group). Because most comprehensibility research has used native English listeners, the scores from English listeners could be used as an estimate of the comprehensibility ratings that would be considered typical of other studies and serve as a way to test for pre-existing differences between the groups, specifically with respect to L2 comprehensibility. Because Montreal is a bilingual city, it was expected that a large number of listeners would either be learning or already be proficient in French and that all listeners would be familiar with French-accented English. Four of the Mandarin listeners were learning French, but none spoke it proficiently; five Hindi listeners were learning French, but none were proficient speakers. Among the English listeners, three were learning French, and seven were proficient speakers. As for the other target languages, one of the French listeners was learning Mandarin, but none of the other listeners spoke any Mandarin or Hindi. Aside from three native French speakers raised in Canada, the French listeners had resided in Canada for a mean of 3.3 years (2 months–13 years). The Mandarin listeners had been in Canada for a mean of 10 months (2 months–3.1 years), and the Hindi listeners for 8 months (1–14 months). Table 1 shows the background characteristics of the four listener groups.

Speakers

The 30 target speech stimuli were selected from an unpublished corpus of 143 L2 English speakers completing several speaking tasks (Isaacs & Trofimovich, 2011). The speakers came from three L1 backgrounds: Mandarin (five women, five men), French (four women, six men), and Hindi (one woman, nine men), referred to as Mandarin-S, French-S, and Hindi-S groups (where “S” stands for “speakers”). All speakers had an English proficiency level high enough to be admitted to credit

Table 2: Background characteristics of speakers

Speaker variables	Mandarin-S <i>M (SD)</i>	French-S <i>M (SD)</i>	Hindi-S <i>M (SD)</i>
Age (years)	23.4 (3.1)	20.7 (2.2)	24.0 (1.4)
Time in Canada (years)	0.7 (0.4)	0.5 (0.2)	0.5 (0.2)
English study (years)	10.6 (3.4)	9.9 (3.1)	14.6 (8.7)
Speaking ability (1–10)	5.3 (1.2)	6.1 (1.3)	7.5 (1.0)

Note: Speaking ability (1 = *extremely poor*, 10 = *extremely fluent*).

programs at the university where the study took place. The target samples included the speakers' performance in a TOEFL iBT integrated task, which requires speakers to listen to a short academic lecture, read a short passage on a similar topic, and then answer a question aloud that relates to both the lecture and the reading. This task was chosen because it requires speakers to use an academic register that language learners encounter, and are expected to produce, when studying in a post-secondary context. Two comparable versions of the task were used, with half of the speakers in each L1 group completing one and half completing the other. While it would have been preferable to have all speakers complete the same task, this was not possible, as the corpus we used did not have enough speakers completing the same task to conduct our study; however, the two tasks were on similar social science topics and of comparable difficulty and, thus, elicited similar responses. The speech samples were edited such that only the first 30 seconds of each were used for the ratings. Table 2 shows the background characteristics of the three speaker groups.

We are including some of the data we analyzed as part of the study here to show the equivalence of the groups at the outset of the study. The procedures around these data will be given in more detail in the next section. Ratings given by the comparison group of L1 English listeners to each of the three L2 speaker groups were compared to determine potential pre-existing differences across the L2 speakers. The measures included L1 English listeners' ratings of comprehensibility, segmental errors, word stress, intonation, and speech rate given to each of the three L2 speaker groups (see below). Repeated-measures ANOVAs revealed significant F ratios for comprehensibility, $F(2, 27) = 15.5$, $p < 0.0001$, $\eta_p^2 = 0.54$; segmental errors, $F(2, 27) = 7.42$, $p = 0.003$, $\eta_p^2 = 0.36$; word stress errors, $F(2, 27) = 11.23$, $p = 0.0001$, $\eta_p^2 = 0.46$; intonation, $F(2, 27) = 9.80$, $p = 0.001$, $\eta_p^2 = 0.42$; and speech rate, $F(2, 27) = 10.17$, $p = 0.001$, $\eta_p^2 = 0.43$. Bonferroni-corrected between-group comparisons further revealed that, for all measures apart from word stress errors, the French and Hindi speakers were not significantly different

Table 3: Means and standard deviations for comprehensibility scores and speech measures given by the English listener group

Speaker variables	Mandarin-S <i>M (SD)</i>	French-S <i>M (SD)</i>	Hindi-S <i>M (SD)</i>
Comprehensibility	339 (92)	636 (199)	677 (133)
Segmental errors	287 (58)	418 (103)	443 (120)
Word stress errors	445 (52)	503 (74)	591 (80)
Intonation	393 (54)	545 (76)	529 (112)
Speech rate	374 (138)	545 (124)	621 (112)

Note: Scores range from 1 (low rating) to 1000 (high rating).

from each other, but both were rated significantly higher than the Mandarin speakers. For word stress, the French and Mandarin speakers did not differ significantly from each other, but the Hindi speakers were rated significantly higher than either group. Thus, the Mandarin speaker group was perceived by the L1 English listeners as being less comprehensible and as having more problems with individual sounds, intonation, and speech rate, compared to the other two L2 speaker groups. Table 3 shows descriptive statistics for all ratings given by the L1 English listeners to the three L2 speaker groups.

