

Research Article

Increase in Speech Recognition due to Linguistic Mismatch Between Target and Masker Speech: Monolingual and Simultaneous Bilingual Performance

Lauren Calandruccio^a and Haibo Zhou^b

Purpose: To examine whether improved speech recognition during linguistically mismatched target–masker experiments is due to linguistic unfamiliarity of the masker speech or linguistic dissimilarity between the target and masker speech. **Method:** Monolingual English speakers ($n = 20$) and English–Greek simultaneous bilinguals ($n = 20$) listened to English sentences in the presence of competing English and Greek speech. Data were analyzed using mixed-effects regression models to determine differences in English recognition performance between the 2 groups and 2 masker conditions.

Results: Results indicated that English sentence recognition for monolinguals and simultaneous English–Greek bilinguals

improved when the masker speech changed from competing English to competing Greek speech.

Conclusion: The improvement in speech recognition that has been observed for linguistically mismatched target–masker experiments cannot be simply explained by the masker language being linguistically unknown or unfamiliar to the listeners. Listeners can improve their speech recognition in linguistically mismatched target–masker experiments even when the listener is able to obtain meaningful linguistic information from the masker speech.

Key Words: informational masking, linguistic masking, simultaneous bilingual

It has been reported that improvements in speech recognition occur when the languages of the target and masker speech are mismatched (i.e., the language of target speech is different than the language of the masker speech). Both Freyman, Balakrishnan, and Helfer (2001) and Tun, O’Kane, and Wingfield (2002) investigated the effectiveness of competing English and Dutch maskers for an English recognition task. Their results indicated improved speech recognition when the masker speech was spoken in Dutch (mismatched to the English target language) rather

than English (matched to the English target language). Since those initial reports, improvement in speech recognition when the target and masker speech are linguistically mismatched compared with linguistically matched has been observed for many different target and masker language combinations (see Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; Calandruccio, Van Engen, Dhar, & Bradlow, 2010; Cooke, Garcia Lecumberri, & Barker, 2008; Rhebergen, Versfeld, & Dreschler, 2005; Van Engen & Bradlow, 2007).

One hypothesis for the observed improvement in recognition is based on the listeners’ inability to understand or unfamiliarity with the masker speech in the mismatched listening condition (Cooke et al., 2008). That is, it has been assumed that listeners can improve their overall recognition score due to less confusion between the target and masker speech when they are unable to understand the linguistic content of the masker speech. More recently, however, it has been proposed that the improvement in recognition may be due to the target and masker speech being linguistically different, rather than linguistically unfamiliar. This was suggested by Van Engen (2010) and Brouwer et al. (2012). Their results indicated that, in addition to monolingual

^aQueens College of the City University of New York

^bDepartment of Biostatistics, University of North Carolina at Chapel Hill

Correspondence to Lauren Calandruccio:
Lauren_Calandruccio@med.unc.edu

Lauren Calandruccio is currently affiliated with Department of Allied Health Sciences, School of Medicine, University of North Carolina at Chapel Hill

Editor: Craig Champlin

Associate Editor: Deborah von Hapsburg

Received November 27, 2012

Revision received March 22, 2013

Accepted September 12, 2013

DOI: 10.1044/2013_JSLHR-H-12-0378

Disclosure: The authors have declared that no competing interests existed at the time of publication.

English speakers, nonnative English speakers also benefited when listening to English speech in the presence of their native language compared with competing English speech. Specifically, their English recognition performance was better when listening in the presence of competing speech that was linguistically mismatched (not English) than linguistically matched (English) speech, even though the listeners were native speakers of the language that composed the masker speech in the mismatched condition. Both reports indicated, however, that the improvement observed for the nonnative listeners was significantly reduced compared with the monolingual control group. Therefore, it remains unclear whether the observed improvement in recognition is due to (a) the target and masker speech being linguistically different, (b) the masker speech being linguistically unfamiliar, or (c) a combination of these two factors.

One way to explore whether the linguistic target–masker mismatch benefit is due to the target and masker speech being linguistically different rather than the masker being linguistically unfamiliar is to test simultaneous bilingual listeners who are proficient in two languages. Simultaneous bilinguals begin learning two languages concurrently from birth (Bialystok & Barac, 2012), typically either learning a different language from each parent (dual-language families) or through the home language and the language of a caregiver. The advantage of testing simultaneous bilinguals in this type of paradigm is that these listeners would be familiar with both the matched and mismatched masker languages, allowing an assessment of whether improvement in speech recognition is due to linguistic differences or linguistic unfamiliarity between the target and masker speech. It was predicted that, if the improvement in speech recognition associated with a linguistic mismatch was due to being unfamiliar with the linguistic content of the masker, then the monolingual group should benefit from the mismatched condition, but the simultaneous bilingual group should not. However, if both groups benefited from the mismatched condition, this would support the hypothesis that speech recognition improvement occurs due to the target and masker being linguistically different.

