

JOURNAL ARTICLE

Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech

Ann R. Bradlow, Michael Blasingame and Kyounghee Lee

Department of Linguistics, Northwestern University, Evanston, Illinois, US

Corresponding author: Ann R. Bradlow (abradlow@northwestern.edu)

The present study provides evidence of a positive correlation between L1 and L2 intelligibility for bilingual talkers. Each talker in a group of Mandarin-English and Korean-English bilinguals was recorded producing simple sentences in each of their languages. The recordings were then presented to native listeners of the language-being-spoken in a test of sentence-in-noise recognition to yield a pair of L1 and L2 intelligibility scores for each talker based on the proportion of words correctly recognized. For all talkers, their L2 speech (i.e., Mandarin- or Korean-accented English) was consistently lower in intelligibility than their L1 speech (i.e., native accented Mandarin or Korean). This dissociation between L1-L2 speech intelligibility was evident in different intelligibility scores at comparable signal-to-noise ratios (as much as 35 percentage points for the Mandarin-English bilinguals) and in the required boost in signal-to-noise ratio to achieve comparable intelligibility (>8–10 dB). Critically, L1 and L2 intelligibility scores within the bilingual talkers were strongly positively correlated: Talkers with relatively high intelligibility in L1 also achieved relatively high intelligibility in L2. These results indicate a persistent influence of talker-specific trait characteristics that combine with, rather than are overwhelmed by, language-specific and dominance-dependent influences in bilingual speech production.

Keywords: speech intelligibility; speech production; bilingualism

1. Introduction

Every spoken utterance conveys linguistic information (what is being said) and indexical information specifying the context in which the sentence should be understood (who is speaking, where and when the utterance is spoken). Decades of acoustic-phonetic analysis of speech signals and experimental studies of speech perception have provided strong evidence for extensive overlap between the acoustic parameters that convey linguistic information and those that convey indexical information. For example, listeners can still identify individual talkers even after information about the vocal source has been removed from the signal (i.e., removal of a traditionally-viewed acoustic correlate of indexical information), as in sine-wave speech (Remez, Fellowes, & Rubin, 1997). And, salient talker-specific variation (i.e., indexical information) is evident in the voice onset time (VOT) cue to phonological voicing, a phonetic timing cue whose production is independent of the size and shape of the talker's vocal tract (Allen, Miller, & DeSteno, 2003; Theodore, Miller, & DeSteno, 2009). Thus, current evidence strongly indicates that linguistic and indexical information are jointly encoded in the multitude of acoustic dimensions that constitute the acoustic speech signal.

In monolingual talkers, talker-specific indexical information and language-specific linguistic information are simultaneously conveyed by speech in a single language.

For bilingual talkers, however, two separate (but interacting) linguistic systems must be actively controlled. This then raises the question of the relationship between talker-specificity and language-specificity in bilingual talkers. Is talker-specific variation tightly constrained by the phonological system of the language being spoken such that, to the extent that talker variation is under talker control, individual talker differences manifest themselves differently in speech production in each of the two languages of a bilingual individual? Or, do individual talkers adopt global, idiosyncratic speech production strategies, or styles, that are evident in their speech regardless of the language being spoken? And, how do the general limitations of speech production in a second-language (L2) versus speech production in the first-language (L1), interact with the constraints of both language-specificity and talker-specificity?

In a prior study (Bradlow, Kim, & Blasingame, 2017), we laid out a strategy for assessing language-independent talker-specificity in bilingual speech production. First, given that acoustic cues to linguistically meaningful contrasts are necessarily constrained by language-specific phonological structure, language-independent talker-specificity is expected to be most salient in global acoustic features (e.g., speaking rate, pitch range). Unlike sub-segmental, segmental, or supra-segmental features that convey phonemic contrasts, global acoustic features are set independently of the phonetic realization of particular phonological categories, and are therefore directly comparable across pairs of languages. Second, in order to establish language-independent talker-specificity beyond features that are a direct consequence of vocal tract anatomy and physiology, it is important to establish status-specific variation across L1 and L2 speech (i.e., variation in the status of the language being spoken as either first acquired, L1, or later acquired, L2) for the parameter in question within individual bilingual talkers. If average L1 and L2 values along a given global parameter are consistent within bilinguals then we cannot eliminate the possibility that variation along the parameter in question is determined entirely by physical features of the talker's vocal tract. Of interest in this research is the extent to which talker-specific traits that are under talker control (i.e., those traits that are not automatic consequences of vocal tract size and shape) are evident across both languages of bilingual individuals. This is not to say that automatic, physically-determined sources of individual variation in speech production are not of interest for understanding L1-L2 interactions in bilinguals, particularly as they relate to identification of strategies and interventions for speech therapy or other technologies and applications that involve talker identification or speech enhancement. However, the present study seeks a novel demonstration of language-independent individual differences at a higher (central/cognitive) level suggesting that individual talker differences may not always be tightly constrained by the phonological system of the language being spoken. Thus, the signature pattern that would establish language-independent talker specificity beyond vocal tract shape and function is a *dissociation* between L1 and L2 in absolute terms, thereby establishing non-automaticity for the given parameter, paired with an *association* of L1 and L2 in relative terms, thereby establishing a language-independent talker-specific trait.

This dissociation/association pattern is illustrated in **Figure 1**. The left panel of this figure, (a), illustrates the predicted pattern of dissociation (a significant difference) between the average L1 and L2 values for a hypothetical speech parameter as measured across a group of talkers in each language. This dissociation in absolute terms demonstrates L1 versus L2 status-specificity by establishing that the parameter in question can vary as a function of factors that are independent of vocal tract size and shape. The right panel, (b), illustrates the predicted pattern of association (a positive correlation) between L1 and L2 values within bilingual individuals on a hypothetical speech parameter (same parameter

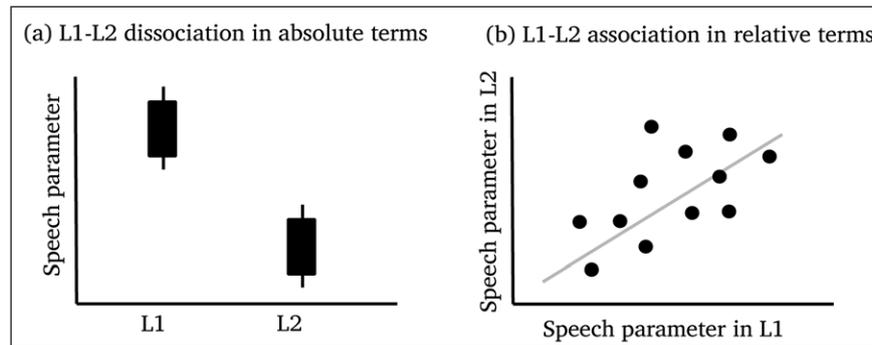


Figure 1: The predicted dissociation/association pattern (hypothetical data).

as in [a]), thereby demonstrating language-independent talker-specificity (i.e., a talker with a relatively high or low value in L1 will also have a relatively high or low value in L2, respectively). Note that in (a), the average values along the given speech parameter can be compared either across different languages within individuals (e.g., L1 Korean versus L2 English within Korean-English bilinguals), or across L1 and L2 speakers of a given language (e.g., L1 English by monolinguals versus L2 English by bilinguals). The association pattern in (b) is necessarily always a within-participant analysis (i.e., each data point represents an individual talker).

Exactly this dissociation/association pattern was observed in our prior work on speaking rate in a group of bilingual talkers ($n = 86$) from various L1 backgrounds ($n = 10$) and with English as their L2 (Bradlow et al., 2017). Using automatically extracted speaking rate measurements (syllables/second) from read and spontaneous speech recordings, this study replicated the well-known and reliable dissociation between L1 and L2 speaking rates in absolute terms (e.g., Guion, Flege, Liu, & Yeni-Komshian, 2000; Baese-Berk & Morrill, 2015, and many others), with L2 speaking rates being significantly slower than L1 speaking rates both within bilinguals (across their L1 and L2) and within L1 versus L2 English (across the L2-English of the bilinguals and the L1-English of a group of monolinguals). This L1-L2 status-specific speaking rate dissociation indicated that the global feature of average speaking rate is under talker control, rather than an automatic consequence of vocal tract structure and function. Critically, this study then established an association between L1 and L2 speaking rate within the bilingual group. L1 speaking rate significantly predicted L2 speaking rate within individuals, indicating a strong role for language-independent talker-specificity in average speaking rate. Relatively fast or slow talkers in L1 were also relatively fast or slow talkers in L2, respectively. (For a similar finding regarding L1 fluency as a predictor of L2 fluency as measured by number of filled pauses, number and duration of silent pauses, and speaking rate, see Towell & Dewaele, 2005; Derwing, Munro, Thomson, & Rossiter, 2009; De Jong, Groenhout, Schoonen, & Hulstijn, 2013).

