Heritage bilingual talkers do not exhibit an intelligibility penalty in either language*

Robert Daland · Mira Oh**
(UCLA · Chonnam National University)

Daland, Robert and Mira Oh. 2014. Heritage bilingual talkers do not exhibit an intelligibility penalty in either language. Linguistic Research 31(3), 403-440. Some speakers have clearer voices than others. In the present study, we investigate whether early bilingualism has systematic repercussions for talker clarity, using a group of Korean-American heritage bilingual talkers, as well as American and Korean (near-)monolinguals. American monolinguals and the bilinguals produced English sentences, which were mixed with white noise at two signal-to-noise ratios and presented to native English listeners. Korean monolinguals and the same bilinguals produced Korean sentences, which were presented to native Korean listeners. In English, there was no monolingual/bilingual difference in talker clarity. In Korean, bilingual talkers were more clear than monolinguals. Post hoc investigation showed that Korean American bilinguals produced the Korean stimuli more slowly than monolinguals, explaining part of the bilingual advantage. Taken together, these results suggest that early bilingual exposure does not cause a talker clarity deficit, in either language. (UCLA · Chonnam National University)

Keywords heritage bilinguals, talker clarity, intelligibility, Korean, English

1. Introduction

What are the consequences of early multiple language exposure? This question is of interest to many people. Over half the world's population has some fluency in more than one language (Crystal, 1997), including early bilinguals (speakers who acquire a second language during early childhood). As a matter of understanding

---

* We wish to acknowledge Syejeong Kim for assistance in running the production and perception studies in Gwangju and three reviewers for their invaluable comments. We also wish to thank Megha Suddara, Marc Garellek, and Midam Kim for commenting on an early draft, as well as Matt Goldrick, Ann Bradlow, and Pat Keating for thoughtful discussion.

** Corresponding author
existing speakers and informing coherent social policies, it is important to study early bilinguals. Moreover, new parents are intensely curious as to whether raising their child bilingual is harmful or helpful. Individuals and society are in a position to make better decisions when we possess more information about the alternatives. Behavioral studies are an excellent way to determine the consequences of single versus multiple early language exposure, and that is the broad topic we take up here.

Early multilingual exposure may have different consequences in different domains. For instance, early bilinguals always acquire the phonetic/phonological contrast system of a language (reviewed later), but do not always acquire a native accent (Flege, Munro, & MacKay, 1995; Flege, Yeni-Komshian, & Liu, 1999). In the present study, we investigate the consequences of early multiple language exposure for talker clarity.

Talker clarity refers to the ease with which a speaker can be understood. Talker clarity is logically related to the concept of intelligibility. Intelligibility refers to the extent to which a listener can recover the linguistic message produced by a talker in a particular discourse setting. Intelligibility can vary from moment to moment, depending on all components of the discourse setting, including talker properties, listener properties, message properties, and environmental properties (Smiljanic & Bradlow, 2009). Talker clarity refers to the specific subset of intelligibility-related properties in a discourse setting that are associated with the talker: inherent properties such as native language background, as well as talker-controlled properties such as speech rate and the use of a clear speech style. In other words, talker clarity is an abstract property that we assign to speakers, based on the observable fact that some talkers are more intelligible on average across a range of discourse settings.

Talker clarity is inherently theoretically interesting because it is necessary for successful language communication: a message must be produced clearly enough that the listener can recover the intended meaning. In addition to general theoretical interest, talker clarity in early bilinguals may be of special interest for two reasons. First, research on talker clarity in early bilinguals fills an important gap in the literature. As reviewed in more detail below, there has been considerable research on intelligibility in monolinguals, and some research on talker clarity in late bilinguals vs. monolinguals, but almost none on early bilinguals. Second, early bilinguals offer a unique opportunity to explore the role of fine-grained phonetic variation in talker clarity. Early bilinguals differ systematically from monolinguals in fine-grained phonetic implementation of
Heritage bilingual talkers do not exhibit an intelligibility penalty... 405

phonological categories (reviewed later), so bilingual-monolingual differences in talker clarity (or the absence thereof) give us important information about the role of particular kinds of phonetic detail in speech perception, and the organization of bilingual speech production systems.

In the following sections, we provide additional background on speech production in early bilinguals and previous intelligibility research. The body of the paper describes an intelligibility experiment whose goal is to compare the clarity of Korean-American early bilinguals to monolingual speakers of both languages they speak (Korean and English). The bilingual talkers for this study belong to a special subset of early bilinguals known as heritage bilinguals, who are recruited because they make up a sizable proportion of college-aged Korean speakers in Los Angeles, where the American data were collected.

2. Intelligibility

There is a long tradition of research on intelligibility in the speech and hearing sciences. One important line of research has focused on the effect of noise and other signal degradations on intelligibility, since this is relevant for listeners with hearing and other communicative disorders. Some research has investigated speech perception in noise for children and non-native listeners, who have less experience with the target language, and/or interference from a non-target language. Most recently, studies have begun to investigate non-native talkers, and/or early bilingual listeners. To date, no study has explicitly compared talker clarity in monolinguals and early bilinguals, a gap which the present study aims to fill.

2.1 Effects of masking noise on intelligibility

One of the most robustly established findings is that louder noise results in lower intelligibility. Noise is standardly quantified using the signal-to-noise ratio (hereafter, SNR) -- the loudness of the signal (in decibels) minus the loudness of the noise. Simpson & Cooke (2005) presented vowel-consonant-vowel sequences to native listeners, where the 'noise' consisted of mixing in recordings from one or more non-target talkers. They found that (equating SNR across conditions) intelligibility
was best when there was only one competing talker, considerably worse with speech-shaped noise (random noise whose spectrum matches the long-term average of speech), but the actual worst-case was an intermediate number of competing talkers. They interpreted this as evidence that intelligibility is most strongly affected by 'perceptual masking' (listeners' inability to compensate for short-term fluctuations in the spectral distribution of the noise), rather than by 'informational masking' (allocation of attentional resources to processing non-target talkers). However, other research has demonstrated that informational masking too plays a role in intelligibility (Sperry, Wiley & Chial, 1997; Brouwer, Van Engen, Calandruccio, & Bradlow, 2012). In the present study, we use white noise as a masker, which controls for perceptual masking across language background, and eliminates informational masking altogether.

### 2.2 Message and listener properties

In general, it is easier to recognize speech that is more predictable (Miller & Isard, 1963; Miller, Heise, & Lichten, 1951; Kalikow, Stevens, & Elliott, 1977). This finding is readily explained by the assumption that listeners apply linguistic and semantic knowledge predictively for speech recognition (see, e.g. review in Norris & McQueen, 2008). However, the predictability of a message depends on the listener's expectations. As Bradlow & Alexander (2007) point out, "the general pattern of experimental findings, and the common experience of non-native listeners, is that the detrimental effects of environmental signal distortion are greater for nonnative than native language speech perception" (p. 2339). The selective disadvantage of non-native speakers appears to be caused by poorer knowledge of the higher-level structure of the language, since highly proficient non-native listeners do not exhibit the expected deficit (relative to monolinguals) when they are identifying nonword CV and VC syllables (Cutler, Webber, Smits, & Cooper, 2004). Early bilinguals pattern with monolinguals in the perception of meaningful speech in noise, and more specifically in their ability to leverage contextual information such as lexical cooccurrence (Mayo, Florentine & Buus, 1997).
2.3 Talker background

One important component in talker clarity is the talker's language background, in particular whether they are a native or non-native speaker of the language. Bent & Bradlow (2003) manipulated talker background, using semantically coherent sentences in noise (the same materials are used in the present study). They found that for native English-speaking listeners, intelligibility is higher when the talker is a native English speaker than when the talker is a high-proficiency late bilingual. However, the same asymmetry does not necessarily obtain for talkers. Specifically, Hayes-Harb, Smith, Bent, & Bradlow (2008) found that native Mandarin listeners were equally good at identifying voiced/voiceless coda consonants whether the talker was a native English speaker or a native speaker of Mandarin with high English proficiency. These results suggest that, relative to monolingual-monolingual dyads, there is a greater decrement to intelligibility if the talker is non-native than if the listener is.

