Non-native perception of regionally accented speech in a multitalker context

Robert Allen Fox, Ewa Jacewicz, Florence Hardjono

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA.

jacewicz.1@osu.edu, fox.2@osu.edu, hardjono.4@osu.edu

Abstract

Noisy listening conditions are challenging to non-native listeners who typically perform poorly while attending to several competing talkers. This study examined whether non-native listeners are able to utilize dialect-related cues in the target and in the masking speech, even if they do not reach the proficiency level of the native listeners. 35 Indonesian-English bilinguals residing in the United States were presented with speech stimuli from two American English dialects, General American English and Southern American English, which were systematically varied both in the target sentences and in 2-talker masking babble at three sound-to-noise ratios (SNR). We found that the non-native listeners were (1) sensitive to dialect-specific phonetic details in speech of competing talkers and (2) performed in a manner similar to native listeners despite their apparent deficit. However, their performance differed significantly when the speech levels of the competing talkers were equal (0 dB SNR). The differential sensitivity of non-native listeners may reflect their inability to separate utterances of competing talkers when there is not enough contrast in their voice levels. In turn, the lack of sufficient contrast may reduce their ability to benefit from the phonetic-acoustic details necessary to encode the signal and comprehend a message.

Index Terms: bilingual, regional accent, speech intelligibility, speech-on-speech masking, American English, Indonesian.

1. Introduction

Multitalker environments are challenging to speech intelligibility. The challenge comes from the interaction of speech and voice characteristics of several competing talkers, which can effectively degrade portions of the target utterances to make them less distinct and more confusable with those of the interfering talkers. While such noisy listening conditions are unfavorable for everyone, they are even more demanding for second language (L2) users. Indeed, proficient L2 listeners are more adversely affected than native listeners and reportedly score between 10 and 30 percentage points worse than the latter [1]. The native language advantage increases in more difficult conditions as the relative level of the masking talkers (in the background) increases.

Research has repeatedly shown that native listeners outperform non-native listeners in experiments involving various combinations of target and masking conditions [2, 3, 4]. The deficit for non-native listeners is primarily attributable to their impoverished knowledge of L2 and interference from their native language (L1) [1] although their reduced performance may also be affected to some extent by their slower processing under increased cognitive load [5]. As is the case with virtually all aspects of L2 acquisition, performance under less favorable listening conditions also improves with increased experience with L2 [4].

One aspect of the apparent deficit of non-native listeners in their ability to understand L2 speech in noisy environments can be related to their limited knowledge of or experience with regional accent variation. Recent experimental research has found that native listeners are aware of dialect-specific phonetic details in speech and are sensitive to disruptions in the expected amount of such variations [6, 7]. The current study examined whether non-native listeners are able to utilize dialect-related cues in the target and in the masking speech, even if they do not reach the proficiency level of the native listeners. The study had two related aims. The first aim was to establish whether, in the competing talker conditions, non-native listeners are sensitive to fine-grained acoustic phonetic details that are present in dialect-specific pronunciation patterns. The second aim was to ascertain whether, when listening conditions deteriorate, the pattern of responses of the non-native listeners responding to specific target-masker combinations parallels that of the native listeners.

These aims arose in part from our previous findings with respect to intelligibility of American English dialects [6]. It was found that listeners from the Midwest who spoke a general variety of American English (GAE) showed higher intelligibility benefit when the target speech was in Southern American English (SAE) and not in their own dialect. The SAE advantage increased as listening conditions deteriorated, which indicates that certain acoustic-phonetic features of the Southern speech caused it to sound more distinct and easier to understand. This was true regardless of whether the target speech was produced by multiple talkers [8] or by one carefully selected representative talker of each dialect [9]. This outcome is contrary to other reports in the literature which found that familiarity and experience with a dialect improves its processing and comprehensibility [10, 11]. Another significant finding was that, in more favorable listening conditions, listeners performed most poorly when both the target and the masker were in their own dialect. Presumably, this matched target/masker combination (GAE/GAE) created more interference at the acoustic-phonetic level of signal processing, which has also been reported for target/masker combinations that shared the same language [12].