In the present study, we seek to extend our investigation of language-independent talker-specificity in bilingual speech production to overall talker intelligibility as assessed by the proportion of words correctly recognized by L1 listeners of the language being spoken. Specifically, in this study we ask if L1 intelligibility is a significant predictor of L2 intelligibility. That is, are relatively high or low intelligibility talkers in L1 also relatively high or low intelligibility talkers in L2, respectively? Because (as discussed more below) variation in overall intelligibility is related to a complex combination of articulatory settings that constitute a particular speaking style, this investigation probes language-independent talker-specificity at a level of speech production control that goes beyond a specific acoustic-phonetic cue (such as speaking rate) and instead focuses attention at a deeper level of planning where various language-specific parameters can be set.

Variation in overall intelligibility and its acoustic correlates has been extensively examined in a large literature on context-induced modifications of speech (see Cooke, King, Garnier, & Aubanel, 2014 for a recent review), including the well-studied phenomenon of listener-oriented clear speech (see Uchanski, 2005; Smiljanic & Bradlow, 2009; Smiljanic, forthcoming, for extensive reviews of the clear speech literature). A general conclusion of this work is that variation in overall intelligibility is related to a wide range of acoustic features: Cooke et al. (2014) list 46 such speech modifications. While this line of work has generally focused on the acoustic correlates of variation in intelligibility across speaking styles within individual talkers, the same features have also been implicated as acoustic correlates of variation in intelligibility across individual talkers (e.g., Bond & Moore, 1994; Bradlow, Torretta, & Pisoni, 1996; Hazan & Markham, 2004). This research on both intra- and inter-talker variability in intelligibility has pointed to both global and segmental acoustic features as significant determiners of overall speech intelligibility. Global features of intelligibility-enhanced speech include (but are not limited to) decreased speaking rate, wider dynamic pitch range, and increased energy in the mid-frequency range of long-term spectra, all of which result in greater acoustic salience (audibility and resistance to noise-related interference). Moreover, these global acoustic enhancements are independent of the phonological structure of the language-being-spoken, and therefore are prime candidates for language-independent talker-specificity in bilinguals.

There is also extensive evidence for a wide range of segmental correlates of variation in speech intelligibility, including vowel space expansion and various consonant contrast enhancements in the spectral and temporal domains (e.g., Ferguson & Kewley-Port, 2002, 2007; Smiljanic & Bradlow, 2008; Tuomainena & Hazan, 2016, and many others). Such segmental correlates of overall speech intelligibility are necessarily implemented in the context of the phonological system of the language being spoken. For example, conversational versus clear speech comparisons of voice onset time (VOT) in English and Croatian (Smiljanic & Bradlow, 2008) showed greater increased VOT for the voiceless category in English but increased pre-voicing for the voiced category in Croatian (for a similar comparison of VOT enhancement in English and Finnish clear speech, see Granlund, Hazan, & Baker, 2011). This cross-language difference in clear speech production strategy nevertheless reveals cross-language consistency with respect to the overarching goal of phonological cue enhancement since in both languages the clear speech production strategy results in a greater acoustic difference between the underlying phonological voicing category contrasts (for a detailed and extensive review of this general point see Smiljanic & Bradlow, 2009, pages 246–250, and references therein). While particular phonological enhancements are inherently language-dependent and therefore not readily transferrable across the two languages of an individual bilingual talker, the general principle of contrast enhancement is language-general and is therefore viable as an individual-level articulatory setting that may be evident in both L1 and L2 speech in bilingual talkers.

In the present study, we hypothesize that talker-specificity manifests in bilingual individuals, at least in part, by a language-independent speech production setting that is related to variation in overall intelligibility. If individual talker variation is at least partially determined by a language-independent talker-specific global speech production setting, then within a group of bilingual talkers, those who are relatively high intelligibility talkers in their L1 will also be relatively high intelligibility talkers in their L2 (association in relative terms). Importantly, this positive correlation should be evident despite consistently lower L2 than L1 intelligibility (e.g., Munro & Derwing, 1995; Flege, Munro, & MacKay, 1995; Derwing & Munro, 1997, and many others). That is, we expect to find the same pattern of absolute dissociation/relative association for overall intelligibility that we found for

speaking rate (Bradlow et al., 2017) and as schematized in **Figure 1** above. This would indicate that the L1-L2 intelligibility association is not a direct consequence of automatic inter-talker variations that are determined by factors related to vocal tract anatomy and physiology. Our interest in this line of research is instead in a level of individual talker variation that is more central/cognitive than peripheral/physical in origin, that is, at a level at which talker-specificity, status-specificity (L1 versus L2), and language-specificity may be mutually modulating.

The notion of a language-independent talker-specific global speech production setting that transcends both (or all) languages of bilingual (or multilingual) talkers is related to, but also distinct from, variation along the hyper-to-hypo-articulation continuum that is at the heart of the H&H theory of phonetic variation (Lindblom, 1990). While the H&H theory explains phonetic variation as a direct consequence of the adaptive nature of speech production, the present research hypothesizes that such adaptation may operate within a persistent (i.e., non-adaptive) setting. The adaptive framework of H&H theory promotes an account of phonetic variation that hinges on the notion of ‘sufficient discriminability,’ a notion that crucially involves a balance between talker- and listener-oriented constraints. H&H theory claims that the phonetic encoding of linguistically meaningful contrasts must be discriminable by a given listener in a given communicative setting, but that the talker need only expend enough articulatory effort to ensure sufficient discriminability for that listener in that setting. The present study hypothesizes that this intra-talker adaptive variation is nested within inter-talker variation that manifests as a talker-specific setting, or window, along the hyper-to-hypo-articulation continuum within which adaptive variation presumably operates. The case of bilingual talkers that is at the center of the present research provides a window into this layer of individual-level variation with non-adaptive origins.

2. Methods

2.1. Overview

We tested overall L1 and L2 intelligibility in two groups of bilingual talkers: a group of Mandarin-English bilinguals and a group of Korean-English bilinguals.¹ Each group of talkers was recorded reading a set of sentences in both their L1 and L2. The recorded sentences were then presented to L1 listeners of each language mixed with speech-shaped noise to mimic the non-optimal listening conditions of typical, real-world speech recognition and to avoid ceiling effects that might compress inter-talker variation. Intelligibility scores in terms of the proportion of words correctly recognized were then assessed for each talker in each language yielding a pair of L1 and L2 intelligibility scores for each individual bilingual talker. These paired scores were then analyzed according to the absolute dissociation/relative association framework described above and as illustrated in **Figure 1**.

¹ Data from the two groups of participants were collected at separate times and locations for reasons having to do with the planning and development of procedures for compilation of the larger corpus of which this study is a part. Specifically, the Korean-English dataset preceded the Mandarin-English dataset. In the intervening time, our plans to develop the larger corpus of L1 and L2 recordings from bilinguals from various L1 backgrounds—the ALLSTAR Corpus as described in greater detail in Bradlow, Kim, and Blasingame, 2017—took increasingly well-specified shape. This situation accounts for some of the methodological differences between the two datasets reported in this paper, including test location (USA for L1 Mandarin intelligibility versus Korea for L1 Korean intelligibility) and signal-to-noise ratios (two conditions (−4 dB and −8 dB) for L1 Mandarin intelligibility versus one condition (−5 dB) for L1 Korean intelligibility). Nevertheless, because the critical comparisons for this study are within each group (relationship between L1 and L2 intelligibility within bilinguals), these relatively minor differences do not undercut interpretation of the data.