Method

Listeners

All listeners had clear ear canals as visualized through otoscopic examination. Hearing thresholds were tested using standard clinical procedures (American Speech-Language-Hearing Association, 2005) using a GSI-61 clinical audiometer. Hearing thresholds were < 20 dB HL bilaterally at octave frequencies between 250 and 8000 Hz for all listeners. Two separate groups of listeners were recruited on the basis of their linguistic experience. The first group of listeners included 20 (16 women and 4 men) monolingual native speakers of American English (M age = 24 years, SD = 5 years). These listeners reported that the only language they were able to speak, read, or write was English. The monolingual English group also reported having no experience with Greek.

The second group of listeners included 20 (13 women and 7 men) simultaneous bilingual speakers of American English and Greek (M age = 22 years, SD = 3 years).

English–Greek bilinguals were recruited from the greater Queens, New York, area. There is a rich history of Greek culture in Queens. It is reported that, in 2000, there were over 43,000 people reporting Greek ancestry in this county (Alexiou, 2013) and that Astoria (a subsection of Queens) had the largest concentration of Greeks outside of Greece or Cyprus (Alexiou, 2013). Greek culture and language is abundant in Queens, where one can easily find Greek churches, restaurants, diners, cafes, taverns, banks, and other services provided by Greek immigrants or Greek Americans.

All participants completed a computer-based questionnaire created by the Linguistics Department at Northwestern University (Chan, 2012) that provided linguistic and demographic information for each participant. Additional questions that are not typically included in the Northwestern University questionnaire were also asked of all English–Greek bilingual listeners to determine whether they were first-generation Greek Americans, whether their parents were able to speak any language other than Greek, and whether they considered themselves equally proficient in English and in Greek. Linguistic histories are described on the basis of five categorical groupings, including language status, language history, language stability, language competency, and demand for language usage, as described in von Hapsburg and Peña (2002). With respect to language status, all 20 listeners considered themselves to be bilingual and reported being fluent in both English and Greek. Their language history indicated that all 20 listeners began learning Greek from birth at home. All listeners were born in the United States and soon after birth were also exposed to English. All listeners reported fluency in both languages during the entire time course of typical language acquisition. Both parents of all participants descended from Greek ancestry, and all listeners attended Greek weekend school. On average, they participated in 9 years of Greek schooling (minimum 7 years of Greek school education). Nine of the 20 listeners also attended Greek dual-language education programs. Regarding language stability, all listeners reported that they reached and maintained fluency levels in both languages. All English–Greek bilinguals reported using both English and Greek languages on a daily basis throughout their entire life course. Their language competency reporting indicated that all listeners were competent in both languages, though 15 of the 20 listeners reported that English was their dominant language. All 20 listeners reported being able to read, write, speak, and listen in both languages without difficulty and reported doing all four types of tasks in both languages regularly. On average, they rated their Greek reading and writing ability as an 8 (indicating *good*) on a 10-point scale, whereas all listeners rated their English reading and writing ability as a 10 (indicating *excellent*). Last, for demand of language usage, they all reported using Greek on a daily basis. On average, they reported using Greek 25% throughout the day. They reported being in more

monolingual situations (either English or Greek) rather than bilingual situations in which they have to switch back and forth quickly between languages, as they are often surrounded by their Greek community while speaking Greek. All listeners reported speaking Greek with their parents, siblings, and other relatives; all listeners lived with their immediate family. Eighteen of the 20 listeners also reported speaking Greek with their friends on a regular basis, and five listeners reported also speaking with their coworkers.

Stimuli

All stimuli were recorded in a double-walled sound-isolated room. Talkers were comfortably seated 1 m away from a Shure SM81 condenser microphone in front of a double-paned window.

Target stimuli were taken from the Basic English Lexicon (BEL) sentences (Calandruccio & Smiljanić, 2012). These sentences use modern, American-English vocabulary and simple syntax. Each BEL sentence has four key words. There are 20 BEL sentence lists, and each list contains 100 key words. An example sentence from the BEL sentences is (with key words represented in capital letters): “These EGGS NEED MORE SALT.” Lists 4, 6, 7, and 14 were randomly chosen for speech recognition testing from the group of 16 sentence lists that have been shown to be equivalent in difficulty at a fixed signal-to-noise ratio (SNR) of -5 dB for monolingual speakers of English (Calandruccio & Smiljanić). List 2 was used to familiarize the listeners with the English recognition task.

Additional stimuli were generated to confirm that the English–Greek bilingual listener group understood Greek (providing evidence that the bilingual group should obtain significant linguistic information from the Greek masker condition). A native Greek speaker was asked to directly translate five BEL lists into Greek (Lists 1, 3, 5, 9, and 12). The instructions to the translator were to keep the meaning of the sentence the same, with the one caveat of retaining four key words per sentence. Two additional English–Greek bilinguals read the transcriptions to ensure proper translations that would be understood by the Greek American community in Queens. Lists 1, 3, 5, and 9 were used for target sentences in the Greek recognition testing. List 12 served as familiarization for the Greek recognition task. Ten English–Greek bilingual talkers (ages 19–23 years) recorded the target stimuli. An informal listening analysis by four additional English–Greek bilinguals rated each talker for their ability to speak Greek with a “native” (Greek, not American) accent. The target talker was chosen on the basis of all four raters, indicating her Greek speech production to be the most nativelike out of the corpus.