2.4 Talker manipulations

When talkers are asked to speak clearly, it generally results in higher intelligibility for at least a subset of listeners (Smiljanic & Bradlow, 2009). However, "the exact articulatory-acoustic cues that contribute to the clear speech advantage remain rather elusive" (p. 244). When speakers are asked to speak clearly, talkers introduce "a wide range of acoustic/articulatory adjustments, including a decrease in speaking rate (longer segments as well as longer and more frequent pauses), wider dynamic pitch range, greater sound-pressure levels, more salient stop releases, greater rms intensity of the non-silent portions (i.e. release burst, frication, and/or aspiration) of obstruent consonants, increased energy in 1000–3000 Hz range of long-term spectra and higher-voice intensity" (pp. 241-242). Smiljanic & Bradlow review a number of studies, concluding that manipulating most of these cues does not cause better perception in listeners (Byrd, 1994; Bond & Moore, 1994; Bradlow et al., 1996; Picheny et al., 1986; Moon & Lindblom, 1994; Ferguson & Kewley-Port, 2002, 2007; Bradlow et al., 2003; Krause & Braida, 2004; Hazan & Markham, 2004; Picheny, Durlach, & Braida, 1989; Stollman, Kapteyn, & Sleeswijk, 1994; Uchanski et al. 1996; Liu & Zeng, 2006). The only factor that has consistently been found to
show an intelligibility gain is adding pauses to natural conversational speech (rather than slowing down the speech rate between pauses or expanding the vowel space). This literature suggests that speech rate makes at most a modest within-talker contribution to clarity; slower speech is not automatically more clear.

3. Bilingual speech production

The speech production system of bilingual speakers has been the topic of considerable research. It has generally been found that early but not late bilinguals acquire the full contrast system of the second/target language. Moreover, early bilinguals generally differ in fine-grained phonetic detail from monolinguals; that is, early bilinguals may express the phonological contrasts slightly differently than monolinguals. These facts can be accounted for by Flege's Speech Learning Model, the most well-known theory on acquisition of speech sound categories (Flege, 2005, and references therein). After describing the SLM, we situate it with respect to the Korean-American bilingual talkers at the heart of the present study.

The contrast system of a language refers to the set of phonological oppositions that can signal meaning differences. For example, the English words *bit* and *beat* are slightly different phonetically, and in English this kind of phonetic difference systematically translates to a word meaning difference. Late bilinguals do not always acquire the full contrast system of the second language. An example is the acquisition of the French /y/-/u/ vowel contrast by native English speakers. Flege (1987) found that French monolinguals distinguish these categories with a robust F2 contrast (/y/: $\approx 2050$ Hz; /u/: $\approx 1200$ Hz; non-overlapping distribution), while American professors teaching French in America (late bilinguals) did not (/y/: $\approx 2000$ Hz; /u/: $\approx 1600$ Hz, overlapping distribution). More specifically, the late bilinguals appeared to know that these vowels were supposed to contrast, but they often produced intermediate F2 values for both /y/ and /u/, so that the intended contrast would not always be recoverable from the phonetics. A similar finding obtained in a study of the acquisition of the rounded central vowel *schwar* by Italian Americans (Flege, Schirru, & MacKay, 2003).

A different pattern has been found for early bilinguals: such speakers always acquire the same contrast system as monolinguals, but differ in the precise phonetic
Heritage bilingual talkers do not exhibit an intelligibility penalty...  

Implementation. For example, Khattab (2000) investigated the production of word-initial voiced and voiceless by children acquiring both English and Egyptian Arabic. She found that, like monolingual controls, early bilingual children produced different VOT values for voiced and voiceless stops, in each language. However, the mean VOT values for English voiced and voiceless stops were different from the corresponding English monolingual VOT values. The same was true for Arabic. Thus, these children acquired the contrast system in each language, while differing in the fine phonetic detail from age-matched monolinguals. Sundara, Polka, & Baum (2006) found a comparable pattern for coronal stops in Canadian English-French bilingual adults: bilinguals were significantly more likely to produce lead voicing (prevoicing) in English voiced stops than monolinguals were, which can plausibly be attributed to bilingualism since voiced stops almost inevitably exhibit lead voicing in Canadian French monolinguals.

3.1 The Speech Learning Model

According to Fleger (2005), the purpose of the Speech Learning Model (hereafter, SLM) is "to account for how individuals learn – or fail to learn – to produce and perceive phonetic segments (vowels, consonants) in a second language" (slide 4). The core assumptions of the SLM are as follows (slides 84-85):

- L2 learners can, in time, veridically perceive the phonetic properties of L2 speech sounds
- as in L1 learning, L2 learning takes time, and is influenced by the nature of the input
- as in L1, production is guided by perceptual representations stored in LTM
- processes of L1 acquisition, including category formation, remain intact in L2 acquisition
- phonetic elements in L1 and L2 exist in a common phonological space, and influence each other

Fleger states that the SLM makes the following predictions (slides 86-89):
the greater the perceived dissimilarity between an L2 sound and a neighboring L1 category, the more likely that a new category will be formed.

category formation for an L2 sound becomes less likely throughout childhood as the L1 representations develop and are elaborated.

when a novel category is not formed in the L2, the L2 sound will be assimilated to an existing L1 category; the merged category will be affected by both L1 and L2 phonetic distributions.

when a novel category is formed in the L2, it may dissimilate from the neighboring L1 category to preserve contrast.

We turn now to situating these points with respect to the talkers in this study.

As evident from the production and accent perception studies by Flege cited above, the SLM is concerned with the acquisition of language-specific acoustic properties, as they reflect (or fail to reflect) native-like phonetic performance. However, the SLM is not intended as a theory of intelligibility in general, or talker clarity in particular. Rather, it makes broad predictions about category structure, which might be interpreted with respect to talker clarity. For example, when the SLM predicts that an individual will not acquire the tense/lax contrast in English vowels, we can infer the talker will have lower clarity since tense/lax is lexically contrastive. However, provided that an individual acquires the same set of phonological contrasts as a monolingual, the SLM does not directly speak to the issue of talker clarity.