Presenting non-native listeners with the two regional variants of American English, GAE and SAE, is an important extension of the current line of research because non-native performance can verify the relative salience of SAE features. The native GAE listeners in [8, 9] were young college students from central Ohio associated with The Ohio State University with no exposure to SAE other than during casual encounters with regional variation through travel or mass media. The non-native listeners tested in the current study are Indonesian-English bilinguals who first learned English at school in Indonesia and later moved to the United States to pursue college education and professional careers. The majority of them studied at The Ohio State University and have lived in central Ohio since their arrival in the US. Their exposure to SAE was no different than that of the native listeners, which creates the most desirable condition for

comparing performance of the two groups using the same testing material.

We expect these bilinguals—who speak Indonesian-accented English and have heard GAE as the most common variety of spoken American English—to be familiar with the phonetic features of GAE, at least as practicable by a non-native speaker. Ruling out their familiarity and experience with SAE, we are left with two possibilities with respect to their predicted performance. First, if certain acoustic-phonetic features of SAE contribute to SAE intelligibility benefit, then the *pattern* of responses of the non-native listeners will be similar to that of the native GAE listeners across all testing conditions. These results will be interpreted as indicating that non-native listeners are sensitive to fine phonetic details in L2. The second possibility is that non-native listeners will perform better when responding to the GAE variety due to their familiarity and experience with this dialect. This outcome would indicate that listening to L2 speech is more selective and acoustic-phonetic cues that are salient to native listeners are mostly ignored. Though unlikely, a third possibility also exists in that non-native listeners will show no preference for either variety, suggesting their lack of sensitivity to regional variation.

2. Methods

2.1. Listeners

35 Indonesian-English bilinguals participated (17 male, 18 female), ranging in age from 22 to 47 years (M = 33.7, s.d. = 7.0). All had been born and had lived in Indonesia prior to their arrival in the US (15 were from Jakarta, 4 from Bandung and the remaining were from other towns). All were fluent in Bahasa Indonesia at the time of testing. In terms of their experience with L2 English, they first learned English in a school setting in Indonesia, beginning from elementary school or even preschool, and later came to the US to attend college. All had at least an undergraduate degree from US institutions, mostly from The Ohio State University. Their exposure to spoken GAE began in college and has continued since then. As is common among immigrant populations, they spoke Indonesian daily at home and with friends, and English at work and outside of their home environment. All were either professionals (engineers, accountants, financial analysts, consultants) or PhD students at The Ohio State University. Based on their detailed background questionnaires collected at the time of testing, they had not traveled extensively in the US nor spent a considerable amount of time in the South.

2.2. Stimulus materials

The same stimulus materials as in [8] were used in the current experiment. Only the essential description is provided here and more details about stimulus creation can be found in [8]. 96 target sentences were selected from the Revised Bamford-Kowal-Bench Standard Sentence Tests [13]. The sentences were read by four middle-age male GAE talkers from Columbus, Ohio and by four middle-age male SAE talkers from Sylva, North Carolina. The talkers were matched for speech tempo and fundamental frequency (F0). The target sentences were mixed with a 2-talker babble, which was created from recordings of spontaneous speech selected from a large corpus. The talkers in the babble included four men (two for GAE and two for SAE) who were of comparable origin, age, speech tempo and F0 as the talkers who read the target sentences. For presentation in the experiment, the level of the masking babble was adjusted relative to the fixed level of the target sentence to create three sound-to-noise ratios (SNRs): +3 dB (the level of the babble was 3 dB less than the level of the target), 0 dB (both levels were equal) and -3 dB (the level of the babble was 3 dB more than the level of the target). These three SNRs represented moderate (+3 dB and 0 dB) and difficult (-3 dB) listening conditions [10].