Speech rating

The listeners were given individual appointments with the researcher. They first rated the 30 target speech samples for comprehensibility. The TOEFL iBT integrated task was explained to the listeners, and they were given a summary of the task to read. They were then instructed that comprehensibility refers to a judgment of listening effort and were warned that the speech samples would be cut off after 30 seconds. The ratings were carried out using a computer-based continuous sliding scale programmed in MATLAB. The scale, which presented the samples in a unique random order for each listener, included two labelled end points, corresponding to the rating of 0 as the left-end point (*hard to understand*) and the rating of 1,000 at the right-end point (*easy to understand*), with no numeric values or intervals labelled. Consistent with prior research (e.g., Saito et al., 2017), listeners were not allowed to replay each file but could proceed to the next sample at their own pace. Listeners were also instructed to make a verbal report after each rating, explaining their reasons behind the comprehensibility ratings for each speaker; these reports were audio-recorded. To minimize researcher bias, the listeners were not given any examples of the types of explanations they could give but were encouraged to speak more if they provided little information during the practice phase.

After the comprehensibility ratings were completed, the listeners proceeded to rate the speech samples again, this time for several speech measures (including degree of foreign accent, which was not analyzed further) using similar 0–1,000 continuous sliding scales:

1. Segmental errors, which refer to vowel and consonant errors (e.g., substituting /d/ for /ð/ in “that”), with 0 corresponding to *frequent* and 1,000 to *infrequent or absent*.
2. Word stress errors, which refer to errors in the placement of stress on words that contain more than one syllable (e.g., saying “COMputer” instead of “comPUter”), with 0 corresponding to *frequent* and 1,000 to *infrequent or absent*.
3. Intonation, which applies to utterances longer than a single word and refers to the expected pitch contours associated with spoken utterances (e.g., using falling pitch to indicate a complete declarative utterance), with 0 corresponding to *unnatural* and 1,000 to *natural*.
4. Speech rate, which refers to how quickly or slowly someone speaks, with 0 corresponding to *too slow or too fast* and 1,000 to *optimal*.

Because previous research suggests there are no overall rating condition effects (i.e., rating speech dimensions separately vs. together within the same set) on L2 speech ratings (O’Brien, 2016), these four ratings were carried out in the same session. The listeners received a handout with explanations and examples of each measure, and the researcher clarified any remaining questions. The procedure was similar to that used for the judgments of comprehensibility, with the exception that there were several scales on the screen and the listeners could hear each speech sample as many times as needed to be confident in their judgments. These types of ratings for specific linguistic variables show strong associations with other linguistic measures (see Saito et al., 2017, for validation of these rated measures against coded analyses of speech). Therefore, ratings were considered an acceptable way to measure the four speech categories, especially because all speech measures, including those carried out by trained coders (e.g., phoneme accuracy or stress placement judgments) and those involving acoustic measurements (e.g., measurements of vowel duration or formant frequencies), involve subjective decisions. For instance, when judging stress placement, one coder may hear misplaced stress, while another thinks stress placement is correct, and two researchers measuring vowel duration may not arrive at the exact same result. Prior to starting the ratings, the listeners were given four practice files (two from each task) using speakers from L1 backgrounds not used in this study. The listeners rated the speech samples using high-quality headphones, and the researcher remained in the room while they

completed the task, but sat at a computer facing away from the listeners in order to minimize their potential discomfort at being overheard.

Analysis

The overall agreement among the L2 listeners, calculated using two-way random intra-class correlations, was high for all five rated measures: comprehensibility ($\alpha = 0.97$), segmental errors ($\alpha = 0.90$), word stress errors ($\alpha = 0.89$), intonation ($\alpha = 0.94$), and speech rate ($\alpha = 0.95$). Therefore, the 10 listeners' ratings within each listener group were averaged to derive a single mean rating per speaker. Pearson correlations and hierarchical regressions were used to investigate (a) relationships between the comprehensibility scores and the four speech measures and (b) contributions of the speech measures to comprehensibility ratings for the L2 listener and speaker groups.