The early bilinguals that were recruited for this study are best described as heritage bilinguals. As discussed in more depth in the Methods section, these speakers were generally exposed to Korean in the home from birth, before or alongside their exposure to English. They were educated largely in English, and estimate their English proficiency to be higher than their Korean proficiency. Thus, these speakers are maximally likely to form distinct categories for sounds that are phonetically similar in both languages, such as Korean tensed stop versus the English voiced stop.
3.2 Studies of Korean–English bilinguals

A number of phonetic studies have been conducted with Korean-English bilinguals specifically, all focusing on the stop systems. In Peninsular Korean speakers born after about 1970, the three laryngeal categories (lax/lenis, tense/fortis, and aspirated) are distinguished word-initially by a combination of VOT and pitch, possibly in combination with phonation (Kim, Beddor, & Horrocks, 2002; Silva, 2006). An early study (Kang & Guion, 2006) found that late bilinguals had merged the English voiced stop category with the Korean tensed stop category, but that early bilinguals did not differ from English monolinguals (in English), nor from Korean monolinguals (in Korean). An important fact is that the participants Kang and Guion classified as 'Korean monolinguals' had resided in the United States for up to 5 months (p. 1674), and were recruited from the American English Institute, which is an intensive English immersion program associated with the University of Oregon. It is now known that second language exposure can affect first language phonetic properties in as little as 6 weeks (Chang, 2012). Thus, this 'monolingual' population would be classified as late bilinguals by many researchers. It follows that claims about early-monolingual differences are problematic; their claims rather apply to early-late bilingual differences. Moreover, subsequent studies have found differences between early bilinguals and monolinguals (Oh & Daland, 2011; Lee & Iverson, 2012). In the next subsection, we discuss recent work on heritage bilingualism with speakers who experienced an abrupt drop in L1 input around primary school, including some studies with Korean heritage bilinguals.

3.3 Heritage bilingualism

Our use of the term 'heritage bilingual' follows Benmamoun, Montrul, & Polinsky (in press): "a heritage speaker is an early bilingual who grew up hearing (and speaking) the heritage language (L1) and the majority language (L2) either simultaneously or sequentially in early childhood... but for whom L2 became the primary language at some point during childhood (at, around, or after the onset of schooling). As a result of language shift, by early adulthood a heritage speaker can be strongly dominant in the majority language, while the heritage language will now be the weaker language." This is an exact characterization of the heritage bilinguals
in the present study -- all acquired Korean or Korean and English in the home during early childhood, and learned English through schooling and their interactions with non-Korean society after entering primary school, while maintaining some use of Korean in the home.

A fascinating series of studies by Au, Knightly, Jun, and Oh have investigated 'childhood memory' of language, in which input in the first language abruptly ceased at some point in early childhood, to be replaced with English. These studies were conducted primarily with Spanish emigres (Au, Knightly, Jun, & Oh, 2002; Au, Oh, Knightly, Jun, & Romo, 2008) and Korean emigres (Oh, Jun, & Knightly, 2003) or adoptees (Oh, Au, & Jun, 2010). The broad pattern that these studies reveal is that even a single year of input in the childhood language confers a substantial advantage in (re-)acquiring the sound system of the childhood language (relative to late bilinguals with no previous experience); but between 4-6 years of childhood exposure are needed to confer a comparable advantage in (re-)acquiring the morphosyntax of the childhood language. The 'sound system' here refers to various properties, including VOT, phonological processes, and accent rating.

Note that the heritage bilinguals in the studies cited differ from the heritage bilinguals talkers in the present study. An inclusion requirement for the current study was that the speakers had continuous exposure to Korean, spoke it at least occasionally with friends or family, and were able to read/write Hangeul. As Benmamoun et al. state, "Heritage speakers vary widely in the degree of their receptive and productive command of the heritage language... It is typical of heritage speakers to have better-developed listening and speaking abilities than reading and writing abilities, a discrepancy that is mainly due to the lack of schooling in the heritage language". The heritage speakers in this study may differ from educated monolinguals in their command of the language as a whole, but they are likely to have a native or near-native sound system.

4. Experiment

In this section, we present an experiment designed to assess whether early (heritage) bilingualism has a systematic effect on talker clarity (in educated adult speakers). Prior to the experiment itself, we situate possible outcomes in the theoretical landscape.
4.1 Predictions

In discussing this work with colleagues, we found a wide diversity of opinions as to what actually was the expected behavior. What this tells us, at the least, is that the experiment conducted here is actually establishing something that the scientific community did not know before, and which could be considered surprising. This section summarizes the predictions that might be drawn from the literature reviewed earlier.

The intelligibility literature contains one study that suggests that early bilingual listeners pattern with monolinguals. The production studies on early bilingual talkers show that they are qualitatively similar to monolinguals, but exhibit some fine-grained phonetic differences. This is a fairly weak basis from which to derive predictions. The strongest inference that can be drawn is that it would be surprising if early bilingual talkers have considerably lower talker clarity than monolinguals in Korean, since the heritage bilingualism studies reviewed earlier suggest a native-like sound system should be acquired.

It is logically imaginable that there would be no bilingual-monolingual difference in either language. The simplest explanation for this outcome would be that early bilinguals do achieve native-like behavior in both languages (as evidenced by being behaviorally indistinguishable from monolinguals) for this task. From one perspective, this is a null result, which many scientists find uninteresting. However, in the present case a null result would be informative, since it would bear on various theoretical issues, such the critical period hypothesis.

Alternatively, if different results are found in each language, the particular pattern that is obtained would presumably derive from the different conditions under which the study population learned and/or uses each language. For example, suppose that early bilinguals are equally as clear as monolinguals in English (since they operate in this language the majority of the time), but less clear than monolinguals in Korean (since they only use this language informally, when interacting with close family members). This would support a usage-based interpretation.

4.2 Talkers

Korean-English heritage bilinguals (5 female, 4 male) and English monolinguals (9 female, 6 male) were recruited from the Los Angeles area. (Three additional
bilinguals were recruited who immigrated to the US at the age of 10 or later. That data is not reported here.) Monolingual Koreans (5 female, 5 male) were recruited from Seoul. All talkers were college-aged (18-25) and were paid 10 USD per recording session or received course credit.

For the Los Angeles talkers, language status was identified through a language background questionnaire (Sundara et al., 2006). Early bilinguals were defined by continuous exposure to English and Korean since the age of 7. English monolinguals were defined jointly as being exposed mainly to English in the home and community, having 2 years or less of academic instruction in any single foreign language (languages studied include Spanish, French, German, and Chinese; talkers with Korean instruction were excluded), and residing for not more than 6 weeks in any single country whose dominant language was other than English. Table 1 summarizes the language backgrounds of the heritage bilinguals; the percentage of time spent speaking each language was not inferrable from their answers to the language background surveys, but talkers were not recruited unless they indicated they used both languages as part of their normal life.

Table 1. Summary of language background data on heritage bilinguals. AOA indicates age of immersion in English (0 means parents spoke English and Korean in the home from birth). Proficiency ratings are the average over self-reported proficiencies in Reading, Writing, Grammar, Vocabulary, Accent, and Fluency on a Likert scale where 1 indicates no ability and 7 indicates native-level ability.