The talker of the target speech was a 21-year-old White Greek–English bilingual woman born in Queens, New York. This talker recorded both the English and Greek BEL target sentences. Both of the target talker’s parents are Greek immigrants and are linguistically proficient only in Greek. She

attended Greek school for 8 years and reported speaking Greek daily. She indicated speaking, writing, and reading in English and Greek equally well. This talker also recorded instructions for the listening task in English and in Greek. The recorded instructions, which were approximately 40 s long, were presented to the bilingual listeners prior to completing sentence recognition testing. This was done in an attempt to focus the participants’ attention to Greek prior to Greek recognition testing.

Competing speech consisted of three-talker female babble spoken in either English or Greek by simultaneous English–Greek bilinguals with similar linguistic histories as the bilingual listener group. To minimize spectral differences between the English and Greek masker conditions that can be caused by using different voices, the same three English–Greek bilingual women were used as the talkers to create both the English and Greek maskers. Three-talker babble has been shown to cause significant additional difficulty for listeners compared with listening conditions in which the target and masker speech are readily segregable (e.g., through spatial separation of the target and masker condition, see Freyman, Balakrishnan, & Helfer, 2004; by stripping the linguistic content of the masker but persevering the average spectral content and low-frequency temporal modulation pattern, see Festen & Plomp, 1990, and Simpson & Cooke, 2005; by using ideal-time frequency segregation techniques, see Brungart, Chang, Simpson, & Wang, 2006).

The English masker consisted of sentences spoken in English taken from the Harvard–Institute of Electrical and Electronics Engineers sentence lists (Rothaus et al., 1969). The same 50 sentences were translated into Greek by a native Greek speaker now living in the United States. Recordings of these materials are available upon request from the first author. The English sentences and the Greek sentences were concatenated together with no silent intervals between sentences. Care was taken so that none of the talkers were speaking the same sentence at the same time, and sentence order was organized so that no two talkers were speaking with the same sentence rhythm. To create the three-talker maskers in both languages, three 50-sentence strings were digitally combined.

Procedure

Listeners were seated in a comfortable chair in a double-walled sound-treated room. The presentation of stimuli was controlled using Cycling 74’s (2011) Max5 (Version 5.1.1) computer software. Target and masker stimuli were mixed in real time using the same custom Max5 software program running on an Apple computer. Stimuli were passed to a MOTU Ultralite input–output firewire device for digital-to-analog conversion (16 bit) and were presented binaurally via Etymotic ER1TM insert earphones, with disposable foam eartips attached.

All listeners completed sentence recognition testing in English in the presence of both English and Greek three-female

talker maskers. The presentation order of the masker conditions was randomized. Prior to participation, each subject was instructed by the target talker (via a recording). In the instructions, subjects were told that they would be listening for the target talker and that their task was to try and ignore the competing talkers and repeat back the speech they thought the target talker said. They were instructed to repeat back anything they heard, even if they did not think the sentence made sense.

Before the sentence testing, listeners were presented with the 25 sentences from BEL List 2 spoken by the same target talker to familiarize them with the task. The first five sentences were presented at an SNR of +1 dB, the next 10 sentences at an SNR of -3 dB, and the last 10 sentences were presented at an SNR of -5 dB (the same SNR used throughout testing). One sentence was presented on each trial. A random portion of the multitalker masker was also presented, beginning 500 ms prior to and ending 500 ms after the target sentence presentation. Throughout the experiment, stimuli were presented at an SNR of -5 dB for both listener groups. Target sentences were presented at a fixed overall average level of 70-dB SPL, whereas multitalker maskers were fixed at 75-dB SPL. During testing, listeners also performed additional recognition testing in the presence of two other babble masker conditions, but those results are not reported in the current article. For the experiment reported in this article, listeners were presented a total of 50 experimental sentences (one BEL list [25 sentences]/masker condition). Listeners' responses were scored online by an examiner outside of the sound-treated room based on four key words per sentence (100 key words per masker condition). Responses were also digitally recorded using a Shure SM81 condenser microphone and Felt Tip's (2011) Sound Studio (Version 4.2) audio software. A second scorer who was a native speaker of English later scored responses. A third native speaker of English reevaluated the scores that were in disagreement. This occurred in less than 2.75% of the trials.

On the same day of testing, the bilingual group also completed a Greek-in-Greek speech recognition test. The order of the English and Greek recognition testing was randomized across the bilingual listeners. That is, 10 of the bilingual listeners completed Greek recognition testing prior to English recognition testing, whereas the other 10 bilingual listeners completed English recognition testing first. The methodology was the same as that used during the English sentence recognition testing, with a few exceptions: (a) List 12 was used for familiarization, (b) the listeners were randomly presented one of the four Greek-translated BEL lists (1, 3, 5, 9) in the presence of a three-talker Greek masker, and (c) a third examiner, who was a native Greek speaker, reevaluated the recorded responses for the scores that were in disagreement between the first two examiners. This occurred in less than 3% of the trials.

