Effect of Speech-Shaped Noise on the Recognition of Malayalam and English Consonants by Malayalam Listeners

Mohan K. Kalaiah, Jayashree S. Bhat, Usha Shastri
Department of Audiology and Speech Language Pathology, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, Karnataka, India

Abstract

Objective: The present study compared recognition of native and non-native consonants in quiet and noise among native speakers of Malayalam. Methods and Material: Fifteen native speakers of Malayalam who had English as the medium of instruction at school participated in the study. Stimuli comprised of 16 vowel-consonants-vowel nonsense syllables spoken by eight native speakers of Malayalam (native consonants) and eight native speakers of American English (non-native consonants). Recognition of native and non-native consonants was studied in quiet and in the presence of speech-shaped noise at signal-to-noise ratios (SNRs) of 8 dB, 0 dB, and −8 dB. The consonant recognition task was carried out as 16-alternative forced-choice procedure, and the responses were stored as confusion matrix. Results: In favourable listening condition (i.e., quiet and 8 dB SNR), the recognition score for native consonants was greater than non-native consonants. In contrast, at 0 dB SNR and −8 dB SNR, the recognition score of non-native consonants was greater than native consonants. Information transfer analysis revealed that the transfer of information was highest for consonant feature manner of articulation and lowest for voicing, across listening conditions for both native and non-native consonants. Conclusions: Recognition of native and non-native consonants were affected differently in the presence of speech-shaped noise among native speakers of Malayalam. In favourable listening condition, recognition of native consonants was better than non-native consonants. However, in challenging listening condition, non-native consonants were found to be recognised better than native consonants.

Keywords: Bilingualism, consonant recognition, noise, non-native, speech recognition

Key Messages:
The present study shows a differential effect of listening conditions on the recognition of non-native consonants compared to native consonants. Native consonants were recognised better than non-native consonants in favourable listening conditions. In challenging listening conditions, non-native consonants were found to be recognised better than native consonants. Poorer recognition of non-native consonants was due to greater difficulty for the recognition of voicing information in non-native consonants, among native Malayalam speakers.

INTRODUCTION

Speech recognition is a “complex phenomenon which involves processing the auditory aspects of the signal as well as language based processing of the information”.[1] Several factors affect the recognition of speech, which includes the degree of redundancy in the stimulus, the presence of background noise and its spectral structure, signal-to-noise ratio (SNR), the age of listener, the rate of presentation of stimuli, knowledge of the language, and cognitive and linguistic abilities of listeners.[2] In ideal listening conditions, speech recognition takes place with little effort and is considered to be automatic. However, in adverse listening conditions such as in the presence of noise and reverberation, the recognition of speech would be challenging. Further, studies have shown that the
recognition of speech was significantly affected among non-native listeners, especially in adverse listening conditions. Poorer speech recognition among non-native listeners has been attributed to the lack of ability to make use of the contextual information available in the speech signal.

Investigations to understand speech recognition difficulties of non-native listeners are not new in the field of speech perception. Nabelek and Donahue compared the recognition of English consonants between native and non-native listeners in quiet and in the presence of reverberation (reverberation time of 0.4, 0.8, and 1.2 s). In quiet, the recognition score was equivalent for both native and non-native listeners. However, in the presence of reverberation, the recognition score of non-native listeners deteriorated rapidly with an increase in the reverberation time. Garcia Lecumberri and Cooke compared the recognition of English consonants between native and non-native listeners in quiet and in the presence of noise. In both quiet and in the presence of noise, the recognition score for English consonants was affected among non-native listeners. Further, in the presence of noise, the recognition score was significantly reduced compared to the quiet listening condition. Other investigators have reported similar findings. Cooke and Garcia Lecumberri compared the recognition of British English consonants in the presence of noise between native speakers of British English and native speakers of other European languages. Among speakers of various European languages, the recognition score for English consonants was dependent on the native language of participants.