2.3. Experimental procedure

Each listener responded to 96 target sentences, evenly divided in 3 experimental blocks corresponding to the 3 SNR levels. The order of experimental blocks was the same for each listener, proceeding from the easiest (+3 dB SNR) through intermediate (0 dB SNR) to the most difficult (-3 dB SNR). In each block, half of the target sentences were read by GAE talkers (4 x 4) and half by SAE talkers (4 x 4). For each dialect, half of the sentences were masked by GAE babble (8 x 2) and half by SAE babble (8 x 2). All 32 sentences in each block were presented in a random order. Upon hearing each sentence, the listeners typed what they heard as the target sentence into a textbox on the computer screen. No repetitions were allowed. The experiment was self-paced. Listeners were tested in a sound-attenuating booth. Sound was delivered diotically over Sennheiser HD 600 headphones at a comfortable listening level. None of the listeners was told about dialect variation in the stimulus speech. To familiarize them with the experiment, 8 practice sentences at +3 dB SNR were first presented. These sentences were different from those used in the experiment.

2.4. Scoring

The digitally stored responses were initially scored adopting the keyword scoring system developed for native listeners [14, 8]. However, based on the error analysis from the current non-native listeners, it became apparent that specific types of errors were predictable due to L1 interference. That is, under increased cognitive load, Indonesian-English bilinguals sometimes defaulted to their L1 grammar when typing English sentences to indicate plural marking, past tense, subject-verb agreement and personal pronouns (each of which are not marked morphologically in Bahasa Indonesia). We then developed a modified version of the original scoring and rescored their responses counting the predictable errors as correct responses. For example, the target sentence "**He listened to** his **father**" (4 keywords in bold) was accepted as correct when spelled "**He listen to** his **father**" because there is no past tense in Indonesian and the concept of the past tense could have been lost in a task that was cognitively taxing. The modified scoring resulted in a modest overall improvement of about 3 percentage points but we felt that this correction is appropriate to account for a possible deleterious interaction of auditory and linguistic factors in L2 users. For each type of scoring, reliability was done on the entire data set by a second experimenter and any discrepancies were resolved prior to data processing.

2.5. Results

Raw scores for each listener were first converted to percent correct and then to rationalized arcsine units, or RAUs [15] to ensure valid assessment of differences across the entire range of the scale after normalizing for ceiling and floor effects. Repeated-measures ANOVA was used to analyze the arcsine

transformed scores with SNR level, target dialect and masker dialect as the within-subject factors. Pairwise post hoc *t*-tests (with Bonferroni adjustment) were used to explore the nature of significant main effects and interactions.

Figure 1 shows the overall mean scores for the bilingual listeners relative to the native GRE listeners reported in [8]. Shown are means for each type of scoring (i.e., original and modified). Not surprisingly, the performance of non-native listeners was comparatively worse at each SNR and the magnitude of the difference was similar across the levels. On average, the non-native listeners scored about 25 percentage points worse than the native listeners. The modified scoring was of benefit mostly at moderate listening conditions and did not improve their scores at the most difficult -3 dB SNR level.

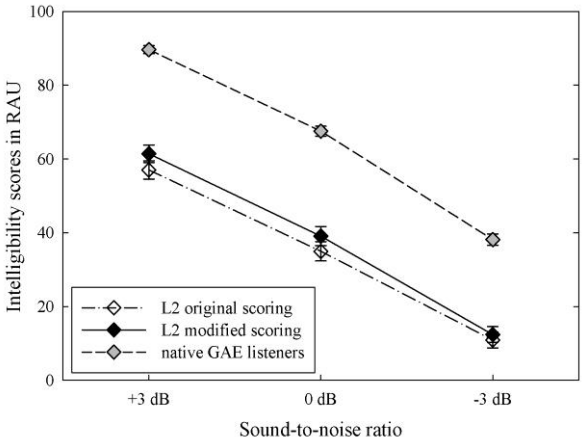

Figure 1: *Means (s.e.) for native GAE and L2 listeners.*

2.5.1. Results for the non-native (bilingual) listeners

We now present the results for the modified scoring only and the results for the original scoring will not be further discussed. Bilingual listeners' responses to all target and masker combinations are displayed in Figure 2. As is evident, intelligibility of the target decreased as the level of the masker increased. The main effect of SNR level was significant [$F(2,68)=403.5$ $p<.001$] and post hoc tests showed that performance at all three levels differed significantly from one another.