All procedures were approved by the Institutional Review Board at Queens College of the City University of New York. Listeners provided written informed consent and were paid for their participation.

Results

Data were transformed into rationalized arcsine units (RAUs; Studebaker, 1985) to normalize the error variance of the performance scores. This transformation was done due to high levels of performance for several listeners for the recognition task. The following statistical analyses are based on percent correct data transformed into RAUs. However, data are displayed using percent correct and not RAUs in all figures.

English and Greek sentence recognition performance was compared for the bilingual listener group. Specifically, recognition of Greek and English speech in the presence of a linguistically matched three-talker masker was compared in two conditions: (a) English target/three-talker English masker and (b) Greek target/three-talker Greek masker. A regression analysis with subject as a random variable tested the main effect of target language (English and Greek). This analysis was based on the following regression model:

$$RAU_{ij} = \beta_0 + \beta_1 I(English) + b_{0i} + c_{ij} \quad (1)$$

where i indexes subject, j indexes target language, I is an indicator function, b_{0i} is the subject-specific random intercept, which follows $N(0, \sigma_b^2)$, and c_{ij} is the random error that follows $N(0, \sigma^2)$. It is assumed that b_{0i} is independent of c_{ij} . Parameter estimates for this regression model are shown in Table 1.

The analysis indicated no significant difference in performance for the bilingual listeners between their ability to recognize BEL sentences spoken in English or Greek in a linguistically matched masker condition (see Figure 1). These data support the assumption that the bilingual group would obtain meaningful linguistic information from the speech in the Greek masker condition.

The next analysis was conducted to determine whether there were differences in English sentence recognition between the two groups of listeners and/or the two masker languages. A mixed-effects regression model was conducted that included the main effects of listener group (monolingual or simultaneous bilingual) and masker language (English or Greek) and the interaction between these two main effects. The final results for the mixed-model analysis were based on the following regression model:

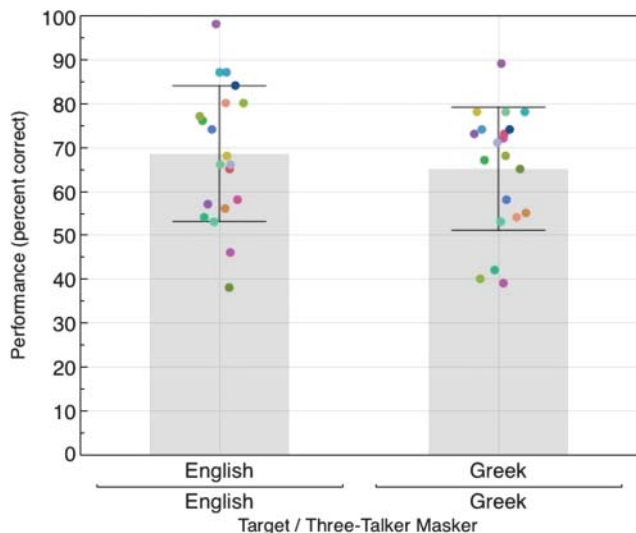
$$RAU_{ij} = \beta_0 + \beta_1 I(Monolingual) + \beta_2 I(English Masker) + b_{0i} + c_{ij} \quad (2)$$

The parameter estimates for the regression model are shown in Table 2.

Table 1. Parameter estimates for the regression model analyzing sentence recognition data from English-Greek bilingual listeners.

Effect	Estimate	SE	Prob t
Intercept	66.882	2.993	<.0001
English target	2.101	1.764	.2484

Figure 1. Mean performance of American English–Greek bilinguals for a sentence recognition task in the presence of a linguistically matched three-talker masker at a fixed signal-to-noise ratio of –5 dB. No significant difference in performance between the two target languages was found. Error bars indicate 1 standard deviation from the mean. Color-coded individual data points are also included.



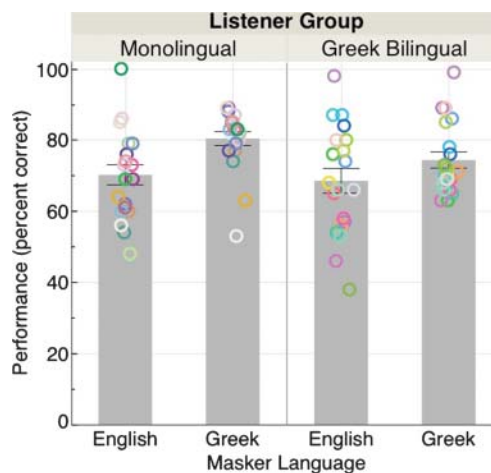
Results indicated a significant effect of masker language, in that the English masker was significantly more effective compared with the Greek masker ($p = .0023$). The main effect of the group was not significant, indicating similar performance levels between the monolinguals and simultaneous bilinguals ($p = .2913$). No significant interaction was found ($p = .4011$). Mean performance and individual performance for both listener groups in both masker languages are shown in Figure 2.