From the findings of above investigations, it is evident that non-native speakers of English have significant difficulty for the recognition of English consonants in the presence of noise, and the degree of difficulty increases with the level of noise. Further, the extent of difficulty for the recognition of English consonants among non-native speakers of English was dependent on the native language of participants. Although many studies have investigated the recognition of English consonants among speakers of various languages, participants in those investigations were mostly speakers of European languages. There is a dearth of studies investigating the recognition of English consonants among speakers of Indian languages. To the best of our knowledge, there are no published studies which have investigated the recognition of English consonants among speakers of Indian languages. Therefore, the present study was carried out to investigate the recognition of English consonants (i.e., consonants spoken by native speakers of Malayalam) by native speakers of Malayalam in quiet and in the presence of noise. Malayalam and English consonants will be referred as native and non-native consonants respectively in the present study. The present study also investigated the effect of SNR on the recognition of Malayalam (native) and English (non-native) consonants.

**SUBJECTS AND METHODS**

**Participants**

Fifteen native listeners of Malayalam (3 males and 12 females) aged between 18 and 23 years (mean age = 19.6 years) participated in the study. All the participants had hearing sensitivity within normal limits in both ears, with pure-tone thresholds less than 15 dB HL at octave frequencies between 250 Hz and 8000 Hz. Immittance evaluation showed normal middle ear functioning, with “A” type tympanogram and acoustic reflex thresholds at normal levels for all the participants. None of the participants had otologic or neurologic problems, exposure to hazardous noise or ototoxic medication, and difficulty understanding speech in the presence of noise. All the participants had English as the medium of instruction at school for at least ten years. Proficiency for understanding English was assessed using Language Experience and Proficiency Questionnaire (LEAP-Q). Self-rated proficiency of all participants was greater than seven, which suggests good to excellent proficiency. The study protocol was approved by Institutional Ethics committee at Kasturba Medical College, Mangalore. Informed consent was obtained from all participants before their participation in the study.

**Stimulus**

Stimuli comprised of 16 consonants (/p/, /t/, /k/, /b/, /d/, /ɡ/, /ʃ/), /l/, /l/, /ɾ/, /ɾ/, /ɾ/, /ɾ/, /ɾ/, /ɾ/) in intervocalic context (VCV), the vowel preceding and following the consonant was always /a/ (e.g., /apa/, /aba/ etc.). These non-sense syllables were produced by native speakers of Malayalam and American English. Eight native speakers of Malayalam (five females and three males) produced non-sense syllables, and the utterances were recorded using Computerized Speech Lab (CSL) model 4150 version 3.2.1. (KAY Elemetrics). The utterances were digitally recorded with a sampling rate of 44100 Hz and 16-bit analog-to-digital converter. All the recordings were carried out in a sound-treated quiet room. All the utterances were independently reviewed by two experts to ensure intelligibility of each stimulus, and those syllables which received poor intelligibility judgment were replaced with new recordings. Non-sense syllables spoken by native speakers of American English were obtained from VCV corpus recorded by Shannon et al.

For investigating consonant recognition in noise, the recorded syllables were mixed with speech-shaped noise to obtain SNRs of 8 dB, 0 dB, and ~8 dB SNR. To generate speech-shaped noise, the VCV syllables spoken by all the talkers
were normalised to the same root mean square (RMS) level, and averaged spectrum was obtained separately for Malayalam and English consonants using all the syllables. Using the averaged spectrum, a finite impulse response function was created, and white noise was passed through the impulse response function to obtain speech-shaped noise. The desired SNR was obtained by adjusting the RMS level of noise with reference to the RMS level of VCV syllables.

**Procedure**

Consonant recognition task was carried out in quiet and in the presence of noise as a closed set identification task. It was carried out as 16-alternative forced-choice procedure. Participants were instructed to identify the consonants in nonsense syllables and provide a response by clicking the button labelled with corresponding consonant sound, shown on the computer screen. Once the response was obtained, the next syllable was presented after a short pause of 1.5 sec. The consonant recognition task was always carried out in the quiet listening condition first and then in the presence of noise. The order of presentation of syllables was randomised across consonants and talkers and delivered to both ears of the participants using Sennheiser HD180 circum-aural headphones. The consonant recognition task was carried out in a quiet room, and all participants completed the task either in one or two sessions. Responses of participants were stored in the form of confusion matrix, separately for each listening condition.