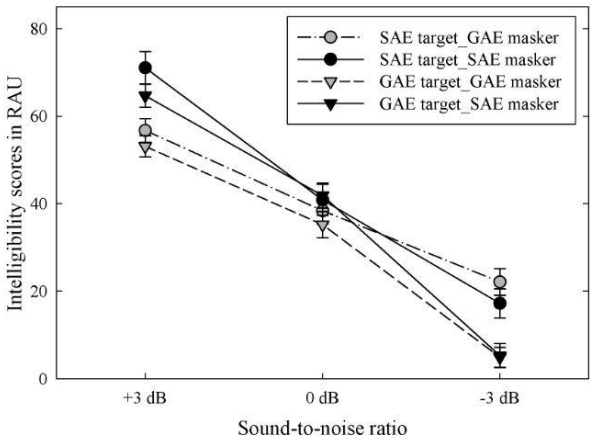

Figure 2: *Means (s.e.) for target and masker at 3 SNRs.*

The significant main effect of target dialect [$F(1,34)=29.1$, $p<.001$] indicated higher intelligibility of the SAE target than the GAE target. However, a significant interaction between SNR and target dialect [$F(2,68)=12.0$, $p< .001$] arose because intelligibility benefit of the SAE target was only at the least and the most difficult SNRs and ceased at 0 dB SNR. This interaction is illustrated in Figure 3. As can be seen, the difference between intelligibility of the SAE target and the GAE target was greatest in the most difficult condition at -3 dB SNR and the SAE target advantage was relatively smaller at +3 dB SNR.

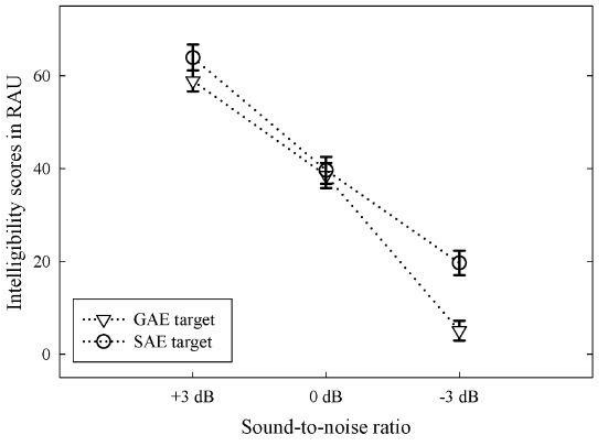

Figure 3: *Means (s.e.) for target dialect at 3 SNRs.*

The significant main effect of masker dialect [$F(1,34)=12.4$, $p=.001$] indicated that GAE was a more effective masker than was SAE. This effect persisted only in moderate listening conditions, however, and dialect in the masker had no influence on intelligibility in the most difficult condition, which was the locus of a significant SNR x masker interaction [$F(2,68)=16.9$, $p<.001$]. This interaction, shown in Figure 4, also revealed that the difference in the effectiveness of the masker was comparatively greater at +3 dB SNR. The remaining interactions in the ANOVA analysis were not significant.

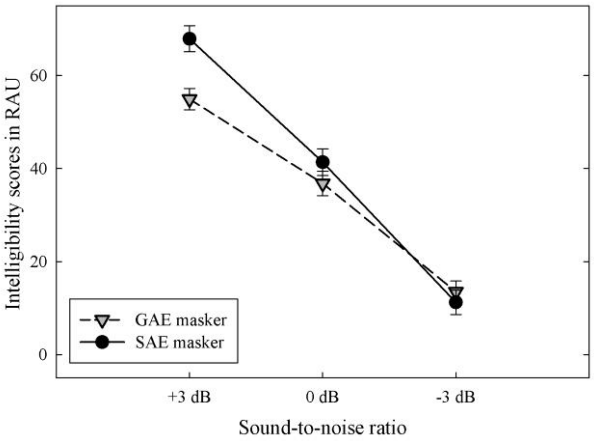

Figure 4: *Means (s.e.) for masker dialect at 3 SNRs.*

2.5.2. Native versus non-native performance

We now examine whether the pattern of responses of the bilingual listeners parallels that of the native GAE listeners. To allow for the comparison, the intelligibility scores for GAE listeners reported in [8] were redrawn here and the patterns for both listener groups are displayed side by side in Figure 5, separately for each SNR level.