<table>
<thead>
<tr>
<th>Talker</th>
<th>AOA</th>
<th>Korean proficiency</th>
<th>English proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*F1</td>
<td>6</td>
<td>6.1</td>
<td>7</td>
</tr>
<tr>
<td>B*F2</td>
<td>0</td>
<td>6.2</td>
<td>7</td>
</tr>
<tr>
<td>B*F3</td>
<td>5</td>
<td>6.8</td>
<td>7</td>
</tr>
<tr>
<td>B*F4</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>B*M1</td>
<td>0</td>
<td>3.8</td>
<td>7</td>
</tr>
<tr>
<td>B*M2</td>
<td>0</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>B*M3</td>
<td>0</td>
<td>5.8</td>
<td>7</td>
</tr>
<tr>
<td>B*M5</td>
<td>0</td>
<td>6.8</td>
<td>7</td>
</tr>
<tr>
<td>B*M6</td>
<td>0</td>
<td>5.8</td>
<td>7</td>
</tr>
</tbody>
</table>

Interestingly, the heritage bilinguals in this study identified their English
Heritage bilingual talkers do not exhibit an intelligibility penalty...  415

proficiency as 7 on all dimensions, while identifying their Korean proficiency as 5-6 on the same dimensions. Some identified themselves as native speakers of English, despite not having been exposed to it during the first several years of their life.

Korean talkers were recruited by a native Korean-speaking research assistant with near-native proficiency in English, who teaches English at the university level. Korean monolingual talkers were recruited only if they had lived in Korea all their lives. Note that such speakers may not be regarded as ‘true’ monolinguals, since English language classes are compulsory in South Korea starting from primary school. However, the principal emphases of these classes are not on listening and speaking, but on reading and writing. Note further that it is customary for many Korean students to attend private, supplementary educational programs called *hagwon* which provide additional instruction in math, English, or other subjects deemed especially important. Thus, the level of English speaking proficiency among college-educated Koreans is very heterogeneous. During the recruitment process, the research assistant assessed the talker's English proficiency; talkers were not recruited if the assistant deemed their English proficiency to be intermediate or above. For example, a talker who did not understand an English question like, “What is your name?” would be deemed low proficiency.

4.3 Items

Items were drawn from the BKB list (Bench, Kowal, & Bamford, 1979), a list originally designed to test for hearing impairment in children. The items all have simple grammatical structures (Subject-Verb-Object, Subject-Verb-Prepositional Phrase) and use highly familiar nouns and verbs, e.g. ‘The boy dropped the ball.’ The same 60-sentence subset was used as in Bent & Bradlow (2003). English monolinguals were recorded reading the BKB sentences aloud. Korean monolinguals were recorded reading a direct translation of the BKB sentences aloud. Bilinguals were recorded reading both the English and the Korean materials. Talkers were instructed to read the sentences aloud at a comfortable pace. They were not aware that their productions would later be mixed with noise. The list of English sentences, together with the Korean translations (in Hangeul), are given in Appendix A.

Every sentence token was mixed with white noise at two noise levels, -4 and -8 dB SNR, with 500 ms of masking noise before and after the stimulus, and subsequently normalized to 75 dB SPL using Praat (Boersma & Weenink, 2012).
4.4 Listeners

Korean native speakers (45) were recruited in Gwangju, S. Korea and received course credit. (Two additional speakers were recruited, but their data was discarded because they were native Chinese speakers.) As noted in the Talker subsection above, Koreans take English classes from primary school on, but the emphasis in those courses is on reading and writing rather than speaking and listening.

English native speakers (33) were recruited from Los Angeles and received course credit or were paid 10 USD for their participation. It was not required that listeners be monolingual, only that they had been exposed to English continuously from birth. (This is as a matter of both practicality and ecological validity. We estimate that around 80% of the participants we recruit received considerable exposure to some language other than English prior to high school, which means this is the population talkers would have to be intelligible to where the study was conducted. Also, restricting to monolinguals would require discarding 80-90% of participants.)

4.5 Lists

Stimulus lists began with a practice block, to familiarize participants with the task and because strong practice effects are observed in this kind of task (Bradlow, p.c.). The practice block consisted of 15 items, half at an SNR of -4 dB and half at -8 dB. The practice items were generally similar to the BKB sentences, and were included only to familiarize listeners with the task. They were produced by the first author (English) or the research assistant (Korean), so as to avoid familiarizing listeners with any of the test voices. The same 15 (English or Korean) items were included in random order.

As for the test items, a separate stimulus list was created for each listener. Each stimulus list contained 60 test items -- exactly one repetition of each BKB item (60 items). The list was divided into four blocks, with SNRs of -8, -4, -8, and -4 respectively. Items were counterbalanced by talker status, so that exactly 30 items were produced by a monolingual talker, and exactly 30 items were produced by a bilingual talker (equal numbers within each noise level). For a given list/item/status, the talker was selected randomly; for example if the item was BKB English sentence
Heritage bilingual talkers do not exhibit an intelligibility penalty... 417

1 and for this list sentence 1 should be a bilingual, a production of BKB sentence 1 was randomly selected from all the available productions by English bilinguals (SNR -8 if block 1 or 3; -4 if block 2 or 4). In other words, talker was randomized subject to the constraint that number of monolingual/bilingual productions was counterbalanced; SNR was counterbalanced by blocking. The order of BKB sentences was randomized within and across lists, so that BKB sentence 1 had an equal likelihood of occurring first, second, or sixtieth at test.

4.6 Procedure

Participants were seated at a testing station wearing headphones. The experiment was administered through Praat. At the beginning of the experiment, listeners were instructed that they would hear sentences in noise, and their task was to type in what they had heard. An individual trial was initiated by a keypress (submission of the preceding response, or end of break). The trial began by presenting a noisy stimulus file over headphones. When the file finished playing, a response box popped up, and the participant typed in what they had heard. Listeners received a break every 15 trials. English listeners received instructions in English, and typed using English orthography. Korean listeners received instructions in Korean, and typed using Hangeul.

4.7 Scoring

Output files were concatenated and postprocessed into a data frame for subsequent analysis in R (R Development Core Team, 2011). Postprocessing consisted of separating the target sentences into language-specific units, and automatically scoring each as an observation using a custom Python script. For English, the unit was the orthographic word. For Korean, the unit was the orthographic syllable. The reason for using syllables in Korean and orthographic words in English was that these are the smallest units that can be reliably scored automatically, in each language respectively. That is, using the orthographic word in English and the orthographic syllable in Korean allows for the greatest precision with a simple automatic scoring method.

An observation was counted as correct if the target unit was contained (as a unit)
in the response after stripping away punctuation and capitalization. For example, if the target sentence was *The boy dropped his ball*, there would be 5 observations (targets: *the*, *boy*, *dropped*, *his*, and *ball*). If the response was *the boy dropped*, then the listener would be scored correct on the first 3 observations (since the response contained *the*, *boy*, and *dropped*), and incorrect on the following 2 (since the response did not contain *his* or *ball*). Similarly, if the target sentence was *그 가족이 집을 샀다* (‘The family bought a house’), there would be 8 observations (since the sentence contains 8 orthographic syllables); if the response was *그 가족이 그를 봤 다*, the listener would be scored correct on the first 4 observations and the final one (since the response contained *그*, *가*, *족*, *이*, and *다*) but not the remaining 3 (since the response did not contain *집*, *을*, and *샀*).