**Data Analysis**

The recognition score for native and non-native consonants was computed for each participant across the listening conditions, and the percent correct recognition score was transformed into rationalised arcsine units (RAU)

The transformed RAU scores were subjected to repeated measure ANOVA, to investigate the effect of languages and listening conditions on the recognition scores. All statistical analysis was performed using Statistical Package for the Social Sciences software version 16.0. In addition, the confusion matrices obtained from participants were subjected to sequential information transfer analysis using feature information xfer software. The features of consonants such as place of articulation, manner of articulation, and voicing were used for information transfer analysis. The values used for place of articulation were bilabial, labiodental, alveolar, palatal and velar; for manner of articulation the values used were stop, affricate, fricative, liquid, and nasal. Voicing had two values, voiced and voiceless. To investigate whether language and listening condition has a significant effect on the transfer of consonant feature information, the data was subjected to repeated measure ANOVA.

**RESULTS**

Figure 1 shows mean consonant recognition scores for native and non-native consonants in quiet and noise listening conditions. The recognition score was highest in favourable listening conditions (quiet and 8 dB SNR) for both native and non-native consonants. While, in challenging listening conditions (0 dB SNR and –8 dB SNR), the recognition score for both native and non-native consonants decreased with a reduction in the SNR. Further, in favourable listening conditions, the recognition score was slightly better for native consonants compared to non-native consonants. In contrast, the recognition score for non-native consonants was better than native consonants in challenging listening conditions. To investigate if the mean recognition scores are significantly different between native and non-native languages (Malayalam and English) and listening conditions (quiet, 8 dB SNR, 0 dB SNR, and –8 dB SNR), scores in RAU was subjected to repeated measure ANOVA. Results showed a significant effect of listening condition [F (3,42) = 650.946, P < 0.001] on the recognition scores, while language had no effect [F(1,14) = 1.174, P = 0.297]. Further, the interaction between language and listening condition was significant [F(3,42) = 42.453, P < 0.001]. Since there was a significant interaction between language and listening condition, repeated measure ANOVA was carried out with listening condition as repeated measures, separately for native and non-native consonants. It revealed a significant effect of listening condition for the recognition score of both native [F(2,31) = 696.406, P < 0.001] and non-native [F(2,30) = 143.261, P < 0.001] consonants. Pairwise comparison using Bonferroni test revealed a significant difference for the recognition scores across the listening conditions for both native and non-native consonants, except for listening conditions between quiet and 8 dB SNR for both native and non-native consonants (native [P = 0.197]; non-native [P = 1]) and quiet and 0 dB SNR of non-native consonants [P = 0.71].
Figure 2 shows the recognition score for each consonant across the listening conditions (Panels A, B, C, and D). Panels E, F, and G show the difference in recognition scores for individual consonants between the listening conditions. In the figure it can be noted that, in favourable listening conditions the recognition score for native consonants was better than non-native consonants, except for consonants /k/, /g/, /n/, /l/, /r/, /s/, /z/, /ʃ/, and /ʧ/, referred as native advantage. The native advantage was greatest for consonants /p/, /b/, /d/ and /v/. In contrast, at 0 dB SNR, the recognition score for most of the non-native consonants was better than native consonants, except for consonants /t/, /d/, /k/, /f/, and /ʃ/. Further, panel F shows that, reducing the SNR from 8 dB to 0 dB SNR had an effect on the recognition score of consonants /p/, /b/, /g/, /m/, /l/, and /ʃ/, while native consonants were greatly affected. On the other hand, at −8 dB SNR, the recognition score for all non-native consonants was better than native consonants, except for consonants /l/ and /ʃ/. Reducing the SNR from 0 dB SNR to −8 dB SNR affected the recognition of all consonants, except for fricatives /s/, /ʃ/, /z/ and affricate /ʃ/ for both native and non-native consonants. This finding in the present investigation shows no native advantage for the recognition of native consonants in challenging listening conditions.

**Information transfer analysis**

Figure 3 shows the transfer of information for consonant features voicing, manner of articulation, and place of articulation for both native and non-native consonants across the listening conditions. In quiet listening condition, the transfer of information was highest for manner of articulation and lowest for voicing for both native and non-native consonants. While, in the presence of noise, the transfer of information reduced with a reduction in the SNR for all consonant features of both native and non-native consonants. Further, in quiet and 8 dB SNR, the transfer of information for consonant features place of articulation and manner of articulation was similar for both native and non-native consonants. However, at 0 dB and −8 dB SNR, the transfer of place of articulation and manner of articulation information was better for non-native consonants. The transfer of voicing information of non-native consonants was greatly affected in quiet and 8 dB SNR conditions compared to native consonants. In contrast to place of articulation and manner of articulation, the transfer of voicing information was higher for native consonants compared to non-native consonants in quiet and 8 dB SNR conditions. Whereas at −8 dB SNR, the transfer of voicing information was better for non-native consonants.