On average, listeners' speech recognition was significantly better when listening in the linguistically mismatched target–masker listening condition than the matched condition. A mismatched benefit score was calculated by subtracting each listener's individual recognition score in the mismatched condition (Greek masker) from their score in the matched condition (English masker). Statistically, both groups benefited to the same degree from the target–masker mismatch condition, though there was a trend for the monolingual group to benefit more from the mismatched condition (see Figure 3).

Table 2. Parameter estimates for the mixed-effects regression model analyzing data from monolingual English and Greek–English bilingual listeners.

Effect	Estimate	SE	Prob t
Intercept	73.95	1.89	<.0001
Listener group	2.03	1.89	.291
English masker	–3.97	1.21	.002
Listener Group \times English Masker	–1.03	0.21	.401

Figure 2. Mean sentence recognition scores for the two groups of listeners (monolingual English and simultaneous English–Greek bilinguals). A significant main effect of masker language was found. No significant differences were observed for the main effect of listener group or the interaction of masker language and listener group. Error bars represent 1 standard error from the mean. Individual color-coded data points for the 20 listeners in each group are also shown.

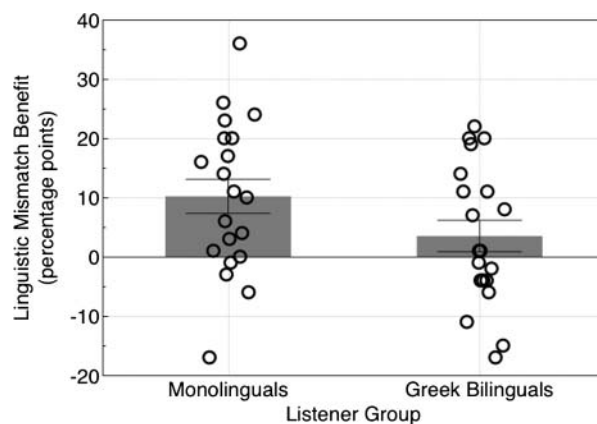


Discussion

Linguistic Mismatched Target–Masker Benefit for Monolingual and Bilingual Listeners

English sentence recognition testing was conducted for a group of monolingual and American English–Greek simultaneous bilinguals in competing English and Greek

Figure 3. On average, speech recognition improved for both listener groups when listening in the mismatched masker (Greek) than the matched masker (English) condition (i.e., there was a significant mismatch benefit for both groups). Mean linguistic mismatch benefit data are illustrated, with errors bars representing 1 standard deviation from the mean. Individual data points are also shown indicating significant variability across listeners.



three-talker maskers. Testing indicated that both groups benefited to the same degree when listening in the mismatched target–masker condition (i.e., English target–Greek masker) compared with the matched target–masker condition (i.e., English target–English masker). These data support the hypothesis that the improvement in speech recognition that has been observed here and in previous studies for linguistically mismatched target–masker listening conditions reflects the target and masker speech being linguistically different, rather than linguistically unfamiliar.

The English–Greek bilinguals were able to understand the Greek competing speech but still showed improvement in their speech recognition when listening to a linguistically mismatched target–masker condition. The idea that listeners benefit from a linguistic mismatch of the target and masker speech even when they are able to gain linguistic information from the speech of both the target and the masker is further supported by data reported in Calandruccio, Dhar, and Bradlow (2010). The data reported by Calandruccio and colleagues indicated that monolingual English speakers were more accurate at recognizing English speech when the masker was highly intelligible Mandarin-accented English than when it was native-accented English. That is, even though the highly intelligible Mandarin-accented English speech had intelligibility ratings from other monolingual listeners of ~90%, the competing Mandarin-accented speech was a less effective masker than the native-accented English speech.