2.2. Talkers

Fourteen Mandarin-English bilinguals (11 males, 3 females; mean age 23 years) were recruited to record sentences both in their L1, Mandarin, and in their L2, English. The Mandarin talkers reported having no history of speech or hearing impairments at the time of recording. Their overall L2 proficiency ranged from 50 to 69 (mean: 59; *SD*: 5.9) on the Versant English Test (Pearson, 2009).² They were all educated through university (up to their undergraduate degree) in Mandarin, and were recorded at Northwestern University (Evanston, Illinois, USA) within one month of arrival in the USA before starting graduate programs at the university.

Ten Korean-English bilinguals (5 males, 5 females; mean age 27 years) were recruited to record sentences in both their L1, Korean, and in their L2, English. The Korean talkers reported having no history of speech or hearing impairments at the time of recording. Their overall L2 proficiency ranged from 34 to 77 (mean: 55.2; *SD*: 11.55) on the Versant English Test (Pearson, 2009). They were all born in Korea and spoke standard Seoul Korean as their first language, and were educated through university (up to their undergraduate degree) in Korean. The Korean group was recorded at Northwestern University. The length of stay for the Korean-English bilinguals in English-speaking countries ranged from 0.1 to 3.5 years with a mean of 1.6 years. All talkers (both Mandarin and Korean) were paid \$10 per hour for their participation.

2.3. Listeners

Native listeners of Mandarin were recruited for the L1 Mandarin speech intelligibility test. None of these listeners had previously participated as talkers for the stimuli recordings. All listeners reported having no history of speech or hearing impairments at the time of intelligibility testing. These Mandarin listeners were members of the Northwestern community who grew up in a Mandarin-speaking country and were educated through undergraduate studies in China or Taiwan. A total of 52 Mandarin listeners (31 males, 21 females, mean age 25 years) participated in the L1 Mandarin intelligibility test, with approximately half of the subjects in each of two signal-to-noise ratio conditions, -4 dB ($n = 27$) and -8 dB ($n = 25$). The Mandarin intelligibility testing was conducted in the Linguistics Department at Northwestern University.

Native listeners of Korean were recruited for the L1 Korean speech intelligibility test. None of these listeners had previously participated as talkers for the stimuli recordings. All listeners reported having no history of speech or hearing impairments at the time of intelligibility testing. These Korean listeners ($n = 20$, 12 males, 8 females; mean age 21 years) were all monolingual undergraduates at Hanyang University (Seoul, Korea). All Korean listeners spoke standard Seoul Korean as their first language and had no experience of living abroad except for four participants who had spent less than six months in English-speaking countries (average: 0.6 months). The Korean intelligibility testing included a single signal-to-noise ratio condition (-5 dB), and was conducted in the Department of Language and Literature at Hanyang University. All Mandarin and Korean listeners were paid for their participation.

Native listeners of English were recruited for both the Mandarin-accented L2 English and Korean-accented L2 English intelligibility tests. Participants were Northwestern University undergraduates who were raised and educated in the USA. None of the participants had any extended previous experience with Mandarin or Korean. Fifty-seven native English

² The range of possible scores on the Versant test is 20 to 80. Scores in the 50–75 range are considered lower-intermediate to advanced proficiency in English. For comparison, note that graduate students at Northwestern University are required to achieve a Versant score of 65 in order to be eligible to serve as a teaching assistant for an undergraduate course.

participants (22 males, 35 females; mean age 20 years) were tested on Mandarin-accented English sentences distributed over two signal-to-noise ratio conditions, 0 dB ($n = 30$) and -4 dB ($n = 27$). Forty native English participants (16 males, 24 females; mean age 19 years) were tested on Korean-accented English sentences distributed over two signal-to-noise ratio conditions, 5 dB ($n = 20$) and 0 dB ($n = 20$). All native English participants were given course credit for their participation.

2.4. Stimuli

For the test of L1 and L2 speech intelligibility by Mandarin-English bilinguals, 112 Mandarin and 112 English sentences were taken from published lists (from the Hearing in Noise Test, HINT; Soli & Wong, 2008). All sentences in both languages were single-clause declarative sentences containing approximately 3–5 content words with appropriate function words in each language (e.g., “The mother heard the baby” in English; “我十分钟后在门口等你” [I will wait for you by the door in ten minutes] in Mandarin).

For the test of L1 and L2 speech intelligibility by Korean-English bilinguals, 100 English sentences were selected from the revised Bamford-Kowal-Bench Standard Sentence Test (BKB-R; Bamford & Wilson, 1979). Like the HINT sentences used for the test with Mandarin-English bilingual talkers, these BKB-R sentences are single clause declarative sentences containing 3 or 4 keywords with several function words (e.g., “The teapot is very hot”). A parallel set of Korean sentences was created for this study by a native Korean speaker (KL) (e.g., “그들은 시계를 보고 있다” [They are looking at the clock] in Korean). English sentences were selected for translation into Korean based on (1) KL’s native speaker judgments of the familiarity of words to non-native Korean speakers, (2) the intelligibility scores of each list reported in Bamford and Wilson (1979), and (3) their translatability into Korean sentences. These selected sentences were translated into Korean, and the translation was then checked and proofread by two additional native Korean speakers in order to ensure translation accuracy and consistency. **Table 1** summarizes the distribution of talkers, sentences, listeners, and signal-to-noise ratios across all conditions in all intelligibility tests.

2.5. Procedures

2.5.1. Recording

The Mandarin-English bilinguals were all recorded as part of a larger corpus of L1 and L2 speech, the ALLSSTAR Corpus³ (Bradlow et al., 2011). Each language (Mandarin and

Table 1: Distribution of talkers, sentences, and listeners across the four intelligibility tests.

| Test condition | Talker L1 | Language stimuli | Listener L1 | Signal-to-noise ratio (dB) |
|----------------|-----------------------|------------------------|-----------------------|----------------------------|
| (1) | Mandarin ($n = 14$) | Mandarin ($n = 112$) | Mandarin ($n = 27$) | -4 dB |
| | | | Mandarin ($n = 25$) | -8 dB |
| (2) | | English ($n = 112$) | English ($n = 30$) | 0 dB |
| | | | English ($n = 27$) | -4 dB |
| (3) | Korean ($n = 10$) | Korean ($n = 100$) | Korean ($n = 20$) | -5 dB |
| (4) | | English ($n = 100$) | English ($n = 20$) | $+5$ dB |
| | | | English ($n = 20$) | 0 dB |

³ Available upon request via the Online Speech/Corpora Archive and Analysis Resource (OSCAAR; <https://oscaar.ci.northwestern.edu/#>).

English) was recorded on a separate day. The Korean-English bilinguals (who were not part of the ALLSTAR Corpus project) were recorded producing the English sentences first and then had a 10-minute break before recording the Korean sentences. All recording sessions were conducted in a sound-attenuated booth using a Shure SM81 Condenser Handheld microphone. The digital speech files were segmented into individual sentence files, and 500 milliseconds of silence were added to both ends of each individual file using Praat (Boersma, 2001; Boersma & Weenink, 2017). The individual sentence stimulus files were all leveled to equate root mean-square (RMS) amplitude across the full set.

2.5.2. Intelligibility testing

As illustrated in **Table 1**, four separate sentence recognition tests were run: (1) native Mandarin listeners with native-accented Mandarin sentences (2 conditions with 2 different signal-to-noise ratios, -4 dB and -8 dB), (2) English listeners with Mandarin-accented English sentences (2 conditions with 2 different signal-to-noise ratios, 0 dB and -4 dB), (3) native Korean listeners with native-accented Korean sentences (1 signal-to-noise ratio condition at -5 dB), and (4) English listeners with Korean-accented English sentences (2 conditions with 2 different signal-to-noise ratios, $+5$ dB and 0 dB).