To investigate whether the transfer of information was significantly different between languages, listening conditions, and consonant features a repeated measure ANOVA was performed. Results showed a significant effect of consonant features [F(1,16) = 25.191, P < 0.001] and listening conditions [F(2,27) = 603.379, P < 0.001] on the transfer of information, while language had no effect [F(1,14) = 2.625, P = 0.127]. Further, results showed a significant three-way interaction between consonant feature, language, and SNR [F(3,43) = 9.052, P < 0.001] and two way interaction between consonant features and language [F(1,19) = 4.794, P = 0.031], consonant features and SNR [F(3,41) = 4.977, P = 0.05], and language and...
SNR $[F(2,27) = 36.3, P < 0.001]$. Since the results showed a significant interaction, a repeated measure ANOVA was carried out separately for native and non-native consonants with consonant features and listening conditions as repeated measures. Results showed a significant effect of consonant features (native $[F(1,18) = 14.411, P = 0.001]$; non-native $[F(1,17) = 19.37, P < 0.001]$) and listening conditions (native $[F(2,25) = 462.762, P < 0.001]$; non-native $[F(3,42) = 171.483, P < 0.001]$) on the transfer of information. The interaction between consonant feature and listening condition was also significant (native $[F(3,45) = 4.799, P = 0.005]$; non-native $[F(3,43) = 7.949, P < 0.001]$). To investigate the effect of listening condition on the transfer of information, repeated measure ANOVA was performed separately for consonant features (place of articulation, manner of articulation, and voicing) of both native and non-native consonants. Results showed a significant effect of listening condition on the transfer of place of articulation (native $[F(2,28) = 331.308, P < 0.001]$; non-native $[F(3,42) = 138.791, P < 0.001]$), manner of articulation (native $[F(1,23) = 319.224, P < 0.001]$; non-native $[F(1,21) = 203.786, P < 0.001]$), and voicing (native $[F(2,24) = 219.215, P < 0.001]$; non-native $[F(3,42) = 35.46, P < 0.001]$) for both native and non-native consonants. Pairwise analysis revealed a significant difference for the transfer of all consonant features across the listening conditions, except for quiet and 8 dB SNR for both native and non-native consonants (native: place ($P = 0.225$), manner ($P = 0.149$), voicing ($P = 1$); non-native: place ($P = 1$), manner ($P = 1$), voicing ($P = 0.698$)). In addition, no significant difference was noted between listening conditions quiet and 0 dB SNR (place ($P = 0.126$); voicing ($P = 1$)) and 8 dB SNR and 0 dB SNR for non-native consonants (manner ($P = 0.084$); voicing ($P = 0.445$)).

**DISCUSSION**

Results of the present study showed that recognition of both native and non-native consonants was better in favourable listening conditions compared to challenging listening conditions. In the presence of noise, the recognition score reduced with a reduction in the SNR for both native and non-native consonants. These results agree with the findings of other investigations$^{[8,13-15]}$. Further, results of the present study showed that recognition of native consonants was better than non-native consonants in favourable listening conditions. However, in challenging listening conditions recognition of non-native consonants was better than native consonants. This finding was not expected since earlier investigations have consistently shown poorer recognition scores for non-native consonants in challenging listening conditions$^{[8,13,14]}$.

Better recognition score for non-native consonants in challenging listening conditions was not expected. This contradictory finding could arise due to differences in the spectrum of native and non-native consonants. Phatak$^{[16]}$ showed that recognition of consonants /t/, /s/, /z/, /ʃ/ was least affected in the presence of speech-shaped noise compared to other consonants. Similar findings have been reported by several investigations in the presence of noise$^{[15-17]}$.

Better recognition in the presence of noise was attributed to higher SNR at high frequencies for those consonants. On the other hand, the contradictory findings could be due to differences in the spectrum of noise used for masking native and non-native consonants. Any difference in the spectrum of noises used for masking native and non-native consonants could result in an unequal amount of masking of consonants, thereby resulting in different recognition scores for native and non-native consonants$^{[16,17]}$.