The verbal reports were transcribed and coded using 40 categories related to different aspects of comprehensibility (e.g., pronunciation, grammar, etc.). However, the current analysis centred only on the comments related to listeners' own language background (L1-Own) or other language backgrounds (L1-Other). Whenever listeners mentioned the L1 background of the speaker, the comment was coded as L1-Own if listeners were commenting about a speaker from their own language background and as L1-Other if they were commenting about a speaker from a different language group. The coding was completed by the first author, and a subset of eight transcripts was independently coded by the second author to ensure that the coding was reliable. There was strong agreement between the first and second coders, and while there were a few differences in coding that had to be resolved through discussion, none of them related to the codes used for this study. In some cases, comments were coded even if an explicit reference to the L1 was not made, provided there was a clear reference to the rating being related to a specific L1. For example, the comment "it's easier to understand 'cause I I've hear this accent before" (Hindi-L 23) was coded as L1-Other even though this listener did not explicitly mention that the speaker was French.

Once the comments were extracted, they were further coded for whether the listener found the L1 of the speaker to be positive, neutral, or negative in terms of its relevance to comprehensibility. A comment was coded as positive even if the listener found the speaker difficult to understand, provided that he or she saw the L1 as beneficial, as in "The speaker 26 is pretty hard to understand but since I'm French I get it a bit" (French-L 8). Similarly, a comment was coded as negative, even though the speech sample was considered easy to understand, if the speaker indicated that the L1 was harmful to comprehensibility.

Comments showing no clear attitudes toward the L1 were given a neutral label, as in “I think um she was again he’s uh Indian and uh but it’s a little different from the previous one, uh because um his vocab and speak some words more clearly” (Mandarin-L 6).

Results

Speech measures and L2 comprehensibility

The first analysis targeted the first research question, namely, the extent to which segmental errors, word stress, intonation, and speech rate relate to L2 listeners’ judgments of comprehensibility for L2 speakers from different L1 backgrounds. To address this question, Pearson correlations were run to compare the relationship between each listener group’s comprehensibility ratings and the same group’s judgments of the four speech measures for each of the three speaker groups (Bonferroni-corrected $\alpha = 0.001$). Differences in which correlations are significant for the combinations of L2 groups would indicate that a listener’s language background plays a role in determining which aspects of speech are most important for comprehensibility. As summarized in Table 4, for the Mandarin speakers, the comprehensibility ratings given by each of the listener groups were not associated with these groups’ ratings of the four speech measures. When rating the French speakers, the French listeners’ comprehensibility rating

Table 4: Pearson correlations between comprehensibility scores and speech measures for each listener–speaker group

Speaker group/rating	Mandarin-L	French-L	Hindi-L
Mandarin-S			
Segmental errors	0.23 ^a	0.47	0.77
Word stress errors	0.15 ^a	0.06	0.80
Intonation	0.67 ^a	0.29	0.50
Speech rate	0.81 ^a	0.52	0.80
French-S			
Segmental errors	0.79	0.90 ^{*a}	0.90 [*]
Word stress errors	0.86 [*]	0.72 ^a	0.89 [*]
Intonation	0.95 [*]	0.78 ^a	0.90 [*]
Speech rate	0.93 [*]	0.84 ^{*a}	0.80
Hindi-S			
Segmental errors	0.86 [*]	0.82	0.83 ^{*a}
Word stress errors	0.90 [*]	0.82	0.82 ^a
Intonation	0.90 [*]	0.74	0.80 ^a
Speech rate	0.77	0.86 [*]	0.95 ^{*a}

^{*} $p < 0.001$ (one-tailed).

^aShared L1 background for speakers and listeners.

was linked to two speech measures (segmental errors, speech rate); the Mandarin listeners' comprehensibility rating was associated with three speech measures (word stress errors, intonation, and speech rate); and the Hindi listeners' comprehensibility rating was correlated with three speech measures (segmental errors, word stress errors, and intonation). Finally, when rating the Hindi speakers, the Hindi listeners' comprehensibility rating was correlated with two speech measures (segmental errors, speech rate); the Mandarin listeners' comprehensibility rating was associated with three speech measures (segmental errors, word stress errors, and intonation); and the French listeners' comprehensibility rating was linked to a single speech measure (speech rate). Thus, whenever significant associations were detected, there was a partial (yet far from perfect) overlap between the speech variables associated with L2 comprehensibility for each of the three listener–speaker group combinations.