For the assessment of Mandarin L1 intelligibility (test 1), native Mandarin listeners completed a sentence recognition task in which they were asked to listen to and repeat as accurately as possible the native-accented Mandarin sentences recorded by the 14 Mandarin-English bilingual talkers. One group of native Mandarin listeners completed the task at -4 dB signal-to-noise ratio ($n = 27$), and a separate group of native Mandarin listeners was tested at -8 dB signal-to-noise ratio ($n = 25$). Similarly, for the assessment of Mandarin-English L2 intelligibility (test 2), native English listeners completed the sentence recognition task (same oral response format as for the L1 intelligibility assessment in test 1) with the Mandarin-accented English sentences recorded by the same Mandarin-English bilingual talkers. One group of native English listeners heard sentences at 0 dB signal-to-noise ratio ($n = 30$), and another at -4 dB signal-to-noise ratio ($n = 27$). These particular signal-to-noise ratios were selected based on previous work in our laboratory with both L1 and L2 speech intelligibility (e.g., Bradlow & Bent, 2002; Smiljanic & Bradlow, 2005, 2011) from which we have determined that, in order to achieve similar levels of speech recognition accuracy, L2 listeners require an approximately 4–5 dB signal-to-noise ratio advantage over L1 listeners when presented with L1 speech, and that L2 speech is similarly disadvantaged relative to L1 speech when presented to L1 listeners. By covering a range of signal-to-noise ratios for both the L1 and L2 intelligibility tests with these Mandarin-English bilinguals, we aimed to identify the boost in signal-to-noise ratio required to bring L2 speech intelligibility up to the level of L1 speech intelligibility. In particular, these data allow us to assess the effect on intelligibility of an 8 dB boost to L2 speech (L1 at -8 dB vs. L2 at 0 dB) as well as a 4 dB boost to L2 speech (L1 at -4 dB vs L2 at 0 dB, and L1 at -8 dB vs L2 at -4 dB). In addition, by including an overlapping signal-to-noise ratio (i.e., -4 dB), we can also compare L1 and L2 intelligibility under identical listening conditions. In this way, we can obtain converging evidence for the overall lower intelligibility of L2 speech relative to L1 speech when presented to L1 listeners.

Each listener in the tests involving the Mandarin-English bilingual talkers (tests 1 and 2 above) heard 112 sentences (14 talkers \times 8 sentences). The sentence stimuli were randomized and presented one at a time with MaxMSP software over Sony MDRV700 headphones in a sound-attenuated booth. During stimulus presentation, MaxMSP mixed the sentences in speech-shaped noise (-4 or -8 dB signal-to-noise ratio for the two conditions of the Mandarin L1 test; 0 or -4 dB signal-to-noise ratio for the two conditions of the Mandarin-accented L2 English test). Each listener heard each instance of a sentence

only as produced by a single talker, but over the course of the entire experiment including all listeners, all 14 talkers had approximately equal exposure for all of their sentences. Listeners' oral responses to the sentence recognition task were recorded as audio files. The Mandarin-English component of this study (test 1 and 2) opted for a spoken response over a typed or written response, as there were input and legibility concerns for the Mandarin responses.

For the assessment of Korean L1 intelligibility (test 3), native Korean listeners completed a sentence recognition task in which they were asked to listen to and write down the Korean sentences as recorded by the 10 Korean-English talkers. For practical testing time and listener recruitment reasons, only one signal-to-noise ratio condition was run for this test (-5 dB signal-to-noise ratio, $n = 20$ listeners) and this test was conducted at Hanyang University in Korea. For the assessment of Korean-English L2 intelligibility (test 4), native English listeners completed the sentence recognition task with the Korean-accented English sentences recorded by the same Korean-English bilingual talkers (same written response format as for the L1 intelligibility assessment in test 3). The Korean-accented English listener group was split in two with one group listening to sentences at +5 dB signal-to-noise ratio, and the other at 0 dB signal-to-noise ratio ($n = 20$ per group). As noted above, our previous work (e.g., Bradlow & Bent, 2002; Smiljanic & Bradlow, 2005, 2011) has indicated that L2 listeners require an approximately 4–5 dB signal-to-noise ratio advantage over L1 listeners to achieve similar levels of speech recognition accuracy and that L2 speech is similarly disadvantaged in terms of intelligibility when presented to L1 listeners. As for the test of Mandarin-English bilinguals' speech discussed above, we included two signal-to-noise ratios for the L2 intelligibility test with the Korean-English bilinguals in an effort to identify the boost in signal-to-noise ratio required to bring L2 speech intelligibility up to the level of L1 speech intelligibility. In particular, these data allow us to assess the effect on intelligibility of boosts of 5 dB (L1 at -5 dB vs. L2 at 0 dB) and 10 dB to L2 speech (L1 at -5 dB vs. L2 at +5 dB). Moreover, the L2 Korean-accented English test at 0 dB signal-to-noise ratio overlapped with one of the L2 Mandarin-accented English L2 tests, and the L1 Korean test at -5 dB signal-to-noise ratio was very closely comparable to the L1 Mandarin tests at -4 dB signal-to-noise ratio. One hundred English sentences and 100 Korean sentences each produced by the 10 Korean-English bilingual talkers were used in the Korean-English sentence recognition task for each respective native listener group (tests 3 and 4). In order to ensure that every listener in each of the two language conditions heard every talker's speech in that language, we constructed 10 stimulus lists for each language, each of which consisted of 100 sentences (10 different sentences \times 10 talkers), containing no overlapping sentences across talkers in a similar fashion to the Mandarin L1 and Mandarin-accented L2 English tests described above. Participants listened to each sentence played over headphones and typed in their response on a computer keyboard. Stimulus presentation was controlled with the software Max/MSP, which allowed for online mixing of the speech and noise at the specified signal-to-noise ratio.

2.6. Data scoring

For the tests involving Mandarin speech (test condition [1] in **Table 1** above), the audio response files from the listeners were scored using the audio software tool MaxMSP, which displayed the text of each sentence along with the audio file (the listeners' responses). Each word was given a score of correct (1) or incorrect (0) by an L1 Mandarin scorer. No distinction was made between keywords and non-keywords for the Mandarin scoring procedure as determining word status in Mandarin is not straightforward (Sproat, Shih, Gale, & Chang, 1996). In order to assure reliable scoring, a subset of scores was compared against those of a second L1 Mandarin scorer. Average scorer agreement was 97.2 percent.

While the typical scoring procedure for tests of English intelligibility in both research and clinical settings relies on measures of keyword recognition (i.e., recognition of content words to the exclusion of intervening grammatical function words), for the test of Mandarin-accented L2 English intelligibility (test condition [2] in **Table 1** above) we opted for a scoring system that matched that of the test of L1 Mandarin intelligibility (test condition [1] in **Table 1** above). That is, we counted both content word and function word recognition accuracy. The rationale behind this choice was that the primary comparison of interest in this study was the within-talker L1-L2 comparison. We therefore wanted a consistent scoring criterion for the L1 and the L2 of the Mandarin-English bilingual talkers. For these tests of Mandarin-accented English, a word was counted as correctly recognized (score of 1) if it completely matched with the sentence transcript from which the talker read during the recording, including all affixes (e.g., plural “s” and past tensed “ed”). In order to assess scoring reliability, responses from each English listener in the 0 dB signal-to-noise ratio condition were scored by two scorers. Inter-scorer agreement was 98 percent, which we deemed high enough to rely on only a single scorer for all listener responses in the -4 dB signal-to-noise ratio condition. As a check on the scoring method adopted for this test, we compared the all-word scoring method and a keyword-only scoring method. Correlation between the two scoring methods was very high ($r = 0.94$, averaged over the two signal-to-noise ratios).

For the test involving Korean speech (test condition [3] in **Table 1** above), the typed response files from the listeners were scored automatically with a Perl script that compared keywords in each participant’s sentence transcriptions with the keywords in the sentences that were read by the talkers. Similarly, the English responses (typed format) in the test of Korean-accented English intelligibility (test condition [4] in **Table 1** above) were also scored automatically with an English-version of the Perl script. By checking keyword for keyword in each sentence, the Perl script identified all responses in which the typed words did not exactly match the target keywords. These ‘incorrect’ words were then checked by hand to see whether the mismatch was due to a recognition error (e.g., “strawberry” for “strawberries” or “cleans” for the word “cleaned”) or a spelling error (e.g., “potatos” for “potatoes”). The former was counted as incorrect, whereas the latter was counted as correct.