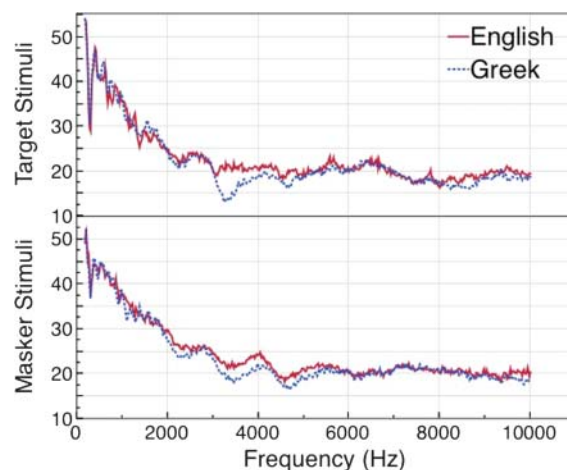
Brouwer et al. (2012) also showed a similar result, in that monolingual English listeners were better at recognizing meaningful English speech in the presence of competing speech composed of anomalous English sentences than meaningful English sentences. Though the listeners would have found individual words and phonemes equally intelligible in both masker conditions, the linguistically mismatched target–masker (meaningful–anomalous) condition allowed listeners to improve their overall recognition. Therefore, the data from Calandruccio et al. (2010), Brouwer et al. (2012), and the current project indicate that the benefit associated with mismatches in the target and masker languages are evident even when the listener understands both languages. Therefore, it seems that rather than an unfamiliarity with the masker language, the improvement in speech recognition that has been observed is more likely (or mainly) due to differences between the target and masker language along some salient set of acoustical cues. For example, different languages often have different rhythm. Though English and Greek both have a dynamic syllable stress, rhythmically Greek is different than English (Arvaniti, 2007). Perhaps differences in rhythm between languages have contributed to the linguistic-mismatch benefit that has been observed in the literature. Another possibility has to do with the difference in syllable rate between masker languages. Greek has significantly more syllables per word than English (Arvaniti, 2007). In the masker stimuli, the number of syllables per word was approximately 50% greater for Greek words than for English words in the masker speech (2.1 and 1.1 syllables per word for the Greek and English masker). In addition, the Greek speech was produced about 25% faster in terms

of the number of syllables per second (an average of an additional four syllables per second between the Greek and English maskers). Therefore, the difference in syllable rate between the English and Greek speech may have provided a cue that listeners used to improve their speech recognition in the mismatched-listening condition. Further research is needed to probe acoustical differences between the target and maskers that have elicited speech recognition improvement.

Controlling for Energetic Masking Differences Between Listening Conditions

Experiments in which differences between masker conditions assessing linguistic-mismatch benefit are investigated have to contend with differences between the masker conditions due to both linguistic differences as well as potential energetic masking confounds. In the current experiment, using the same voices to create the different masker conditions, we minimized the long-term average spectral differences across maskers. Only minor long-term average spectral differences were observed between the maskers created using English and Greek speech (see Figure 4). Freyman et al. (2001) used this technique when creating English and Dutch two-talker maskers and showed similar long-term average speech spectra (LTASS) between the two language maskers. In some experimental conditions, using the same talkers is not feasible. For example, in Calandruccio, Dhar, and Bradlow (2010), native English speakers were used to create two-talker maskers. The effectiveness of this masker was later compared with different levels of Mandarin-accented speech. In this case, the same talkers could not be used to generate the different masker conditions. In a situation like this, another possible way to control for differences in LTASS is to normalize the LTASS of each masker by the average LTASS of the masker conditions (see Brouwer et al., 2012, for a full description of performing this method).

Figure 4. Long-term average speech spectra for Greek and English targets and maskers.



In the current study, we also tried to control for differences in low-rate temporal modulations by monitoring the order in which we concatenated the sentences and the number of talkers that were used within the masker conditions. Two-talker maskers have been shown to cause the greatest amount of informational masking compared with other maskers with different numbers of competing talkers (Freyman et al., 2004; Rosen, Souza, Ekelund, & Majeed, 2013). For this experiment, three talkers instead of two talkers were used because significant differences in low-frequency temporal modulations were observed between the two-talker English and the two-talker Greek masker created during pilot testing. The inclusion of one additional talker to create a three-talker masker allowed for similar numbers of relatively long masker envelope minima between the English and Greek maskers providing similar dip-listening opportunities between the masker conditions. Specifically, we analyzed the number of relatively long masker envelope minima between the English and Greek maskers. The distribution of masker envelope values was calculated by computing the Hilbert envelope of the two three-talker speech maskers (Greek and English). The envelopes were then low-pass filtered using a second-order Butterworth filter with a 40-Hz cutoff. The result allowed for a quantitative assessment of masker envelope minima available to the listener. Similar proportions of relatively low envelope values between the Greek and English maskers indicated comparable opportunities for dip listening (Festen & Plomp, 1990) between the Greek and English masker conditions.

An additional advantage of the current experimental design included using two groups of listeners with different linguistic experiences who were able to complete the recognition task at similar performance levels. This allowed both groups to be tested using the same SNR. Hearing-impaired listeners have been shown to benefit less from masker fluctuations than normal hearing listeners (Festen & Plomp, 1990). Many explanations for this result have been proposed, such as reduced ability to “listening in the dips,” but more recently it has been attributed to increased speech recognition thresholds in stationary noise (Bernstein & Grant, 2009; Christiansen & Dau, 2012). Calandruccio et al. (2010) showed a similar effect indicating that a greater linguistic-mismatch benefit was observed at more difficult (−5 dB) rather than easier SNR (−3 dB). Both Van Engen (2010) and Brouwer et al. (2012) used different SNRs to test their native and nonnative English-speaking listeners. Because nonnative speakers of English in these studies acquired English later in life, different SNRs were necessary to equate performance levels. Van Engen used a difference as large as 8-dB SNR between listener groups, and Brouwer et al. used a difference of 2-dB SNR between groups. Therefore, the monolinguals in both studies may have had more opportunities for dip listening, allowing for a greater linguistic-mismatch benefit compared with the nonnative English listeners. In addition, the decreased linguistic-mismatch benefit observed in these previous studies compared with the current study may be related to differences in linguistic experience with the target language between their nonnative

English speakers and the native English bilingual speakers tested in the current experiment.