The automatic scoring method used here has both advantages and disadvantages as compared with scoring by hand. The primary advantages are the scoring method is fast, consistent, and replicable. The computer can compute the score in milliseconds and will not make scanning, alignment, or forgetting errors like humans do. The primary disadvantage is that our implementation of the automatic scoring method is not flexible with respect to the spelling systems of either language. Homophones (e.g. target: *dear*; response: *deer*) are not counted as correct in English even though the listener recovered the surface phonological string correctly (this is also true for Korean). Additionally, if the target contained multiple tokens of the same unit, the listener would get them all correct if they got one. The stimuli contained relatively few instances where either of these is an issue (the main repeated tokens were *the* in English and *다* in Korean). Thus, the amount of noise introduced by this method was small, and more importantly, there is no reason to believe that automatic scoring would favor bilinguals over monolinguals (or vice versa), relative to manual scoring. The authors deemed automatic scoring to be a favorable tradeoff.

5. Results and discussion

To give an overview of the most important aspects of the data, mean proportion of items correct is shown in Fig. 1 with standard errors.
Heritage bilingual talkers do not exhibit an intelligibility penalty... 419

Figure 1. Mean proportion of items correct at two noise levels, \(-4\) and \(-8\) dB SNR

There is a robust effect of noise in both languages, a monolingual/bilingual difference for both noise levels in Korean, and the lack of monolingual/bilingual difference in English at the low noise level. Later, we will demonstrate that the apparent noise-by-language-status interaction in English is marginally significant.

The significant effect of noise validates the task, demonstrates that the task is sufficiently sensitive as to replicate an expected effect. As for the question of interest, whether early language exposure has any systematic effect, the results suggest a bilingual benefit in Korean, and no difference in English. This pattern invites an explanation in terms of the specific circumstances under which the bilinguals learned and use each language. We will develop this analysis in greater depth below, but first, we briefly describe the statistical methods we use.

5.1 Analysis method: logistic mixed-effects regression

Mixed-effects regression is called 'mixed-effects' because it allows for fixed
effects (typically, the experimental manipulations of interest) and random effects (item-level, subject-level, and other factors known or believed to affect the dependent variable, but outside the research question). Mixed-effects regression allows for both continuous and discrete independent variables as well as variable nesting and interaction. Thus, unlike ANOVA and other classical statistical models, rich model structures can be fit without violating key regression assumptions (such as the assumption that residuals are uncorrelated). Because mixed-effects models can load some variance unexplained by fixed effects onto the random effects, they are considerably more powerful than classical statistical tests, where 'powerful' here is used in the statistical sense of 'more likely to find an effect'. Thus, mixed-effects regression is superior to ANOVA in terms of flexibility and sensitivity. A final virtue of mixed-effects regression models is that non-nested models can be compared using various measures of model fit, including the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These measures are equal to -2 times the data likelihood, plus a penalty for the number of free parameters. Because a change of 2 (1.96) in log-likelihood is sufficient to render a nested model comparison significant under the \( t \)-test, a difference of 4 in AIC/BIC is normally deemed weak evidence for the superiority of one model over another, and differences of more than 10 are generally considered strong evidence (Royall, 1997; for additional discussion of the merits of mixed-effects regression, see Pinheiro & Bates, 2000). Logistic regression is the proper choice when the dependent variable is binary. In logistic regression, the statistical model attempts to predict the log-odds of a correct response using a linear model. In the present case, the observations include correct/incorrect values at the level of individual words (English) or syllables (Korean), which is a binary value. That is why logistic regression was used.

In the remainder of the paper we adopt the convention that small caps refer to particular regression models, and italics refer to a particular dependent variable. For example, ‘\( snr \)’ indicates a factor in the regression analysis, while ‘SNR’ retains the customary technical meaning of signal-to-noise ratio. Separate models were run for each language, since the dependent unit being analyzed differed across languages (English: orthographic word; Korean: orthographic syllable), and because the two datasets used different talkers/listeners/items and were collected at two different laboratories in two different countries. All calculations were done using R 2.14.0 (R Development Core Team, 2011) on a ThinkPad Edge running Windows 7.
5.2 Analysis: first pass

The Baseline model was specified as follows: the fixed effects structure consisted of a single coefficient for \( \text{snr} \); and the random effects structure consisted of intercepts for \( \text{item} \) and \( \text{listener} \), with a corresponding slope for \( \text{snr} \) for each. The model also contained a random intercept for \( \text{talker} \) and a random slope for \( \text{snr} \). Note that \( \text{snr} \) was treated as a categorical variable rather than a numerical one, since the only levels were -4 and -8. That is, the random effects structure was ‘maximal’ in the sense specified by Barr, Levy, Scheepers, & Tily (2013) and recommended by Baayen, Davidson, & Bates (2008): a random intercept for each participant and item and a corresponding random slope for every combination of fixed effects. Put simply, the BASELINE model is assuming that SNR is the only manipulation of interest, and asking whether it is significant. However, it also acknowledges that there might be unexplained variance associated with each item, talker, and listener, as well as acknowledging that particular items, talkers, and listeners might be specially vulnerable to increased noise.

The BILINGUAL model differed from the BASELINE model by adding talker status \( \text{isBi} \) (talker status – monolingual or bilingual). The fixed effect structure consisted of \( \text{snr} \), \( \text{isBi} \), and an interaction coefficient. As before, the random effects structure was ‘maximal’ in the sense that there were random intercepts for \( \text{item} \) and \( \text{listeners} \), with slopes for all four combinations of \( \text{snr} \) and \( \text{isBi} \). The model also included a random intercept for \( \text{talker} \) with a random slope for \( \text{snr} \); it was not possible to cross this random effect with \( \text{isBi} \) since language background is a fixed property of each talker. Put simply, the Bilingual model is assuming that SNR and language status are the only two manipulations of interest, but it acknowledges that particular items, talkers, and listeners might be specially vulnerable to noise, and particular items and listeners might exhibit unexplained variation with respect to language status.

Table 2. Summary of logistic mixed-effects models: fixed effects and AIC/BIC. Categorical variables expressed using contrast coding, with baselines \( \text{snr}=-4 \), and \( \text{isBi}=\text{monolingual} \). Significance values – ***: \( p<0.001 \), **: \( p<0.01 \), ns: \( p>0.1 \), ms: \( 0.05<p<0.1 \).
Table 2 shows the fixed effects, significance level, and AIC/BIC for both the BILINGUAL and BASELINE models, as well as a more complex model DURATION that will be described later.

5.2.1 Method verification: the effect of noise

It is expected that intelligibility is lower when there is more noise. The inclusion of a noise manipulation was intended to validate the experimental method. In other words, if we don't see a noise effect, there is probably something wrong with the experiment. If we do see a noise effect, that proves the experiment is sufficiently sensitive as to get at least one effect that is highly expected based on all previous studies. As illustrated in Fig. 1, and confirmed by the negative snr coefficients in the Baseline and all other models, there was a significant effect of snr in the expected direction: word/syllable identification was significantly worse in the high-noise condition, for both Korean and English. The decrement in performance from a 4 dB increase in noise provides a baseline, against which potential other effects might be compared. Moreover, since the statistical analysis shows a non-null effect of noise,
Heritage bilingual talkers do not exhibit an intelligibility penalty... we conclude that the experiment is sufficiently sensitive to obtain some positive effects.