For all talkers (Mandarin-English and Korean-English bilinguals) in both of their languages (L1 of Mandarin or Korean, and L2 of English for all talkers), an overall intelligibility score in terms of average percent words correctly recognized was then determined in each language at each of the signal-to-noise ratios tested.

2.7. Acoustic analyses

While the primary focus of the present study was on overall intelligibility as an index of a language-independent talker-specific global speech production setting, we also include analyses of a set of acoustic parameters that are related to vocal source characteristics, specifically fundamental frequency (F0) average and range, and the slope of the long term average speech spectrum (LTASS) in the mid-frequency range. The purpose of these analyses was to investigate a set of parameters that are quite closely associated with talker specificity but that, in contrast to speaking rate (the focus of our previous study) and overall intelligibility (the focus of the present study), are more directly related to vocal anatomy and physiology. That is, while speaking rate and overall intelligibility are closely related to patterns of speech articulation and vary significantly depending on the language being spoken (i.e., on the L1 or L2 language ‘state’), these source features (F0 mean, F0 range, and LTASS slope) are more likely to remain relatively impervious to the influence of language-specific and dominance-dependent ‘state’ characteristics in bilingual speech.

Thus, we can begin to delineate talker-specific traits that are both language- and state (L1 vs. L2)-independent (e.g., F0 average, F0 range, and LTASS slope) from talker-specific traits that are language-independent but state-specific (e.g., speaking rate and overall intelligibility).⁴ All of these source-related characteristics were measured automatically from the digital speech files using Praat scripts. F0 mean and range were measured across the full set of sentences in each language by each talker. LTASS slope was measured as the change (drop) in energy from the 0- to 1-kHz range to the 1- to 4-kHz range, a measure that is related to vocal effort.

3. Results

3.1. Overview of main patterns of results

Figure 2 displays the intelligibility scores and acoustic measurements for all talkers in each of their two languages. For each parameter, two plots are shown, a boxplot comparison of the L1 and L2 values and a scatterplot relating the L1 and L2 values within individual talkers. In each of the scatterplots, the filled and open symbols are for the talkers in the Mandarin-English and Korean-English bilingual talker groups, respectively. The top-left quadrant shows the data from the intelligibility tests, the bottom-left quadrant shows

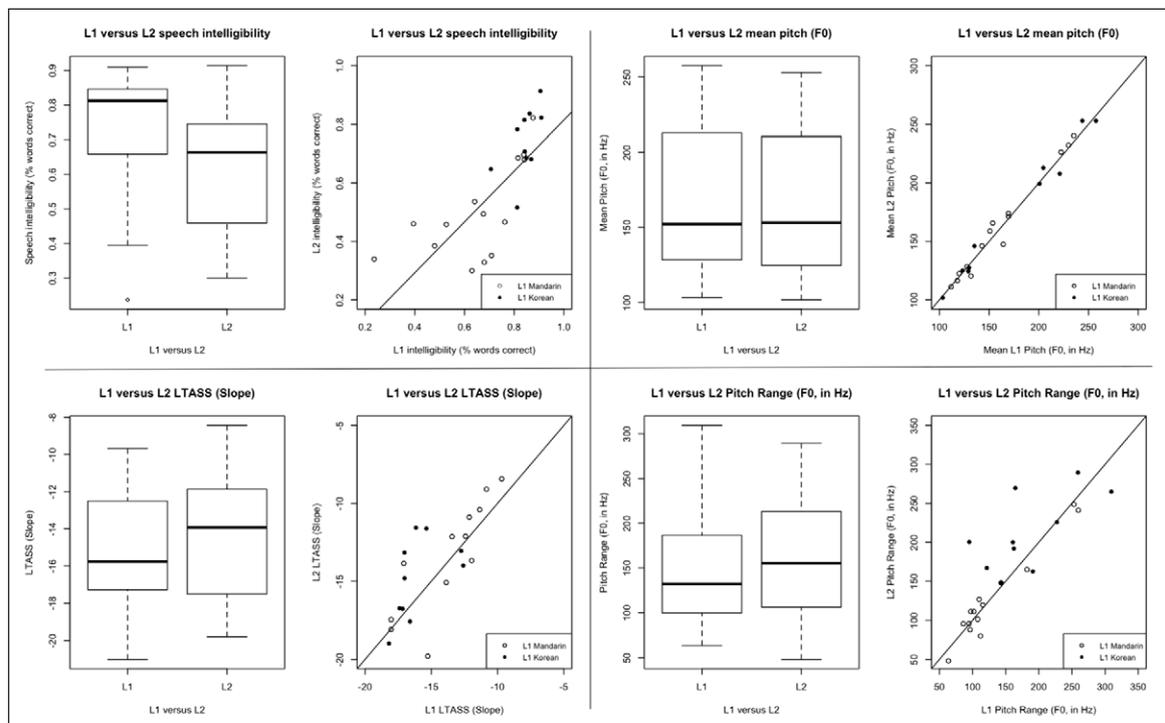


Figure 2: Intelligibility scores and acoustic measurements for all talkers in each of their two languages. L2 intelligibility scores are from higher signal-to-noise ratio test conditions relative to L1 test conditions (boost of 4 dB and 5 dB for the Mandarin-English and Korean-English groups, respectively). Lines on the scatterplots reflect $y = x$.

⁴ Note that the relation between speaking rate and intelligibility is complex rather than straightforward (for extensive discussion of this point see Uchanski, 2005; Smiljanic & Bradlow, 2009; Smiljanic, forthcoming, and the references therein). While clear/high-intelligibility speech is typically (though not always) slower than conversational/low-intelligibility speech, speaking rate variation alone does not account for intelligibility variation. Moreover, many of the acoustic-phonetic characteristics of clear speech are not rate-related. Indeed, across all materials in the present study, the correlation between speaking rate (automatically detected acoustic syllables/second) and word recognition accuracy is non-significant ($r = 0.13$). By focusing on variation in intelligibility as indexed by native listener recognition accuracy, the present study compares L1 and L2 speech samples within and across bilingual talkers in terms of the combination of rate-related and rate-independent articulatory-acoustic features that account for variation in intelligibility.

the LTASS slope data, the top-right quadrant shows the F0 mean data, and the bottom-right quadrant shows the F0 range data. Note that the L2 intelligibility scores shown in **Figure 2** are from test conditions in which the L2 productions were tested at a higher signal-to-noise ratio (boost of 4 dB and 5 dB for the Mandarin-English and Korean-English groups, respectively). Specifically, for the Mandarin-English bilingual group, the L1 and L2 intelligibility scores displayed in the figure are from the tests with signal-to-noise ratios of -8 dB and -4 dB, respectively, and for the Korean-English bilingual group, the L1 and L2 intelligibility scores displayed in the figure are from the tests with signal-to-noise ratios of -5 dB and 0 dB, respectively. (Data from all signal-to-noise ratio conditions are presented in **Table 2** and discussed in detail below.)

3.2. Intelligibility data

The intelligibility data (**Figure 2**, top-left) follow the hypothetical absolute dissociation/relative association pattern (see **Figure 1** above) quite closely. On average, the L2 intelligibility scores are lower than the L1 intelligibility scores even with the signal-to-noise ratio boost of 4–5 dB for the L2 speech. In addition, the adjacent scatterplot indicates a positive correlation between these L1 and L2 intelligibility scores. This pattern suggests that, while L1 intelligibility exceeds L2 intelligibility (demonstrating L1 vs. L2 status-specificity), talkers who are relatively high intelligibility in their L1 are also relatively high intelligibility in their L2 (demonstrating language-independent talker-specificity).

Table 2 shows summary statistics for the full set of intelligibility data as measured in all four of the tests represented in **Table 1**. Specifically, **Table 2** shows the Mandarin-English bilinguals' L1-Mandarin intelligibility as assessed by L1 Mandarin listeners at -8 dB and -4 dB signal-to-noise ratio, and their L2-English (Mandarin-accented English) intelligibility as assessed by English listeners at 0 dB and -4 dB signal-to-noise ratio. **Table 2** also shows the Korean-English bilinguals' L1-Korean intelligibility as assessed by L1 Korean listeners at -5 dB signal-to-noise ratio, and their L2-English (Korean-accented English) intelligibility as assessed by English listeners at 0 dB and +5 dB signal-to-noise ratio. All scores in **Table 2** represent averages (with standard deviations in parentheses) over all talkers, listeners, and sentences.