Simultaneous Versus Nonnative Bilinguals

Most research that investigates bilinguals and their ability to understand speech in noise has focused on listeners who speak English as a second language (Cooke et al., 2008; Stuart, Zhang, & Swink, 2010; von Hapsburg, Champlin, & Shetty, 2004). That is, their English-language abilities were acquired *after* the acquisition of their first or native language. This type of bilingual speaker has been categorized in the literature as either a sequential bilingual or, more recently, as an emerging bilingual (Garcia, 2009). The majority of data indicates a deficit for these types of bilingual listeners understanding their second language in competing noise compared with their native language (e.g., Black & Hast, 1962; Gat & Keith, 1978; van Wijngaarden, Steeneken, & Houtgast, 2002). This has been shown for even highly experienced sequential bilinguals who acquired their second language early in life (Rogers, Lister, Febo, Besing, & Abrams, 2006; Shi, 2010).

Though the projections through 2020 of linguistic diversity within the United States continue to dramatically increase (Shin & Ortman, 2011), there are relatively few data on speech perception in simultaneous bilinguals. Mayo, Florentine, and Buus (1997) included three listeners in their data set who were bilingual since infancy. However, the small sample size made it difficult to make interpretations on the basis of this grouping alone. In 2010, Shi reported data for a larger simultaneous bilingual listener group. Specifically, Shi evaluated native and nonnative target word recognition on an English sentence recognition test. In this study, data were reported for monolingual listeners and for simultaneous bilinguals (or using Shi’s definition, listeners who reported having learned both English and a second language prior to age 2 years). Using the Speech Perception in Noise (SPIN) test (Bilger, Nuetzel, Rabinowitz, & Rzezchowski, 1984; Kalikow, Stevens, & Elliott, 1977), the impact of competing speech and reverberation on the ability to recognize high- and low-predictable words within the SPIN sentences was assessed. Listeners were tested in the presence of a 12-talker babble at two different SNRs (+6 and 0 dB) and two different reverberation times (1.6 and 3.6 s). Eight simultaneous bilinguals participated in the study (five Spanish speaking, one Russian, one Greek, and one Haitian Creole). There was a nonsignificant (nine-RAU) trend for better performance in monolingual than in simultaneous bilingual listeners.

The current experiment did not include reverberation. The SNR used in the current study, however, was 5 dB more difficult than was used in Shi (2010). The more difficult SNR allowed for similar performance levels between these data and those reported in Shi. The bilinguals and monolinguals in the current study performed equally well. Some potential differences to note between our study and Shi’s study are the number of talkers used in the competing speech (three talkers vs. 12-talker babble, respectively), the sample size ($n = 20$ vs. $n = 8$, respectively), the stimuli (the BEL

sentences vs. the SPIN test), and potentially the listener population. For example, we both examined “simultaneous bilinguals”; however, all of our listeners spoke the same two languages and had similar linguistic experiences (e.g., attending Greek school). In addition, all of the English–Greek bilinguals were born in the United States. It is not clear whether the listeners in Shi were immigrants or born in the United States, which could have affected early language development. The data reported in the current article indicated that, for this task, the simultaneous bilinguals performed equally well when listening to English speech in competing speech noise compared with their monolingual English counterparts (see Figure 2, English masker condition). However, it should be emphasized that all of the Greek bilinguals who participated in this study are Americans, and 15 of the 20 bilinguals reported that English is their dominant language. Therefore, though all of the listeners learned both languages from birth, making these listeners simultaneous bilinguals, they are most likely not balanced bilinguals (i.e., they are most likely not equally proficient in both languages), as evidenced by self-reported linguistic data. It would be interesting to investigate how these simultaneous bilinguals are able to recognize Greek speech in noise compared with Greek monolinguals. That is, though these listeners are proficient in Greek, because Greek is not their dominant language, these bilinguals may not be as good as their Greek monolingual counterparts when listening to Greek in noise, especially for recognition tasks that include less common vocabulary and complex syntax. Last, the stimuli used in this experiment, the BEL sentences, were designed to minimize linguistic biases between native and nonnative speakers of English. The differences in recognition performance between monolingual and simultaneous bilingual listeners that were observed in Mayo et al. (1997) and Shi were not observed in these data. This finding may support the use of the BEL sentences for speech recognition testing when trying to eliminate or minimize linguistic biases for bilingual listeners.

Conclusion

These data add to the growing evidence that listeners can obtain a linguistic target–masker mismatch benefit to improve their overall recognition score even when the listener is able to obtain meaningful linguistic information from the matched and mismatched masker speech. The differences between the mismatched language conditions need to be further probed so that the acoustical cues that listeners are able to use to improve their speech recognition can be identified. In addition, understanding intersubject variability will be an important step forward to help determine whether listeners are using different strategies to perform these recognition tasks.