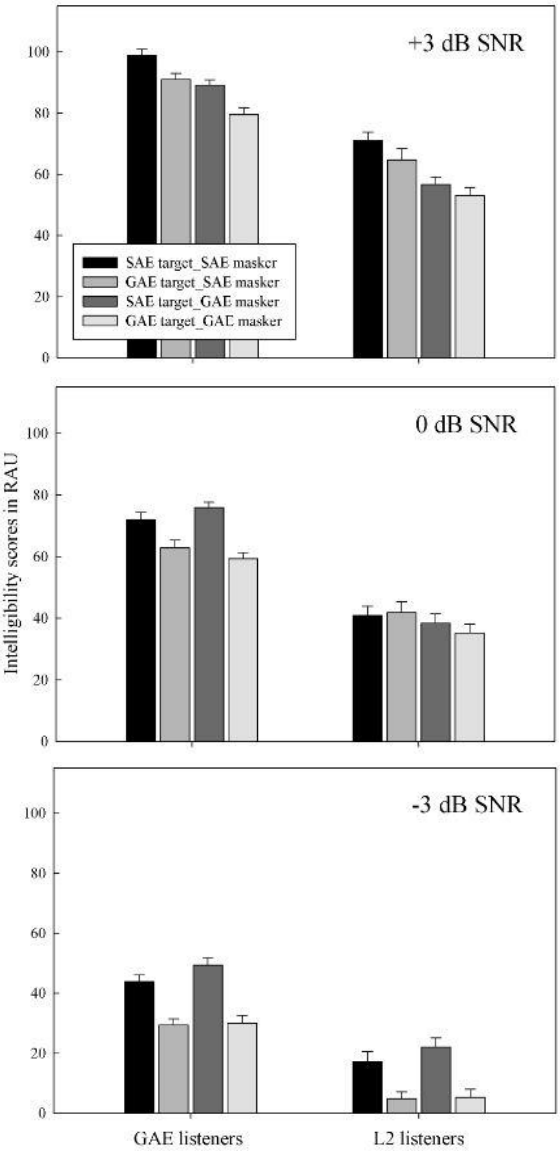

Figure 5: *Means (s.e.) for GAE listeners and L2 listeners.*

It is apparent from Figure 5 that both groups performed similarly at +3 dB SNR and -3 dB SNR and their responses differed primarily at 0 dB SNR. In particular, at +3 dB SNR, both groups performed best when the SAE target was masked by SAE babble and worse when the GAE target was masked by GAE babble. When listening conditions deteriorated at -3 dB SNR, both groups responded better to the SAE target than to the GAE target and the scores were slightly higher when the masker was in GAE. The patterns were clearly different at 0 dB SNR. The bilingual listeners were unaffected by dialectal variation in either target or masker speech whereas the native GAE listeners still benefitted when the target was in SAE. In general, the pattern of their responses at 0 dB SNR was similar to that at -3 dB SNR.

3. Discussion

The current study had two aims. First, it sought to establish whether non-native listeners are sensitive to fine-grained acoustic phonetic details in dialect-specific pronunciation patterns under challenging listening conditions. The second aim was to determine whether the pattern of non-native listeners' responses is similar to that of native listeners, especially when listening conditions deteriorate. The current results provide evidence that non-native listeners are both sensitive to phonetic details in speech of competing talkers and are also able to perform in a manner similar to native listeners despite their apparent deficit in L2 speech recognition in noisy background, which was also confirmed in this study.

It is evident that acoustic-phonetic features of SAE were salient for the bilingual listeners—as they were salient for the native listeners—and contributed to SAE intelligibility benefit, which we interpret as indicating that non-native listeners are able to detect fine phonetic details in L2. An unexpected finding was that their responses to the target-masker combinations examined here differed from those of the native listeners at 0 dB SNR, when the voice levels of all competing talkers were equal. This differential sensitivity to L2 at 0 dB SNR may guide future experiments aimed on defining the nature of the L2 speech processing deficit. It may help to better understand how L2 listeners process L2 speech, that is, how they encode the auditory input and how they decode the linguistic message in the presence of competing talkers. It is possible that non-native listeners cannot effectively separate utterances of competing talkers when there is not enough contrast in their voice levels. The lack of a sufficient contrast may reduce their ability to benefit from the phonetic-acoustic details in order to "follow" a particular talker and comprehend a linguistic message.