Statistical analyses of these intelligibility data show a clear dissociation of L1 and L2 intelligibility for both groups of bilinguals. Note that for all statistical analyses, intelligibility scores were expressed in log odds scores to account for the non-linearity inherent in intelligibility differences across the full range of possible scores (i.e., the effect of an increase in word recognition accuracy depends upon where that increase occurs, e.g., an increase of 10 percentage points from 85% is not equivalent to a 10 percentage point increase from 55%). Under consistent listening conditions of -4 dB signal-to-noise

Table 2: Overall intelligibility scores for all talkers in both their L1 and L2 at each of the signal-to-noise ratios (SNRs) that were tested. Numbers represent average word recognition accuracy across all talkers, sentences, and listeners in each condition. Standard deviations are shown in parentheses.

| | L1 intelligibility | | L2 intelligibility | | |
|-----------------------------|--------------------|-----------------|--------------------|-----------------|----------------|
| | -8 dB SNR | -4 dB SNR | -4 dB SNR | 0 dB SNR | N/A |
| Mandarin-English bilinguals | 65.1% (18.6) | 85.6% (12.7) | 50.0% (16.3) | 73.4% (12.6) | N/A |
| Korean-English bilinguals | N/A | -5 dB SNR | N/A | 0 dB SNR | +5 dB SNR |
| | N/A | 84.1% (5.8) | N/A | 74.1% (11.5) | 78.1% (8.1) |

ratio, the L1 Mandarin speech and L2 English speech of the Mandarin-English bilingual talkers differed by approximately 35 percentage points, a highly significant difference (average of 50% versus average of 85.6%, $t(13) = 11.38, p < .01$). A boost of 4 dB to the L2 English speech of these Mandarin-English bilingual talkers (signal-to-noise ratio of 0 dB) was insufficient to reach the intelligibility level of their L1 Mandarin speech at -4 dB signal-to-noise ratio (average of 73.4% versus average of 85.6%, $t(13) = 5.15, p < .01$). A drop of 4 dB to the L1 Mandarin speech of these Mandarin-English bilingual talkers (signal-to-noise ratio of -8 dB) resulted in L1 Mandarin intelligibility scores that were lower than the L2 English intelligibility scores at 0 dB signal-to-noise ratio (average of 65.1% versus average of 73.4%, $t(13) = -2.75, p < .05$), but higher than the L2 English intelligibility scores at -4 dB signal-to-noise ratio (average of 65.1% versus average of 50%, $t(13) = 4.16, p < .01$). Thus, these data indicate that, in these relatively unfavorable (but realistic) listening conditions, a boost of somewhere between 4 dB and 8 dB is required to bring L2 speech into the intelligibility range of L1 speech.

Similarly, for the Korean-English bilinguals, L1 Korean intelligibility in the -5 dB signal-to-noise ratio was higher than L2 English intelligibility in the 0 dB signal-to-noise ratio condition (average of 84.1% versus average of 74.1%, $t(9) = 3.7, p < .01$), indicating that even with a 5 dB difference in signal-to-noise ratio, the L2 English of Korean-English bilinguals was substantially less intelligible than L1 Korean. A boost of 10 dB to the L2 speech was also insufficient in order to reach comparable levels of intelligibility across their two languages (average of 84.1% versus average of 78.1%, $t(9) = 3.03, p < .05$). Taken together, these data demonstrate, as expected and as evident in **Figure 2** and **Table 2**, a clear dissociation within the two groups of bilinguals between L1 and L2 intelligibility scores in absolute terms.

In order to analyze the effects on L2 (English) intelligibility, we built a linear model with fixed effects for L1 intelligibility and native language (Korean or Mandarin, contrast coded as 0.5 and -0.5, respectively), and their interaction. The dependent variable, L2 intelligibility score, was transformed to log-odds space ($\log((p)/(1-p))$) where p is the averaged word recognition accuracy of each talker in each signal-to-noise ratio condition and language. In this analysis we focus on L1 Mandarin and L1 Korean intelligibility as assessed in the -8 dB and -5 dB signal-to-noise ratio conditions, respectively; and on L2 Mandarin-accented English and L2 Korean-accented English intelligibility from the -4 dB and 0 dB signal-to-noise ratio conditions, respectively (matching the data shown in **Figure 2** above). These conditions are the more difficult listening conditions in each case and therefore exhibited greater variance than their more favorable signal-to-noise ratio counterpart conditions. Critically, each data point represents a L1, L2 intelligibility pair for a given talker.

Results of the linear modeling revealed a significant main effect of L1 intelligibility on L2 intelligibility ($B = .8, SE B = .23, t(20) = 3.56, p < .01$). That is, for any given talker, if his or her speech intelligibility in the L1 was relatively high, then his or her speech intelligibility in the L2 was also predicted to be relatively high. There was no significant main effect of L1 (Korean or Mandarin) on L2 intelligibility ($B = .1, SE B = .33, t(20) = .3, p = .76$), indicating no significant difference in L2 speech intelligibility based on L1 language background as either Mandarin or Korean (despite the difference in signal-to-noise ratios). The two-way interaction was also non-significant ($B = -.45, SE B = .45, t(20) = -.99, p = .34$).⁵

⁵ The influence of language background is included in this analysis in order to verify the generality of the L1-L2 relationship within individual talkers (i.e., as a justification for combining our two groups of participants for the sake of the within-talker L1-L2 analysis that is the central point of this study); however, it should be noted that the strength of a main effect or interaction involving L1 background would likely vary

In order to verify this pattern with more closely matching testing conditions across the two groups of bilinguals, we conducted a second analysis that focused on the L1 Mandarin and L1 Korean intelligibility scores from the -4 dB and -5 dB signal-to-noise ratio conditions, respectively; and L2 Mandarin-accented English and L2 Korean-accented English intelligibility both from -0 dB signal-to-noise ratio conditions. The results of this analysis were consistent with the analysis reported above. The critical main effect of L1 intelligibility on L2 intelligibility was significant ($B = .8$, $SE B = .24$, $t(20) = 3.41$, $p < .01$). The main effect of L1 (Mandarin or Korean) was not significant ($B = .25$, $SE B = .37$, $t(20) = .67$, $p = .51$), and the two-way L1 (Mandarin or Korean) X L1 intelligibility interaction was also not significant ($B = -0.45$, $SE B = .47$, $t(20) = -.95$, $p = .35$).

This pattern of results—a dissociation between L1 and L2 intelligibility in absolute terms paired with an association of L1 and L2 intelligibility in relative terms—provides supporting evidence for overall intelligibility as an index of a talker-specific speech production trait that persists across status-specificity (L1 versus L2 status) and language-specificity (in this case, English versus either Mandarin or Korean).

3.3. Acoustic measurement data

Table 3 shows the average value for each of the acoustic parameters measured across all talkers and sentences for each of the two groups of bilingual talkers in both their L1 and L2. For each parameter, we built a linear model with fixed effects for Status of the task language (L1 versus L2, contrast coded as 0.5 and -0.5 , respectively) and Group (Mandarin-English versus Korean-English contrast coded as 0.5 and -0.5 , respectively) with random intercepts for talkers. We then used model comparison to assess significance by comparing the model with the factor of interest to a model with just that factor deleted ($n-1$ changes). For F0 mean and LTASS slope, the statistical analyses showed no main effects or interactions, indicating that these parameters remain stable across L1 and L2 speech production for both groups of bilingual talkers. For F0 range, there was no main effect of L1 vs. L2 Status, but there was a significant main effect of Group ($\chi^2(1) = 8.17$, $t = -2.99$, $p < .01$), with the Korean group producing a greater F0 range (average across L1 and L2 of 198 Hz) than the Mandarin group (average across L1 and L2 of 129Hz). There was also a significant Group by Status interaction ($\chi^2(1) = 5.09$, $t = 2.28$, $p < .05$), reflecting a greater difference between L1 and L2 F0 range for the Korean talkers (> 25 Hz) compared to the Mandarin talkers (< 4 Hz). This Group-based difference for F0 range may reflect different functions of pitch variation in Korean, Mandarin, and English resulting in greater L1-L2 differentiation for the Korean-English bilinguals compared to the Mandarin-English bilinguals. However a full exploration of this difference would