5.2.2 English

If bilinguals exhibited an overall difference in speech clarity when speaking English, this should have been evident in two ways. First, the fixed-effect coefficient $isBi$ should differ significantly from 0 in the BILINGUAL model; second, the parameter-penalized log-likelihood measures AIC and BIC should have decreased, relative to the BASELINE model. Instead, we observe no significant main effect for language and a marginally significant interaction. Moreover, the model fit improved according to the AIC (a numeric decrease), but deteriorated according to the BIC (a numeric increase), which means that the added parameters in the BILINGUAL model explain a non-negligible amount of variance, but the variance explained per datum is not necessarily sufficient to justify the added parameters. As this was a planned (confirmatory) comparison, the statistics as a whole do not give reason to reject the null hypothesis. In short, these heritage bilinguals have the same level of talker clarity as English monolinguals when they are speaking English.

5.2.3 Korean

If bilinguals exhibited an overall difference in speech clarity when speaking Korean, this should have been evident in the same ways as above. Indeed, we observe a significant main effect of bilingualism, a marginally significant interaction with SNR, and a large improvement in both AIC and BIC relative to the BASELINE model. The coefficient for $isBi$ is positive and of approximately the same magnitude as the coefficient for $snr$, indicating that at low noise the bilingual talkers were significantly more intelligible than monolingual talkers. The marginally significant interaction term was negative, and of a smaller magnitude than the main fixed effect term, implying that the 'bilingual advantage' was numerically attenuated (though still present) in the high-noise condition.

These results clearly show that in this experiment, Korean-American heritage bilinguals were more intelligible than Korean monolinguals when they were speaking Korean, but (the same) bilinguals were equally intelligible as English monolinguals
when they were speaking English. To the extent that these results generalize to other bilingual populations and communicative settings, they suggest an answer to this paper’s research question: heritage bilingualism does not cause an intelligibility penalty for talkers, in either language.

Prior to accepting this conclusion, we deemed it wise to conduct some followup analysis. We found it rather surprising that bilingual talkers would be more clear than monolinguals in Korean, the language in which they rated their own proficiency as less than monolinguals, while being equally clear as monolinguals in English, the language in which they rated their own proficiency as equal to monolinguals. Of course, it is possible that self-rated proficiency might not correlate with intelligibility, but we did not expect that lower self-rated proficiency would actually result in higher intelligibility.

One important assumption was that the speech materials were 'comparable'. The stimulus items here were all read speech, elicited in a laboratory setting. Although participants were not instructed that their speech would be mixed with noise, our expectation was that these materials would be read with moderate care. Crucially, we assumed that bilinguals would read the materials of each language with the same level of care as each monolingual group, and that bilinguals would read the English and Korean materials with the same level of care. However, in hindsight, this assumption was suspect. As noted above in the Methods section, the bilinguals were educated largely in English and generally identified English as their primary language (e.g. through self-rated proficiency). Reading Hangeul was part of the criteria for inclusion in the study, but it stands to reason that these speakers had a great deal of practice reading English out loud (for example, in American classrooms), and comparatively little practice reading Hangeul out loud. It is natural to slow down when facing an unfamiliar or unpracticed task. Thus, when early bilinguals were asked to read the Korean materials out loud, they may have been more careful than with the English materials (Goldrick, p.c.). As a result, their Korean productions may have exhibited numerous acoustic adjustments that caused a higher level of intelligibility. We tested for this possibility by measuring the vowel space and rate-of-speech in all talkers' production -- crude proxies for spectral and temporal hyperarticulation, respectively.
5.2.4 Vowel space

The vowel space measure was defined as the square root of the area of the triangle defined by the cardinal vowels /a/, /i/, and /u/ in Bark-transformed F1 by F2 space (Johnson, Flemming, & Wright, 2004). Note that we use /a/ here to refer to the low, central vowel that English and Korean both possess. Vowel measurements were done by hand, using Praat's default settings (Burg method; pre-emphasis from 50 Hz; maximum formant 5500 Hz; 5 formants; window length 0.025 s, dynamic range 50.0 dB). To reduce measurement variability, measurements were taken averaged over a window encompassing the vowel center and as much of the vowel as appeared to be steady-state.

For each vowel, 8 tokens were measured (except for English /u/, for which only 5 tokens were available). For English, stressed vowels were measured. The F1 and F2 of /a/ were measured in the following items: ball (sentence 19), boxes (sent. 15), closet (47), dog (2), dropped (1), father (11), hot (36), shopping (20). For /i/, the following items were measured: cheese (20), clean (3), clearing (23), mirror (38), police (23), sweet (49), three (54), tree (27). For /u/ the following items were measured: broom (58), fruit (5), new (51), shoe (53), shoes (32). (Note that for some readers, the vowel in mirror is lax rather than tense; however, for the purposes of comparing speaker vowel spaces, that does not really matter. As a technical matter, there is no tense/lax contrast before r so the phonological analysis is inherently uncertain. More fundamentally, the purpose is not to estimate the speaker’s /i/ but to get some estimate of their vowel space, and the same thing was done across all speakers, so the use of mirror for /i/ does not contribute any special error to just one set of speakers.) For Korean, word-initial vowels were measured (as determined by the orthographic standards of Korean). The F1 and F2 of /a/ were measured in the following items: 가족이 ‘family’-NOM (17), 갔다 ‘go’-PAST (40), 나무에 ‘tree’-LOC (27), 남동생에게 ‘brother’-DAT (52), 다리가 ‘leg’-NOM (45), 마셨다 ‘drink’-VB (57), 아버지는 ‘father’-TOP (11), 찾아다 ‘find’-VB (4). For /i/, the following items were measured: 기차가 ‘train’-NOM (56), 빨간줄 ‘broom’-ACC (58), 시계가 ‘clock’-NOM (41), 신발이 ‘shoe’-NOM (32), 씻었다 ‘wash’-VB (7), 이상한 ‘odd’ (22), 집을 ‘house’-ACC (17), 친구가 ‘friend’-NOM (14). For /u/ the following items were measured: 누군가가 ‘somebody’-NOM (42), 물이 ‘water’-NOM (21), 부러졌다 ‘be broken’-VB (45), 부엌 ‘kitchen’ (41), 수위가
‘janitor’-NOM (58), 우체부가 ‘mailman’-NOM (43), 움직인다 ‘move’-VB (56), 주고 ‘give’-GERUND (39). The F1 and F2 of each token were converted to the Bark scale using the equation $\text{Bark} = \frac{26.81}{1+1960/\text{Hertz}}-0.53$ (Traunmüller, 1990). By-speaker averages were calculated for each cardinal vowel (for bilinguals, one set for each language), and the by-speaker regression variable vowelSpace was defined as the square-root of the area of the resulting triangle.

To give the reader an idea of the vowelSpace variable, Fig. 2 shows the average F1 and F2 values for the cardinal vowels of each of the speaker groups. The vowel spaces are shown as distinct triangles for each group. The figure is shown on the Hertz scale since readers are unlikely to be familiar with typical Bark scale values for the cardinal vowels. (The Bark transformation was used for the regression because it is psychoacoustically well-motivated, and because it was expected to equate variance between F2 and F1 measurements.)