The current results are suggestive of this possibility in light of the pattern at +3 dB SNR. In particular, the GAE target/masker combination may not have provided enough contrast for L2 listeners who were most accustomed to hearing GAE speech. This could be one explanation for why their performance was poorest in this condition. They performed better when there was a mismatch between target and masker dialect and best when both target and masker were in SAE, which was acoustically distinct from GAE. The GAE target/masker combination was also detrimental to the native GAE listeners who, in general, seemed to benefit from a mismatched target/masker dialect conditions. Apparently, the mismatch provided a greater acoustic contrast which enhanced intelligibility also for native listeners.

This particular outcome is also in agreement with studies which used native and non-native (or foreign-accented) speech combinations and showed that native English listeners performed poorly in the presence of the English masker but their performance improved when the target an masker speech we linguistically different [14, 16]. The current study shows that similar effects apply also to the dialect of the same language.

4. Conclusions

Non-native listeners are sensitive to phonetic details in speech of competing talkers as evidenced in their responses to regional dialect variation. They are also able to perform in a manner similar to native listeners although their intelligibility scores under challenging listening conditions are comparatively lower. Their ability to attend to fine-grained acoustic details in regional accents seems to decline when the speech levels of the competing talkers are equal.

5. References

[1] Cooke, M., Garcia Lecumberri, M. L., and Barker, J., "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception", Journal of the Acoustical Society of America, 123: 414–427, 2008.

[2] van Wijngaarden, S. J., Steeneken, H. J., and Houtgast, T., "Quantifying the intelligibility of speech in noise for non-native talkers", Journal of the Acoustical Society of America, 112: 3004-3013, 2002.

[3] Shi, L. F., "Perception of acoustically degreaded sentences in bilingual listeners who differ in age of English acquisition", Journal of Speech, Language and Hearing Research, 53: 821-835, 2010.

[4] Pinet, M., Iverson, P., and Huckvale, M., "Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity", Journal of the Acoustical Society of America, 130: 1653-1662, 2011.

[5] Clahsen, H, and Felser, S. "How native-like is non-native language processing?", Trends in Cognitive Science 10: 564-570, 2006.

[6] Jacewicz, E., and Fox, R. A. "The effects of cross-generational and cross-dialectal variation on vowel identification and classification", Journal of the Acoustical Society of America, 131: 1413-1433.

[7] Clopper, C. G., and Bradlow, A. R., "Perception of dialect variation in noise: Intelligibility and classification", Language and Speech, 51: 175-198, 2008.

[8] Jacewicz, E., and Fox, R. A. "The effects of dialect variation on speech intelligibility in a multitalker background", Applied Psycholinguistics, doi: 10.1017/S0142716413000489, 2014.

[9] Jacewicz, E., and Fox, R. A. "Regional accents affect speech intelligibility in a multitalker environment", ISCA Interspeech, Lyon, France, 2081-2085, 2013.

[10] Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K., "Comprehension of familiar and unfamiliar native accents under adverse listening conditions", Journal of Experimental Psychology: Human Performance and Perception, 35: 520-529, 2009.

[11] Sumner, M., and Samuel, A. G. "The effect of experience on the perception and representation of dialect variants", Journal of Memory and Language, 60: 487-501, 2009.

[12] Calandruccio, L., Dhar, S., and Bradlow, A. R., "Speech-on-speech masking with variable access to the linguistic content of the masker speech", Journal of the Acoustical Society of America, 128: 860-869, 2010.

[13] Bench, J., Kowal, A., and Bamford, J., "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology", 13: 108-112, 1979.

[14] Van Engen, K. J., and Bradlow, A. R., "Sentence recognition in native- and foreign-language multi-talker background noise", Journal of the Acoustical Society of America, 121: 519–526, 2010.

[15] Studebaker, G., "A 'rationalized' arcsine transform. Journal of Speech, Language, and Hearing Research", 28: 455–462.

[16] Calandruccio, L., and Zhou, H. "Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance", Journal of Speech, Language and Hearing Research, in press, 2014.