Table 3: Average acoustic measurements for all talkers in both their L1 and L2. Numbers represent averages across all talkers and sentences in each condition. Standard deviations are shown in parentheses.

| | Mandarin-English bilinguals | | Korean-English bilinguals | |
|-------------|-----------------------------|-------------------|---------------------------|-------------------|
| | L1 | L2 | L1 | L2 |
| F0 mean | 160.48 Hz (40.89) | 161.52 Hz (43.05) | 174.81 Hz (56.49) | 175.07 Hz (55.05) |
| F0 range | 130.16 Hz (58.89) | 127.08 Hz (56.28) | 183.35 Hz (64.14) | 211.92 Hz (47.63) |
| LTASS slope | -14.72 (3.66) | -14.22 (3.83) | -16.03 (1.92) | -14.83 (2.5) |

depending on a multitude of possible group-wise factors that are not explored (or relevant to) the present study, including overall proficiency and testing conditions.

require more extensive datasets with materials designed to more directly compare both word-level and phrase-level prosody across the languages. Critically, for the purposes of the present study, L1-L2 correlations for each of these parameters are extremely high ($r = .99, .85, \text{ and } .8$ for F0 mean, F0 range, and LTASS slope, respectively) indicating that within bilingual individuals there is very little variation along these parameters across L1 and L2 speech. Thus, in contrast to the absolute dissociation/relative association pattern that we observe for overall intelligibility, for all of the acoustic parameters that we measured in this study (F0 mean, F0 range, and LTASS slope), we observed extensive overlap between the group-wise L1 and L2 values (i.e., no L1-L2 dissociation in the boxplots in **Figure 2**) along with an extremely tight association between the L1 and L2 values within individual bilingual talkers (i.e., near perfect association of L1 and L2 values in the scatterplots in **Figure 2**). This pattern suggests that these source-related features exhibit a talker-specific trait that is both language-independent and status-independent.

4. Discussion

The present study provides evidence of a dissociation between L1 and L2 speech intelligibility in absolute terms and an association of L1 and L2 speech intelligibility in relative terms within a group of Mandarin-English and Korean-English bilinguals. That is, while all of the bilingual talkers were more intelligible in their L1 than in their L2 for native listeners of the language being spoken, talkers who were relatively high intelligibility in their L1 were also relatively high intelligibility in their L2.

4.1. Variation in L2 speech intelligibility

It is unsurprising that bilinguals vary substantially in their L2 speech intelligibility. Indeed, wide differences in all aspects of L2 performance is a hallmark of L2 learning. This variability stems from well-known interactions between the structures of the particular co-existing L1 and L2, as well as the general challenges of second language learning regardless of the particular source and target languages involved. Language-specific L1-to-L2 transfers presumably result in commonalities across groups of bilinguals such that particular foreign accents can be distinguished from each other (e.g., talkers of Chinese-accented English should be distinguishable from talkers of French-accented English by native English listeners). However, it is important to note that studies of identification and discrimination of foreign accents have yielded variable patterns, and have demonstrated a strong influence of degree of foreign accentedness on classification of accents into L1-based groups (i.e., groups with the same native language background). For example, Atagi and Bent (2013) found higher classification accuracy for high-intelligibility foreign-accented English sentences than for low-intelligibility foreign-accented English sentences, suggesting that, despite the fact that low-intelligibility sentences presumably contain more L1-to-L2 phonetic transfers, the greater processing demands of attending to difficult-to-understand sentences interferes with the classification task. Nevertheless, the critical role of language-specific L1-to-L2 phonetic and phonological transfer in L2 speech production has been extensively documented (see Davidson, 2011, for a review).

In addition to language-specific sources of variation, L2 speech production variability derives from cognitive features of bilingualism and the related challenges of speaking in an L2. Such status-specificity (L1 versus L2 status) is reflected in features such as the typically slower speaking rate of L2 speech relative to L1 speech regardless of the particular L1 and L2 involved (e.g., Guion, Flege, Liu, & Yeni-Komshian, 2000; Baese-Berk & Morrill, 2015, and many others).

4.2. Variation in L1 speech intelligibility

It is also well-established that native talker (i.e., L1) speech intelligibility varies substantially across individuals even for a standard set of simple spoken sentences when presented to native listeners drawn from the same population as the talkers, all of whom have no known speech, hearing, or language deficits. Interestingly, this prior work on variation in L1 intelligibility has indicated substantial individual variation in overall intelligibility in the absence of any explicit or implicit instruction to enhance intelligibility (i.e., inter-talker variation in baseline intelligibility, e.g., Bradlow et al., 1996; Hazan & Markham, 2004), as well as in the extent to which talkers modify their speech in response to a communication barrier (i.e., inter-talker variation in adaptive intelligibility, e.g., see Schum, 1996; Ferguson, 2004, 2012). As discussed above, acoustic-phonetic comparisons of both inter- and intra-talker variability have identified a wide range of acoustic correlates of intelligibility (see Cooke et al., 2014 for a recent review). In general, these intelligibility-related acoustic-phonetic variations can be viewed as being guided by the goals of greater acoustic salience and audibility at the utterance level in combination with enhanced phonetic contrast between phonologically distinct categories. While the distal source of individual variation in L1 speech production—both baseline and adaptive—remains unclear, such variation is likely determined by a combination of peripheral mechanisms (i.e., physical characteristics of the vocal tract and other speech production organs) and central mechanisms (including cognitive, linguistic, and social components).

4.3. Relation between variation in L2 and L1 speech intelligibility

Due to the different underlying sources of L2 and L1 variation, it is conceivable that there would be no connection on an individual level between variation in L1 intelligibility and variation in L2 intelligibility. Yet, the present data showed a strong positive association between relative intelligibility in L1 and L2 in bilingual talkers, such that variation in L1 intelligibility predicts variation in L2 intelligibility. This pattern strongly suggests that talker-specificity is not necessarily constrained by either language-specificity or L1 versus L2 status-specificity. Instead, overall speech intelligibility is sensitive to a talker-specific trait characteristic that combines with, rather than is overwhelmed by, language-specific and dominance-dependent influences on bilingual speech production. That is, the data suggest that the sound shape of bilingual speech production is modulated by a confluence of at least three streams of influence: one that flows from an interaction between the underlying phonological structures and patterns of phonetic implementation of the particular L1 and L2 (language-specificity), a second that flows from the general challenge of managing the relative activation and suppression of two (or more) languages in any given speech communication situation (L1 versus L2 status-specificity), and a third that is related to an individual-level speech production setting that, together with automatic features of an individual's speech production anatomy and physiology, establishes persistent talker trait characteristics (talker-specificity).⁶

Within specific communicative settings, this confluence of individual-level and group-level influences on bilingual speech production will presumably modulate (and be modulated by) the adaptive mechanisms that seek to balance talker-oriented and listener-oriented constraints (i.e., variation along the hyper-to-hypo-articulation continuum of Lindblom's H&H theory of phonetic variation, Lindblom, 1990). For example, we might

⁶ An alternative, though unlikely, alternative account for the present findings is that the voice characteristics of individual talkers are differentially affected by noise such that some talkers' speech samples are better masked than those of other talkers regardless of the language being spoken.

speculate that talker-specific characteristics may be attenuated in hyper-articulated, clear speech where discriminability of linguistic contrasts is prioritized. Similarly, the expression of individual-level traits across the time-course of second language learning and across various types of bilingualism (i.e., across bilinguals with varying balances of proficiency in their languages) is also likely to vary. It remains for future research to determine exactly how individual-level and group-level influences in bilingual speech production are integrated with both the short-term dynamics of speech communication in specific conversational situations and with the longer-term dynamics of second language acquisition.

Finally, growing evidence of language-independent talker-specificity in bilingual speech production raises the question of whether we might observe a parallel link between L1 and L2 speech perception. There is a small but expanding literature indicating such a link (e.g., see Díaz et al., 2008; Earle & Arthur, 2017), which together with the present project, might suggest a deep neural source for individual differences in speech and language function and learning that manifests in both L1 and L2 production and perception.