![Figure 2. Average F1 and F2 values for the cardinal vowels of each speaker group](image)

Bilinguals’ vowel space does not differ appreciably in size from monolinguals’ in either language (English: $t=0.7903$, $df=20$, $p=0.4386$; Korean: $t=0.8865$, $df=15$, $p=0.3893$). This is not the pattern we expect if spectral hyperarticulation accounts for
the bilingual advantage in Korean only. Thus, it seems likely that \textit{vowelSpace} cannot account for the Korean bilingual advantage. Indeed, during exploratory data modeling we found that \textit{vowelSpace} and interaction terms were always nonsignificant (that is, when \textit{vowelSpace} was added to or crossed with the fixed effects from the \texttt{BASELINE} and \texttt{BILINGUAL} models). Since the Korean bilingual advantage was quite sizable, even a crude measure like this one should reveal some effect, if spectral hyperarticulation was really responsible for the bilingual advantage. We conclude that spectral hyperarticulation plays at best a minor role in the bilingual advantage, which is consistent with Smiljanic & Bradlow's (2009) review demonstrating no evidence of a causal relationship between spectral hyperarticulation and talker clarity. (Note that the F2 average for English /u/ is unusually high; fronted /u/ is a well-known dialect feature of Californian English; Hagiwara, 1995; Katseff, Houde, & Johnson, 2006).

5.2.5 Rate of speech

Rate of speech was assessed using a custom Praat script which divided each production into intonational phrases, operationalized by parsing phrase boundaries at silences over 200 ms in duration. The script reported total number of phrases and the duration of each phrase. This information was combined to yield the total speaking duration for each sentence as produced by each talker. The durations were then converted to a \textit{z}-score on a by-sentence basis, so that the regression variable \texttt{zDur} represents how long a particular speaker took to articulate a sentence, relative to how long other speakers took to articulate the same sentence on average.

To give the reader an idea of how \texttt{zDur} is distributed, Fig. 3 shows a violin plot, with a separate violin for each talker group. A violin plot is a hybrid of the boxplot and the histogram. Like the boxes in a boxplot, the violins are usually oriented vertically; the top and bottom of the violin denote the range of the data after outliers are stripped out; the thick line inside the violin represents the middle quartiles, and the thick dot at the center represents the median. However, instead of a simple box, the outer border of the violin represents the density of the data; in other words, the violin shape is like a histogram that has been smoothed and flipped on its side. (Unlike boxplots, violin plots make it readily apparent if the data exhibit gross nonnormality; violins may still be plotted side-by-side for easier comparison.)
As evident from Fig. 3, bilinguals exhibited an overall slower duration than monolinguals when they were speaking Korean (but not when they were speaking English). Moreover, the difference was not driven by just a few items or speakers. The middle two quartiles hardly overlap, which implies that bilinguals employed a slower speech rate than monolinguals across-the-board. This is the pattern that we expect $z_{Dur}$ to exhibit, if it is to explain the Korean-only bilingual advantage. However, to determine whether speech rate plays a causative role, it is necessary to conduct a statistical comparison. If the apparent bilingualism advantage in Korean was caused entirely by temporal hyperarticulation, the inclusion of $z_{Dur}$ should cause the bilingualism coefficient $isBi$ to become not significantly different from 0 in Korean.

5.3 Analysis: second pass

The final regression model considered, Duration, included $z_{Dur}$ fully crossed with snr and $isBi$ as fixed effects. The random effects structure was the same as for the Bilingual model (because the model did not converge with random slopes for $z_{Dur}$). The results of that model were already reported in Table 2. As the reader
Heritage bilingual talkers do not exhibit an intelligibility penalty... 429

may find the Duration regression table somewhat opaque, Table 3 gives the predicted log-odds of a correct response (English: word; Korean: syllable) for monolinguals and bilinguals in low and high noise, at slow, moderate, and fast speech rates ($z_{Dur}$: +1, 0, -1, respectively); nonsignificant predictors are not included in the expected log-odds calculations, and the corresponding percent correct is shown in parentheses.

Table 3. Predicted log-odds of correct response in Duration model as a function of language (English, Korean), noise level (SNR: -4 dB, -8 dB), speech rate ($z$-scored duration: positive $z$-score indicates slower speech), and talker status (monolingual, bilingual). Insignificant terms not included.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th>Korean</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-4</td>
<td>-8</td>
<td>-4</td>
<td>-8</td>
</tr>
<tr>
<td>monolingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>1.69 (84.4%)</td>
<td>1.09 (74.9%)</td>
</tr>
<tr>
<td>bilingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>2.04 (88.5%)</td>
<td>1.75 (85.2%)</td>
</tr>
<tr>
<td>Moderate ($z_{Dur}=0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>monolingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>1.10 (75.0%)</td>
<td>0.25 (56.2%)</td>
</tr>
<tr>
<td>bilingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>2.02 (88.3%)</td>
<td>1.48 (81.5%)</td>
</tr>
<tr>
<td>Fast ($z_{Dur}=+1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>monolingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>0.51 (62.5%)</td>
<td>-0.59 (35.7%)</td>
</tr>
<tr>
<td>bilingual</td>
<td>1.58 (82.9%)</td>
<td>0.72 (67.3%)</td>
<td>2.00 (88.1%)</td>
<td>1.21 (77.0%)</td>
</tr>
</tbody>
</table>

In both languages there is the expected drop in log-odds in higher noise. Note that the English log-odds are constant with respect to both language status and speech rate. The Korean log-odds show a bilingual benefit across all speech rates and noise levels. The benefit is especially noticeable in faster speech and at higher noise levels. Put another way, bilinguals' talker clarity is more resistant than monolinguals' to degradation from manipulations that affect intelligibility, but only in Korean.

5.3.1 English

Fig. 3 showed that bilinguals spoke English at about the same speech rate as monolinguals did, on average. As confirmed by the nonsignificant main effect and
interaction terms for \( zDur \), speech rate variation did not contribute significantly to talker clarity. While the AIC showed some improvement in model fit relative to the Baseline model (despite the absence of significance in any coefficients not already present in that model), the BIC shows a serious deterioration in model fit. Overall, the Baseline model appears most appropriate for these data, because the only predictor which consistently explains enough variance in the English data to justify the ‘cost’ in model complexity is \( snr \). Thus, the Duration results give additional evidence that monolinguals and bilinguals speaking English are equally clear.

### 5.3.2 Korean

In contrast to the English results, there was a significant effect of speech rate in the Korean results. Specifically, as expected by the hyperarticulation account, there was a positive coefficient for \( zDur \), meaning that when talkers took longer to produce a sentence (i.e. spoke more slowly), that sentence was, on average, easier to perceive. Moreover, there was a significant improvement in model fit (both AIC and BIC), relative to both the Baseline and Bilingual models. This suggests that for the Korean data, the Duration model provides the best explanation, supporting the hypothesis that a slower speech rate did cause higher intelligibility in these talkers. However, the inclusion of \( zDur \) did not ‘explain away’ the bilingualism benefit. Even after speech rate was included in the model, bilingual talkers were still more intelligible than monolingual talkers on average.