Shared background benefit

The next analysis targeted the second research question, that is, the extent to which shared L1 background contributed to the variance in L2 comprehensibility scores given by L2 listeners to L2 speakers after accounting for the contribution of the four speech measures. For this analysis, three hierarchical multiple regressions were run (one per L2 listener group), with comprehensibility rating as the outcome variable. In each regression, the four speech measures (segmentals, word stress, intonation, speech rate) were entered into the regression model in Step 1 in order to estimate the amount of variance in comprehensibility accounted for by the characteristics of L2 speech as rated by a given listener group. Then, language background, coded as a dummy categorical variable with the matching listener–speaker background as a reference group, was added in Step 2 in order to determine the unique contribution of language background to L2 comprehensibility judgments by each listener group. In each case, if language background significantly improved the fit of the model, it would suggest that that language background is contributing to comprehensibility in a way that is not captured in typical speech measures.

For the Mandarin listeners, speech rate and word stress errors together accounted for 73% of the total variance in the comprehensibility scores given to the three L2 speaker groups (see Table 5 for full regression results). Adding language background in Step 2 produced a 6% improvement in the model fit. The Mandarin listeners downgraded both the French speakers relative to the Mandarin speakers, $t = -2.65$, $p = 0.014$, and the Hindi speakers relative to the Mandarin speakers, $t = -2.21$, $p = 0.037$. The Mandarin listeners rated the Mandarin speakers'

Table 5: Hierarchical regression results for Mandarin listeners

Variable	Model 1			Model 2		
	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β
Segmentals	0.51	0.28	0.25	0.26	0.32	0.13
Word stress	0.82	0.41	0.34*	0.78	0.37	0.32*
Intonation	-0.05	0.40	-0.03	0.49	0.45	0.27
Speech rate	0.73	0.25	0.50**	0.61	0.23	0.42*
Mandarin vs. French				-0.12	0.05	-0.29*
Mandarin vs. Hindi				-0.14	0.06	-0.31*
R^2		0.73			0.79	
R^2 change		0.77			0.06	
<i>F</i>	(4, 25)	20.99***		(2, 23)	3.99**	

Note: Language background was entered as three dummy variables, with Mandarin serving as the reference group.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Hierarchical regression results for French listeners

Variable	Model 1			Model 2		
	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β
Segmentals	0.73	0.24	0.43*	0.65	0.29*	0.38*
Word stress	-0.05	0.29	-0.03	0.07	0.33	0.04
Intonation	0.30	0.25	0.21	0.18	0.26	0.13
Speech rate	0.45	0.20	0.37*	0.39	0.23	0.32
French vs. Hindi				-0.04	0.06	-0.08
French vs. Mandarin				-0.10	0.06	-0.21
R^2		0.82			0.83	
R^2 change		0.85			0.02	
<i>F</i>	(4, 25)	34.14***		(2, 23)	1.56	

Note: Language background was entered as three dummy variables, with French serving as the reference group.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

comprehensibility 124 points higher compared to the French speakers, and 136 points higher compared to the Hindi speakers.

For the French listeners, segmental errors and speech rate together accounted for 82% of the total variance in the comprehensibility scores given to the three L2 speaker groups. Adding language background in Step 2 did not produce a significant improvement in model fit (see Table 6).

For the Hindi listeners, speech rate alone accounted for 86% of the total variance in comprehensibility scores given to the three L2

Table 7: Hierarchical regression results for Hindi listeners

Variable	Model 1			Model 2		
	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β
Segmentals	0.31	0.28	0.20	0.25	0.27	0.16
Word stress	0.38	0.35	0.22	0.54	0.33	0.31
Intonation	-0.09	0.33	-0.05	-0.07	0.32	-0.04
Speech rate	0.75	0.19	0.63**	0.52	0.21	0.44
Hindi vs. French				-0.02	0.04	-0.04
Hindi vs. Mandarin				-0.12	0.06	-0.23*
R^2		0.86			0.88	
R^2 change		0.88			0.02	
<i>F</i>	(4, 25)	46.22		(2, 23)	2.69	

Note: Language background was entered as three dummy variables, with Hindi serving as the reference group.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

speaker groups (see Table 7). Again, adding language background in Step 2 did not produce an overall significant improvement in model fit. Nevertheless, the Hindi listeners showed a significant tendency to downgrade the Mandarin speakers relative to the Hindi speakers, $t = -2.10$, $p = 0.047$, rating the comprehensibility for the speakers of their own L1 background by 119 points higher.