Acknowledgments

Partial support was provided by National Institutes of Health Grants R01ES021900, UL1RR025747, and P01CA142538, awarded

to Haibo Zhou. This project could not have been completed without the dedication of Anastasia (Natasha) Pashalis. We are also grateful to all of the research assistants in the Speech and Auditory Research Laboratory at Queens College, especially Rosemarie Ott, Jennifer Weintraub, Evangelia Tsirinkis, and Michelle D’Alleva.

References

- Alexiou, N. (2013). *Greek immigration in the United States: A historical overview*. Retrieved from <http://www.qc.cuny.edu/Academics/Degrees/DSS/Sociology/GreekOralHistory/Pages/Research.aspx>
- American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry* [Guidelines]. Retrieved from www.asha.org/policy
- Arvaniti, A. (2007). Greek phonetics: The state of the art. *Journal of Greek Linguistics*, 8, 97–208.
- Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 125, 3358–3372.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122, 67–73. doi:10.1016/J.Cognition.2011.08.003
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, 27, 32–48.
- Black, J. W., & Hast, M. H. (1962). Speech reception with altering signal. *Journal of Speech and Hearing Research*, 5, 70–75.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, 131, 1449–1464. doi:10.1121/1.3675943
- Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 120, 4007–4018.
- Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, 128, 860–869. doi:10.1121/1.3458857
- Calandruccio, L., & Smiljanić, R. (2012). New sentence recognition materials developed using a basic non-native English lexicon. *Journal of Speech, Language, and Hearing Research*, 55, 1342–1355. doi:10.1044/1092-4388(2012/11-0260)
- Calandruccio, L., Van Engen, K., Dhar, S., & Bradlow, A. R. (2010). The effectiveness of clear speech as a masker. *Journal of Speech, Language, and Hearing Research*, 53, 1458–1471. doi:10.1044/1092-4388(2010/09-0210)
- Chan, C. L. (2012). NU-subDB: Northwestern University Subject Database [Web application]. Department of Linguistics, Northwestern University. Retrieved from <https://babel.ling.northwestern.edu/nusubdb2/>
- Christiansen, C., & Dau, T. (2012). Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. *The Journal of the Acoustical Society of America*, 132, 1655–1666.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123, 414–427. doi:10.1121/1.2804952

- Cycling '74.** (2011). *Max5* (Version 5) [Computer software]. Walnut, CA: Author. Available from <http://cycling74.com>
- Felt Tip.** (2011). *Sound Studio* (Version 4.2) [Computer software]. New York, NY: Author. Available from <http://felttip.com/ss/>
- Festen, J. M., & Plomp, R.** (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88, 1725–1736.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S.** (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109, 2112–2122.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S.** (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115, 2246–2256.
- Garcia, O.** (2009). *Bilingual education in the 21st century: A global perspective*. West Sussex, United Kingdom: Wiley-Blackwell.
- Gat, I. B., & Keith, R. W.** (1978). An effect of linguistic experience: Auditory word discrimination by native and non-native speakers of English. *Audiology*, 17, 339–345.
- Kalikow, D. N., Stevens, K. N., & Elliott, L. L.** (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, 61, 1337–1351.
- Mayo, L. H., Florentine, M., & Buus, S.** (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686–693.
- Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A.** (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, 118, 1274–1277.
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B.** (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, 27, 465–485. doi:10.1017/S014271640606036x
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A.** (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133, 2431–2443.
- Rothausser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M.** (1969). IEEE recommended practices for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17, 225–246.
- Shi, L. F.** (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech, Language, and Hearing Research*, 53, 821–835. doi:10.1044/1092-4388(2010/09-0081)
- Shin, H. B., & Ortman, J. M.** (2011, April 21). *Language projections: 2010 to 2020*. Paper presented at the Federal Forecasters Conference, Washington, DC.
- Simpson, S. A., & Cooke, M.** (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, 118, 2775–2778.
- Stuart, A., Zhang, J., & Swink, S.** (2010). Reception thresholds in quiet and noise for monolingual English and bilingual Mandarin-English listeners. *Journal of the American Academy of Audiology*, 21, 239–248.
- Studebaker, G. A.** (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, 28, 455–462.
- Tun, P. A., O’Kane, G., & Wingfield, A.** (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17, 453–467.
- Van Engen, K. J.** (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, 52, 943–953. doi:10.1016/j.specom.2010.05.002
- Van Engen, K. J., & Bradlow, A. R.** (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121, 519–526.
- van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T.** (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *The Journal of the Acoustical Society of America*, 112, 3004–3013.
- von Hapsburg, D., Champlin, C. A., & Shetty, S. R.** (2004). Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *Journal of the American Academy of Audiology*, 15, 88–98.
- von Hapsburg, D., & Peña, E. D.** (2002). Understanding bilingualism and its impact on speech audiometry. *Journal of Speech, Language, and Hearing Research*, 45, 202–213.