Since the recognition score was better for non-native consonants in challenging listening conditions, we speculated that the level of some components of speech might be higher for non-native consonants compared to native consonants. Figure 4 shows the spectrum of speech-shaped noise used for masking native and non-native consonants and VCV syllables spoken by native speakers of Malayalam and American English. It shows that the spectrum of speech-shaped noise used for masking native and non-native consonants was similar. In contrast, the VCV
syllables spoken by Native American speakers have greater energy at high frequencies and slightly lower energy at low frequencies compared to syllables spoken by native Malayalam speakers. Therefore, in the presence of noise, the effective SNR at high frequencies would be higher for non-native consonants compared to native consonants. Based on this observation, better recognition score for non-native consonants in challenging listening conditions could be attributed to relatively higher SNR at high frequencies for non-native consonants compared to native consonants. However, in favourable listening conditions, higher SNR at high frequencies did not result in better recognition scores for non-native consonants compared to native consonants. This finding could be attributed to the difference in use and weightage of perceptual cues during recognition of consonants.\(^{[18,19]}\)

To conclude, findings of the present study show that the recognition of native and non-native consonants are differently affected in the presence of speech-shaped noise among native speakers of Malayalam. In favourable listening conditions, recognition of native consonants was better than non-native consonants, while in challenging listening conditions the recognition of non-native consonants was better than native consonants.

**Financial support and sponsorship**
Nil.

**Conflicts of interest**
There are no conflicts of interest.

**REFERENCES**

1. Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 1977;61:1337-51.

2. Bialystok E. Acquisition of literacy in bilingual children: A framework for research. Lang Learn 2007;57:45-77.

3. Bradlow AR, Alexander JA. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J Acoust Soc Am 2007;121:2339-49.
4. Nábělek AK, Donahue AM. Perception of consonants in reverberation by native and non-native listeners. J Acoust Soc Am 1984;75:632-4.
5. Garcia Lecumberri ML, Cooke M. Effect of masker type on native and non-native consonant perception in noise. J Acoust Soc Am 2006;119:2445-54.
6. Cooke M, Garcia Lecumberri ML, Scharenborg O, van Dommelen WA. Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages. Speech Commun 2010;52:954-67.
7. Garcia Lecumberri ML, Cooke M, Cutugno F, Giurgiu M, Meyer BT, Scharenborg O, et al. The non-native consonant challenge for European languages. Proc Annu Conf Int Speech Commun Assoc INTERSPEECH 2008;1781-4.
8. Broersma M, Scharenborg O. Native and non-native listeners’ perception of English consonants in different types of noise. Speech Commun 2010;52:980-95.
9. Marian V, Blumfeld HK, Kaushanskaya M. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. J Speech Lang Hear Res 2007;50:940-67.
10. Shannon R.V., Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. J Acoust Soc Am 1999;106:L71-4.
11. Studebaker GA. A “rationalized” arcsine transform. J Speech Hear Res 1985;28:455-62.
12. Wang MD, Bilger RC. Consonant confusions in noise: A study of perceptual features. J Acoust Soc Am 1973;54:1248-66.
13. Cutler A, Weber A, Smits R, Cooper N. Patterns of English phoneme confusions by native and non-native listeners. J Acoust Soc Am 2004;116:3668-78.
14. Cutler A, Garcia Lecumberri ML, Cooke M. Consonant identification in noise by native and non-native listeners: effects of local context. J Acoust Soc Am 2008;124:1264-8.
15. Kalaiah MK, Thomas D, Bhat JS, Ranjan R. Perception of consonants in speech-shaped noise among young and middle-aged adults. J Int Adv Otol 2016;12:184-8.
16. Phatak SA, Allen JB. Consonant and vowel confusions in speech-weighted noise. J Acoust Soc Am 2007;121:2312-26.
17. Phatak SA, Lovitt A, Allen JB. Consonant confusions in white noise. J Acoust Soc Am 2008;124:1220-33.
18. Flege JE. Language experience in second language speech learning: In honor of Janes Emil Flege. Bohn OS, Munro MJ, editors. Studies in Second Language Acquisition. Amsterdam: John Benjamins Publishing Company; 2007.
19. Strange W. Cross-language studies of speech perception: A historical review. In: Strange W, editor. Speech Perception and Linguistic experience: Theoretical and Methodological Issues. Baltimore: York Press 1995 3-45.