Verbal reports about language background

The last analysis focused on the final research question, which concerned the degree to which L2 listeners overtly attribute their comprehensibility judgments to the language background of L2 speakers. Analyses of verbal reports indicated that the French listeners overall made the most comments relating to the language backgrounds of the speakers (55 comments from eight raters), followed by the Mandarin listeners (28 comments from nine raters), with the Hindi listeners reporting far fewer comments (six comments from three raters). The lower number of comments for some groups, compared to others, is not reflected in the larger data set. With all 40 coded categories included, the percentage of each group's total comments in the study was roughly similar (Mandarin-L = 24%, French-L = 25%, Hindi-L = 23%, English-L = 29%). And within each listener group, not all listeners commented on language background, with more Mandarin listeners (Own-L1 = 6; Other-L1 = 6) and French listeners (Own-L1 = 2; Other-L1 = 6) providing comments, compared to the Hindi listeners (Own-L1 = 1; Other-L1 = 2).

Both the Mandarin and the French listeners made positive comments far more often when talking about their own L1 backgrounds,

Table 8: Frequencies of the L1-Own category comments for L2 listener groups

Comments	Mandarin-L	French-L	Hindi-L	All L2 listeners
Positive	13	20	1	34
Neutral	1	3	0	4
Negative	2	3	1	6

with 82% of the Mandarin listeners' comments and 77% of the French listeners' comments being positive (see Table 8). For example, a French listener noted, "Uh I have not difficulties to understand him but uh I'm I'm think, I think it's a French so it's easy to me to to understand with his uh with his accent. Uh he has a French pronunciation of the words so I so I understand uh what he said" (French-L 20). There were also comments from three listeners suggesting that a speaker would be hard to understand for those who did not share his or her L1, as in "He's Chinese so I can understand what, what he's talking about. But maybe for some other people who is not familiar with Chinese accent, with Chinese like logic, maybe it's a little bit hard to understand" (Mandarin-L 5). The negative comments about listeners' own background often reflected difficulties the listeners attributed to speakers from their language background. For example, a Mandarin listener commented, "I guess he's a Chinese because his pronunciation is not very very good" (Mandarin-L 17). As there were only two Own-L1 comments in the Hindi-L group, it is difficult to generalize for this group.

In terms of talking about other L1 backgrounds (see Table 9), all three L2 listener groups made far more negative than positive comments (Mandarin-L = 75%, French-L = 75%, Hindi-L = 100%). The French listeners' negative comments were divided fairly evenly between the Mandarin (42%) and Hindi speakers (58%). However, both the Mandarin and Hindi listeners made far fewer negative comments about the French speaker group, compared to the other groups (22% and 25%, respectively). Not all negative reports indicated that the listeners struggled to understand the speaker. For example, one French listener commented, "The speech was understandable even though the Indian accent was present" (French-L 24). However, many of the comments did indicate that the listeners struggled to understand the speech because of the speaker's L1. For example, a Hindi listener reported, "Speaker number 14's uh comprehensibility was hard to understand since he had a Chinese accent for sure" (Hindi-L 16).

Table 9: Frequencies of the L1-Other category comments for L2 listener groups

Mandarin-S	Mandarin-L	French-L	Hindi-L	All L2 listeners
Positive	0	0	0	0
Neutral	0	0	0	0
Negative	0	10	3	13
French-S				
Positive	1	0	0	1
Neutral	0	0	0	0
Negative	2	0	1	3
Hindi-S				
Positive	0	1	0	1
Neutral	2	2	0	4
Negative	7	14	0	21
All L2 speakers				
Positive	1	1	0	2
Neutral	2	2	0	4
Negative	9	24	4	37

Discussion

The current study examined (a) how pronunciation, prosody, and fluency characteristics of L2 speech contribute to L2 listeners' judgments of comprehensibility for L2 speakers from different L1 backgrounds, (b) whether there exists a comprehensibility benefit beyond what can be attributed to segmental, prosodic, and fluency characteristics of L2 speech, and (c) whether L2 listeners attribute their comprehensibility judgments to the language background of L2 speakers. With respect to the relationship between L2 comprehensibility and the four speech measures, there were differences in which speech measures were significantly related to comprehensibility judgments for different combinations of L2 listeners and speakers. No speech measures were linked to the Mandarin speakers' comprehensibility scores for any of the listener groups, and different clusters of speech measures were associated with the comprehensibility scores given by the listeners to the other speaker groups.