Acknowledgements

We gratefully acknowledge the essential technical support of Chun-Liang Chan, and research assistance of Vanessa Dopker, Emily Kahn, and Kelsey Mok. We are also grateful to Professor Taehong Cho, Director of the Hanyang Phonetics & Psycholinguistics Laboratory (HPPL), for generous help with the recruitment of Korean participants and access to experiment running facilities. Work supported by Grant R01-DC005794 from NIH-NIDCD.

Competing Interests

The authors have no competing interests to declare.

References

- Allen, J. S., Miller, J. L., & DeSteno, D. 2003. Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113, 544–552. DOI: <https://doi.org/10.1121/1.1528172>
- Atagi, E., & Bent, T. 2013. Auditory free classification of nonnative speech. *Journal of Phonetics*, 41, 509–519. DOI: <https://doi.org/10.1016/j.wocn.2013.09.003>
- Baese-Berk, M., & Morrill, T. 2015. Speaking rate consistency in native and non-native speakers of English. *Journal of the Acoustical Society of America*, 138(3), EL223–EL228. DOI: <https://doi.org/10.1121/1.4929622>
- Bamford, J., & Wilson, I. 1979. Methodological considerations and practical aspects of the BKB sentence lists. In: Bench, J., & Bamford, J. (eds.), *Speech-hearing tests and the spoken language of hearing-impaired children*, 148–187. London, UK: Academic Press.
- Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Boersma, P., & Weenink, D. 2017. Praat: Doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from: <http://www.praat.org/>.
- Bond, Z. S., & Moore, T. J. 1994. A note on the acoustic phonetic characteristics of inadvertently clear speech. *Speech Communication*, 14(4), 325–337. DOI: [https://doi.org/10.1016/0167-6393\(94\)90026-4](https://doi.org/10.1016/0167-6393(94)90026-4)
- Bradlow, A. R., Ackerman, L., Burchfield, L., Hesterberg, L., Luque, J., & Mok, K. 2011. Language- and talker-dependent variation in global features of native and non-native speech. In: *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 356–359.

- Bradlow, A. R., & Bent, T. 2002. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112(1), 272–284. DOI: <https://doi.org/10.1121/1.1487837>
- Bradlow, A. R., Kim, M., & Blasingame, M. 2017. Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers. *Journal of the Acoustical Society of America*, 141, 886–899. DOI: <https://doi.org/10.1121/1.4976044>
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255–272. DOI: [https://doi.org/10.1016/S0167-6393\(96\)00063-5](https://doi.org/10.1016/S0167-6393(96)00063-5)
- Cooke, M., King, S., Garnier, M., & Aubanel, V. 2014. The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language*, 28, 543–571. DOI: <https://doi.org/10.1016/j.csl.2013.08.003>
- Davidson, L. 2011. Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass*, 5(3), 126–139. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00266.x>
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. 2013. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 1–21.
- Derwing, T. M., & Munro, M. J. 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1–16. DOI: <https://doi.org/10.1017/S0272263197001010>
- 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. DOI: <https://doi.org/10.1017/S0272263109990015>
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. 2008. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), 16083–16088. DOI: <https://doi.org/10.1073/pnas.0805022105>
- Earle, F. S., & Arthur, D. T. 2017. Native phonological processing abilities predict post-consolidation nonnative contrast learning in adults. *The Journal of the Acoustical Society of America*, 142, EL525–EL531. DOI: <https://doi.org/10.1121/1.5013141>
- Ferguson, S. H. 2004. Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116, 2365–2373. DOI: <https://doi.org/10.1121/1.1788730>
- Ferguson, S. H. 2012. Talker differences in clear and conversational speech: Vowel intelligibility for listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 55, 779–790. DOI: [https://doi.org/10.1044/1092-4388\(2011/10-0342\)](https://doi.org/10.1044/1092-4388(2011/10-0342))
- Ferguson, S. H., & Kewley-Port, D. 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 112, 259–271. DOI: <https://doi.org/10.1121/1.1482078>
- Ferguson, S. H., & Kewley-Port, D. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, 50, 1241–1255. DOI: [https://doi.org/10.1044/1092-4388\(2007/087\)](https://doi.org/10.1044/1092-4388(2007/087))
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. 1995. Accent, intelligibility, and comprehensibility. *Journal of the Acoustical Society of America*, 97(5), 3125–3134. DOI: <https://doi.org/10.1121/1.413041>
- Granlund, S., Hazan, V., & Baker, R. 2011. An acoustic-phonetic comparison of the clear speaking styles of Finnish–English late bilinguals. *Journal of Phonetics*, 40(3), 509–520. DOI: <https://doi.org/10.1016/j.wocn.2012.02.006>

- Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. 2000. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21(2), 205–228. DOI: <https://doi.org/10.1017/S0142716400002034>
- Hazan, V., & Markham, D. 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, 116, 3108–3118. DOI: <https://doi.org/10.1121/1.1806826>
- Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J., & Marchal, A. (eds.), *Speech Production and Speech Modeling*, 403–439. Dordrecht, Netherlands: Kluwer Academic.
- 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. DOI: <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Pearson. 2009. Official Guide to Pearson Test of English Academic. London: Longman.
- Remez, R. E., Fellowes, J. M., & Rubin, P. E. 1997. Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651–666. DOI: <https://doi.org/10.1037/0096-1523.23.3.651>
- Schum, D. J. 1996. Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the American Academy of Audiology*, 7(3), 212–218.
- Smiljanic, R. (Forthcoming). Clear speech perception. In: *Handbook of speech perception*, Pardo, J., Remez, R. E., & Pisoni, D. B. (eds.).
- Smiljanic, R., & Bradlow, A. R. 2005. Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, 118(3), 1677–1688. DOI: <https://doi.org/10.1121/1.2000788>
- Smiljanic, R., & Bradlow, A. R. 2008. Temporal organization of English clear and plain speech. *Journal of the Acoustical Society of America*, 124(5), 3171–3182. DOI: <https://doi.org/10.1121/1.2990712>
- Smiljanic, R., & Bradlow, A. R. 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass*, 3(1), 236–264. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00112.x>
- Smiljanic, R., & Bradlow, A. R. 2011. Bidirectional clear speech perception benefit for native and high proficiency non-native talker-listener pairs: Intelligibility and accentedness. *Journal of the Acoustical Society of America*, 130(6), 4020–4031. DOI: <https://doi.org/10.1121/1.3652882>
- Soli, S. D., & Wong, L. L. 2008. Assessment of speech intelligibility in noise with the Hearing in Noise Test. *International Journal of Audiology*, 47(6), 356–361. DOI: <https://doi.org/10.1080/14992020801895136>
- Sproat, R., Shih, C., Gale, W., & Chang, N. 1996. A stochastic finite-state word-segmentation algorithm for Chinese. *Computational Linguistics*, 22(3), 377–404.
- Theodore, R. M., Miller, J. L., & DeSteno, D. 2009. Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America*, 125, 3974–3982. DOI: <https://doi.org/10.1121/1.3106131>
- Towell, R., & Dewaele, J. M. 2005. The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In: Dewaele, J. M. (ed.), *Focus on French as a foreign language*, 210–239. Clevedon, UK: Multilingual Matters.
- Tuomainen, O., Hazan, V., & Romeo, R. 2016. Do talkers produce less dispersed phoneme categories in a clear speaking style? *Journal of the Acoustical Society of America*, 140(4), EL320–EL326. DOI: <https://doi.org/10.1121/1.4964815>
- Uchanski, R. M. 2005. Clear speech. In: Pisoni, D. B., & Remez, R. E. (eds.), *Handbook of speech perception*, 207–235. Malden, MA: Blackwell Publishers.

How to cite this article: Bradlow, A. R., Blasingame, M. and Lee, K. 2018 Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9(1): 17, pp. 1–20, DOI: <https://doi.org/10.5334/labphon.137>

Submitted: 22 January 2018 **Accepted:** 02 August 2018 **Published:** 09 October 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access icon, which is a stylized padlock with a circular arrow around it.