In summary, the statistical analysis shows that bilingual talkers had higher talker clarity than Korean monolinguals, but equal talker clarity as English monolinguals. There was an interaction with speech rate, where talking more slowly is associated with greater talker clarity in Korean. The results were equivocal as to exactly how much of the bilingual benefit arose from speaking more slowly, versus how much was intrinsic to being bilingual. But some amount of bilingual advantage remained even when speech rate was controlled for statistically. Thus, at the minimum, these results show that these heritage bilinguals were never less intelligible than monolingual peers, in either language.
6. General discussion

This paper presented the results of an intelligibility study, in which English and Korean listeners heard English and Korean sentences in noise (respectively) and were asked to write down the words they heard. Intelligibility was assessed as (the log-odds of) successfully recovering a word (English) or syllable (Korean). Two manipulations were included: SNR (signal-to-noise ratio) and the language background of the talker (monolingual or bilingual). As expected, listeners correctly recovered more words (English) or syllables (Korean) in low noise (-4 dB SNR) than in high noise (-8 dB SNR) in both languages. However, there was an interaction between language and language background. Heritage bilingual talkers did not differ in intelligibility from English monolinguals when speaking English. Heritage bilinguals were more intelligible than monolingual speakers when speaking Korean. This finding was somewhat surprising, since the bilingual talkers reported their Korean proficiency to be slightly lower than their English proficiency. Follow-up acoustic measurements showed that bilinguals spoke more slowly than monolinguals when they were reading Hangeul, but spoke at the same speech rate as monolinguals in English. This might be attributed to the fact that early bilinguals were educated in English, so they have more experience reading English sentences out loud than Korean ones. Statistical modeling shows that bilinguals’ slower speech rate contributed to their greater intelligibility in Korean, but did not ‘explain away’ the monolingual/bilingual difference. Therefore, it is safe to conclude that bilingual talkers were not less clear than monolinguals, in either language.

6.1 Implications for Child Language Acquisition

We draw two implications from the present study. First, exposing a child to a second language during their formative years is unlikely to negatively impact their intelligibility in the first language. Second, native-level intelligibility can be attained as long as the child receives sufficient exposure by early childhood.

The population in the present study consisted of Korean-American heritage bilinguals, who were exposed to Korean from birth in the home and local community. All such speakers were immersed in English when they began primary school or earlier. These speakers went on to achieve a level of talker intelligibility
in Korean that is at least on par with their monolingual peers, indicating that English exposure did not harm their intelligibility as Korean talkers. In other words, bilinguals achieved a native level of intelligibility in both languages, despite rather different exposure histories.

The properties that were shared across both English and Korean histories were were some level of maintenance during adulthood, and a significant level of immersion prior to or beginning at primary school age. It is not clear whether these are necessary properties to achieve native-level intelligibility, but this study suggests that they are sufficient. In other words, to ensure that a child has native-level talker clarity in a language, one can speak it at home, or one can educate the child in the language from primary school on. It is a novel conclusion of this study that immersion as late as 6-7 years of age is sufficient to acquire native-like talker intelligibility.

6.2 Relation to existing work on bilingualism

Finally, the present study intersects with a line of work on monolingual/bilingual differences in speech production (Chang, 2012; Flege, 1991; Kang & Guion, 2006; Khattab, 2000; Oh & Daland, 2011; Sundara et al., 2006; Tahta et al., 1981). Generally such studies have found that early bilinguals exhibit the same phonological contrasts as monolinguals, but differ in fine-grained phonetic detail. The Korean vowel spaces we observed in Fig. 2 might be taken as an example of this phenomenon. The bilingual /i/ appeared to be somewhat lower and fronter than the corresponding monolingual vowel (possibly reflecting influence from the American /i/). But this example is merely illustrative – the materials were not designed for measuring vowel spaces, so a more careful study would be needed for confirmation. The point is, it is logically imaginable that fine-grained production differences would have coarse-grained perceptual consequences. The present study shows that at least one kind of coarse-grained perceptual consequence – a significant decrement/increment in talker intelligibility – did not occur in either language. While highly speculative, we would suggest that language acquisition during early childhood is constrained in some way so as to avoid talker clarity deficits. That is, even though early bilinguals exhibit fine-grained phonetic differences from monolinguals, the kind of differences that early bilinguals exhibit may be ones which do not cause clarity deficits.
Heritage bilingual talkers do not exhibit an intelligibility penalty...

References


Heritage bilingual talkers do not exhibit an intelligibility penalty...


Heritage bilingual talkers do not exhibit an intelligibility penalty...


Appendix A: The BKB sentences (English and Korean)

1. The children dropped the bag.  
아이들이 가방을 떨어뜨렸다

2. The dog came back.  
개가 돌아왔다

3. The floor looked clean.  
바닥이 깨끗해 보였다

4. She found her purse.  
그녀가 자기 지갑을 찾았다

5. The fruit is on the ground.  
과일이 땅에 있다

6. Mother got a saucepan.  
엄마는 냄비를 샀다

7. They washed in cold water.  
그들은 찬물로 씻었다

8. The young people are dancing.  
젊은이들이 춤추고 있다

9. The bus left early.  
버스가 일찍 떠났다

10. The ball is bouncing very high.  
공이 아주 높이 튀어 오르고 있다

11. Father forgot the bread.  
아버지자는 빵을 잊어 버렸다

12. The girl has a picture book.  
그 소녀는 사진책을 갖고 있다

그 소년은 자기 책을 잊어 버렸다

한 친구가 점심을 먹으러 왔다

15. The match boxes are empty.  
성냥갑이 비었다

16. He climbed his ladder.  
그는 자기 사다리를 탔다

17. The family bought a house.  
그 가족이 집을 샀다

18. The jug is on the shelf.  
항아리가 선반 위에 있다

19. The ball broke the window.  
공이 유리창을 깼다

20. They are shopping for cheese.  
그들은 치즈를 사려고 쇼핑하고 있다

21. The pond water is dirty.  
연못 물이 더럽다

22. They heard a funny noise.  
그들은 이상한 소리를 들었다

23. The police are clearing the road.  
경찰이 길을 막고 있다

24. The bus stopped suddenly.  
버스가 갑자기 멈췄다

책에 이야기가 나와 있다

26. The young boy left home.  
어린 소년이 집을 떠났다

27. They are climbing the tree.  
그들은 나무에 오르고 있다

28. She stood near her window.  
그녀는 창문 옆에 섰다

29. The table has three legs.  
그 책상은 다리가 세 개다

30. A letter fell on the floor.  
편지 하나가 바닥에 떨어졌다

31. The five men are working.  
남자 다섯 명이 일하고 있다
32. The shoes were very dirty.
33. They went on a vacation.
34. The baby broke his cup.
35. The lady packed her bag.
36. The dinner plate is hot.
37. A dish towel is by the sink.
38. She looked in her mirror.
39. The good boy is helping.
40. They followed the path.
41. The kitchen clock was wrong.
42. Someone is crossing the road.
43. The mailman brought a letter.
44. They are riding their bicycles.
45. He broke his leg.
46. The milk was by the front door.
47. The shirts are hanging in the closet.
48. The chicken laid some eggs.
49. The orange was very sweet.
50. He is holding his nose.
51. The new road is on the map.
52. She writes to her brother.
53. The football player lost a shoe.
54. The three girls are listening.
55. The coat is on a chair.
56. The train is moving fast.
57. The child drank some milk.
58. The janitor used a broom.
59. The ground was very hard.
60. The buckets hold water.