In terms of the effect of shared language background on L2 comprehensibility, after the variance from the four speech measures was accounted for, language background explained an additional 6% of unique variance in comprehensibility ratings for the Mandarin listeners. That is, the Mandarin listeners downgraded both the French and the Hindi speakers, relative to the Mandarin speakers. For the French and Hindi listeners, language background did not significantly contribute to the comprehensibility scores, although the Hindi listeners

tended to downgrade the Mandarin speakers relative to the Hindi speakers. In verbal reports, the L2 listeners overtly considered L1 background as a factor in rating L2 comprehensibility. This happened more frequently for the Mandarin and French listeners, who made more comments related to the L1s of the speakers, than for the Hindi listeners, who made fewer. The L2 listeners were also more likely to make positive comments about L1 background when sharing an L1 with the speaker, and to make negative comments when the L1s did not match.

Speech variables and listener–speaker language background

There were clear L1 background effects with respect to the relationship between listeners' judgments of L2 comprehensibility and listener-rated pronunciation, prosody, and fluency characteristics in L2 speakers' output (see Table 4). There was a partial (yet far from complete) overlap in the speech variables associated with ratings of L2 comprehensibility across the three listener groups. This was particularly true for L2 speakers of French and Hindi, where a mismatch in language background between listeners and speakers was characterized by either a narrower or a wider range of speech measures associated with L2 comprehensibility, relative to the shared listener–speaker L1 background. For instance, for the French listeners evaluating the French speakers, segmental errors and speech rate were the two speech measures associated with L2 comprehensibility. However, for the Mandarin and Hindi listeners, the French speakers' comprehensibility was linked to only one of these two measures, and these listener groups associated the French speakers' comprehensibility with two additional variables, which did not seem to be relevant to the French listeners. If ISIB effects arise due to shared phonological knowledge between listeners and speakers (e.g., Hayes-Harb et al., 2008), then shared comprehensibility benefits likely stem from a similar knowledge overlap. With no common language background available, listeners likely resort to the linguistic cues that appear most relevant for comprehension.

The Mandarin speakers were the only group for which no speech measure correlated with comprehensibility ratings given by any listener group. However, compared to the French and the Hindi speakers, the Mandarin speakers were also rated the weakest in four out of five speech ratings they received from the English listeners (Table 3), and the ratings given to the Mandarin speakers by the other L2 listeners generally featured lower values, compared to the ratings given to other speakers (Table 4). Thus, it may have been difficult for the listeners to detect any salient relationships between speech variables and L2 comprehensibility for the (low-proficiency) Mandarin speakers,

particularly if listeners overall struggled with understanding these speakers. These results align well with findings from research showing that the linguistic variables that contribute to comprehensibility can vary based on several factors, including speakers' L1 (Crowther, Trofimovich, Saito, & Isaacs, 2015), speaking task being targeted (Crowther, Trofimovich, Isaacs, & Saito, 2015), and speakers' proficiency level (Saito et al., 2017).

Although the current data set revealed clear L1 background effects for L2 listeners evaluating L2 speech, such that the pronunciation, prosody, and fluency characteristics of L2 output relevant to comprehensibility may be specific to a particular listener–speaker combination, the current findings also implied some similarities. For example, as shown in Table 4, for the French and Hindi listener–speaker combinations, the two speech measures most strongly correlated with comprehensibility were identical (segmental errors, speech rate), and these same measures featured the largest number of significant associations across all listener groups (four, in each case). It may be that, regardless of listener–speaker combinations, some aspects of segmental accuracy and some characteristics of speech fluency, such as speech rate, might be generally relevant to L2 listener perception of comprehensibility. If some speech dimensions feeding into listener judgments of L2 comprehensibility, such as segmental accuracy and speech rate, are similar across various L2 speaker–listener combinations, then these dimensions could serve as instructional foci in language classrooms composed of learners with different language backgrounds who will be using their L2 with different interlocutors.

Benefits of shared L1 background

Findings from ISIB research indicate that the benefit of a shared L1 can be moderated by proficiency (e.g., Hayes-Harb et al., 2008; Xie & Fowler, 2013). The current results are consistent with these findings, in that the Mandarin listeners enjoyed the greatest benefit from having a shared L1 background with the speakers whose oral ability level was relatively low. Although an independent measure of L2 speaking proficiency (e.g., TOEFL score) for the listeners was not available, the Mandarin listeners also self-reported the lowest comprehensibility ratings of the three groups (see Table 1). Thus, listeners who share an L1 background with lower proficiency speakers may perceive that speech as easier to understand than the speech of speakers from other L1 backgrounds. This has implications for learners from the same L1 background who are studying in a common L2, such that these learners may have an inflated sense of how comprehensible each other's speech will sound to interlocutors who do not share their language

background (see Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016, for evidence of low-proficiency L2 listeners overestimating their comprehensibility).

One interesting question is what exactly brought about a unique comprehensibility benefit from shared listener–speaker L1 background, beyond what was explained by shared understanding of the four speech measures. Because this study targeted extemporaneous speech, it is likely that additional contributions to comprehensibility stemmed from aspects of L2 speech not captured in this study. For example, linguistic dimensions such as grammar, vocabulary, and discourse richness contribute to comprehensibility (e.g., Crowther, Trofimovich, Saito, & Isaacs, 2015; Saito et al., 2017). While ISIB research is typically interested only in specific phonological aspects of a shared interlanguage, other speech commonalities that are part of a shared language background could be important to comprehensibility in real-world interaction. There is also evidence that factors unrelated to speech can play a role in speech judgments, such as bias and listener expectations (e.g., Kang & Rubin, 2009). For example, Kang (2012) found that variance in students' proficiency ratings of international teaching assistants could be explained partially by measures of prosody, but that the listeners' native-speaker status and other background variables (e.g., experience with L2 speakers) also contributed to the ratings.

While it is interesting to consider what could account for the variance explained by the shared language background for the Mandarin participants, the effects of a shared language background in this study were very small, appearing for only one language group (Mandarin) and explaining only an additional 6% of the variance. These findings, when combined with ISIB-related studies showing only minimal evidence of shared language-background effects, suggest that it may be time for research to move past looking for a shared intelligibility or comprehensibility benefit and instead focus on how specific aspects of speech may contribute to comprehensibility for different combinations of listener–speaker groups.

Verbal reports as a window on shared L1 benefits

The verbal report data only partially supported what was found in the regression analyses. The Mandarin listeners made several comments related to an improved understanding of L2 speech from their own language background and a decreased understanding of speakers from other backgrounds, which was reflected in the regression model for the Mandarin listeners. However, language background had no predictive power in regression models for the French listeners, yet the

French listeners made the most comments about the ease with which they could understand speech from their own L1 background. The Mandarin and French listeners also made negative comments about understanding the Hindi speakers, which was not reflected in the quantitative analysis. Thus, a relatively poor agreement between the results of quantitative analyses of speech ratings and descriptive coding of verbal reports suggests that what listeners report may not be what actually influences their speech ratings. Hayes-Harb and Hacking (2015) investigated reasons underlying L1 English listeners' accent judgments of L2 English speakers. They found that the raters frequently strayed from simply analyzing the speech to "describe the social-cultural backgrounds of the speakers" (p. 62). At least in this data set, judging from overall high proportions of variance in comprehensibility explained by speech measures (73–86%), it is most likely that the characteristics of the speech which made a particular speaker easy or difficult to understand were captured mostly through the speech measures.

Limitations and conclusion

There are several limitations that should be addressed. First, the sample size was limited, making correlation and regression analyses exploratory in nature. With more participants, future research could more easily compare different listener–speaker groups. Further, tighter controls over participant proficiency would allow for a clearer understanding of the role of proficiency in comprehensibility judgments. Finally, a more explicit set of instructions accompanying verbal reports may have yielded clearer findings. The lack of direction given to the listeners when making verbal reports was intentional and was done to prevent researcher bias. However, interviews or focus groups may provide more appropriate measures when looking for specific information that listeners consider in L2 speech.

As English continues to spread around the globe, with an increasing number of English users being L2 speakers, it is important that research investigating comprehensibility and intelligibility reflect this complexity. This study found that different L1 listener–speaker combinations resulted in overlapping yet non-identical linguistic variables contributing to comprehensibility ratings. Further, comprehensibility ratings were uniquely associated with a matching listener–speaker L1 background benefit, at least for the speakers with lower L2 comprehensibility. Finally, listeners often attributed the language background of the speakers to the ease or difficulty they experienced when rating speakers, though this was not always reflected in the actual speech

ratings. The entire concept of having speech that is easy or difficult to understand is multifaceted, comprising qualities of speech and characteristics of the listener, not to mention many variables relating to the context of a given interaction. While it is impossible for any study to capture the full range of these complexities, the closer we come to understanding comprehensibility in a wide range of contexts, the more potential there will be to help language learners communicate successfully in whichever setting they plan to use their L2.

Correspondence should be addressed to **Jennifer A. Foote**, Faculty of Extension, University of Alberta, 1-024 Enterprise Square, 10230 Jasper Ave., Edmonton, AB T5J 4P6; e-mail: jfoote@ualberta.ca.

Acknowledgements

This research was supported by the Social Sciences and Humanities Research Council of Canada. We would like to thank our participants for their involvement in this study. We would also like to thank our anonymous reviewers who provided valuable feedback and improved the final version immensely.

References

- Algethami, G., Ingram, J., & Nguyen, T. (2011). The interlanguage speech intelligibility benefit: The case of Arabic accented English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2010* (pp. 30–42). Ames, IA: Iowa State University.
- Bent, T., & Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114(3), 1600–1610. <https://doi.org/10.1121/1.1603234>
- 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>
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, 99(1), 80–95. <https://doi.org/10.1111/modl.12185>
- Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, 49(4), 814–837. <https://doi.org/10.1002/tesq.203>
- Deterding, D., & Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes*, 25(3–4), 391–409. <https://doi.org/10.1111/j.1467-971X.2006.00478.x>
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423. <https://doi.org/10.2307/3588487>

- Gass, S.M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. <https://doi.org/10.2307/3588378>
- Harding, L. (2008). Accent and academic listening assessment: A study of test-taker perceptions. *Melbourne Papers in Language Testing*, 13(1), 1–33.
- Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, 29(2), 163–180. <https://doi.org/10.1177/0265532211421161>
- Hayes-Harb, R., & Hacking, J.F. (2015). Beyond rating data: What do listeners believe underlies their accentedness judgments? *Journal of Second Language Pronunciation*, 1(1), 43–64. <https://doi.org/10.1075/jslp.1.1.02hay>
- 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(4), 664–679. <https://doi.org/10.1016/j.wocn.2008.04.002>
- Isaacs, T., & Trofimovich, P. (2011). International students at Canadian universities: Validating a pedagogically-oriented pronunciation scale. Unpublished corpus of second language speech.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83–103. <https://doi.org/10.1093/applin/23.1.83>
- 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.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. <https://doi.org/10.1016/j.system.2010.01.005>
- 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. <https://doi.org/10.1080/15434303.2011.642631>
- Kang, O., & Rubin, D. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28(4), 441–456. <https://doi.org/10.1177/0261927X09341950>
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English.

- Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Major, R.C., Fitzmaurice, S.F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36(2), 173–190. <https://doi.org/10.2307/3588329>
- Matsuura, H. (2007). Intelligibility and individual learner differences in the EIL context. *System*, 35(3), 293–304. <https://doi.org/10.1016/j.system.2007.03.003>
- Matsuura, H., Chiba, R., Mahoney, S., & Rilling, S. (2014). Accent and speech rate effects in English as a lingua franca. *System*, 46, 143–150. <https://doi.org/10.1016/j.system.2014.07.015>
- Munro, M.J., & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- 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(4), 451–468.
- Munro, M.J., & Derwing, T.M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531. <https://doi.org/10.1016/j.system.2006.09.004>
- Munro, M.J., Derwing, T.M., & Morton, S.L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <https://doi.org/10.1017/S0272263106060049>
- O'Brien, M.G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64(4), 715–748. <https://doi.org/10.1111/lang.12082>
- O'Brien, M.G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, 38(3), 587–605. <https://doi.org/10.1017/S0272263115000418>
- 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(4), 439–462. <https://doi.org/10.1093/applin/amv047>
- Smith, B.L., Bradlow, A.R., & Bent, T. (2003). Production and perception of temporal contrasts in foreign-accented English. In M.J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona, Spain* (pp. 519–522). Barcelona, Spain: Causal Productions.
- Smith, B.L., & Hayes-Harb, R. (2009). Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of the Acoustical Society of America*, 125(4), 2761. <https://doi.org/10.1121/1.4808694>

- 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(1), 115–120. <https://doi.org/10.1016/j.wocn.2010.11.005>
- Smith, L.E., & Bisazza, J.A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning*, 32(2), 259–269. <https://doi.org/10.1111/j.1467-1770.1982.tb00971.x>
- Stibbard, R.M., & Lee, J.I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *Journal of the Acoustical Society of America*, 120(1), 433–442. <https://doi.org/10.1121/1.2203595>
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905–916. <https://doi.org/10.1017/S1366728912000168>
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(1), 122–140. <https://doi.org/10.1017/S1366728914000832>
- Winke, P., & Gass, S. (2013). The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, 47(4), 762–789. <https://doi.org/10.1002/tesq.73>
- 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(5), 369–378. <https://doi.org/10.1016/j.wocn.2013.06.003>