# LISTENING IN A SECOND LANGUAGE

*Z. S. Bond, Thomas J. Moore and Beverley Gable*

Ohio University, Athens, OH
Armstrong Laboratories, Wright-Patterson AFB, OH

## ABSTRACT

Native and non-native listeners identified English words and sentences in six different listening conditions. When they heard speech mixed with noise or when they had to use linguistic knowledge to respond, non-native listeners suffered greater performance decrements than native listeners. Their performance appears to be data-driven ('bottom-up') requiring full specification of the acoustic-phonetic information relevant for selecting a particular word.

## 1. INTRODUCTION

Findings concerning the intelligibility of a second language (L2) have been quite consistent. Whenever listening conditions deteriorate in some way, non-native listeners, regardless of their level of proficiency in L2, experience difficulty. This finding has been consistent across a number of different methods of assessing intelligibility and a variety of difficult listening conditions.

A number of researchers have used standardized intelligibility tests to compare the word identification abilities of native and non-native listeners. This has been done for speech in noise [2,10,22], digitally encoded and synthesized speech [11,14,17] and reverberation [18,22]. All found that performance differences between native and non-native listeners were accentuated when listening conditions were made more difficult, with the non-native listeners' performance being relatively poorer.

Also employed have been tasks involving sentence understanding. Studies using noise as a masker [1,4,5,8,21] have, with one exception [9] reported that non-native listeners have intelligibility scores much lower than those of native listeners for equivalent listening conditions and require a signal-to-noise ratio (SNR) approximately 3 dB better to achieve equivalent intelligibility. Gaies, et al. [9] reported minimal differences in intelligibility between native and non-native listeners, about 3%, but they presented their materials at a relatively high SNR (+ 7 dB). Other listening conditions employed with sentence materials were time compression [6], temporal interruptions [1], and digital processing [15,16,17]. While the intelligibility of interrupted sentences was not different for the native and non-native listeners, both time compression and digital processing resulted in the non-native listeners making many more errors than did the native listeners

## 2. PURPOSE

The general issue motivating this work is the speech recognition and speech understanding abilities of non-native listeners. The present study addressed three specific questions, comparing native and non-native listener performance.

1. The effects of noise and band-pass filtering representative of that found in many common communication systems on speech recognition scores.
2. The contribution of linguistic knowledge. This question can be addressed by employing both word and sentence materials with the same listeners.
3. The variability associated with the talkers. If native and non-native listeners use similar strategies for speech understanding, they should find the same talkers relatively more or less intelligible.

## 3. METHODS

### 3.1. Talkers

Five young males served as talkers, all experienced members of the Armstrong Laboratory's Voice Communication Evaluation Panel. They each recorded Modified Rhyme Test (MRT) lists [12] and setence materials [7].

### 3.2. Listeners

Fifty-six native and 53 non-native listeners heard the MRT and sentence recordings under one of six listening conditions. All listeners were full time students at Ohio University. The non-native listeners were from Japan, China or Korea. All were fluent enough to follow college level lectures and class discussion in English.

### 3.3. Materials

The MRT training materials consisted of one MRT list of 50 words per talker and the test materials consisted of a different MRT list for each talker. The sentence training materials consisted of two sentences per talker and the test materials consisted of ten different sentences per talker.

Six sets of tapes were created (1) a 'clear' condition in which the recorded speech was not modified; (2) and (3) conditions in which pink noise was added to the recorded speech at SNRs of +3 and 0 dB; (4) a condition in which the 'clear' tape was band-pass filtered (300 to 3000 Hz); (5) and (6) where the recorded speech was band-pass filtered and pink noise at +3 and 0 dB SNR was added.

### 3.4. Procedure

Listeners were tested in small groups. Each group received one listening condition. Each listener was located at an individual station containing a recorder, earphones and microphone. First, they listened to the MRT training tape and then to the test recording. Both the training and the test recordings consisted of five different MRT lists, each one read by a different talker in the same listening condition. Listeners recorded their responses by choosing which one of the six possible rhyming words they had heard. After the MRT lists, listeners heard training and test sentences produced by each of the five talkers. Two different response modes were employed for the sentence materials. Approximately half of the listeners wrote the sentences; half repeated each sentence after they heard it. Listeners were randomly assigned to one of these two groups.

## 4. RESULTS

**Statistical Analysis.** The percent correct response to MRT words and the percent correct identification of five key words in sentences were employed as the dependent measures. These data were submitted to three separate Analyses of Variance corresponding to the three response modes, MRT words, written sentences, and spoken sentences. For each speech mode, 2 x 6 x 5 Analysis of Variance (linguistic background x listening condition x talker) was performed. Post-hoc comparisons were made using t-tests and Tukey HSD.

### 4.1. Language Background

The main effect for language background was significant in all three response modes: MRT words, written sentences and spoken sentences. In all three, the native listeners outperformed the non-natives.

### 4.2. Talkers

The main effect for talker was significant for the three response modes. Fig. 1 displays the percent correct responses for native and non-native listeners to MRT words. Native listeners reported from 83 to 87 percent of the words correctly for the four more intelligible talkers, but identified only 68 percent of the words for Talker 4. The non-native listeners identified fewer words correctly than the native listeners, and Talker 4 was the least intelligible at 52 percent, while the other four talkers were more intelligible, 70 to 75 percent. When listeners were responding to sentences, Talker 1 was the least intelligible to both native and non-native listeners, in both response modes.

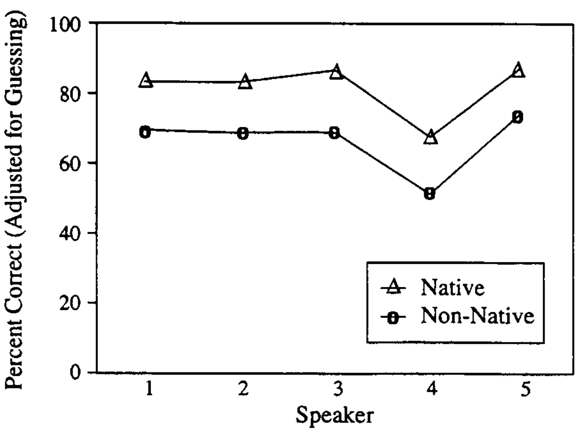


**Figure 1.** Percent correct MRT words for native and non-native listeners as a function of talker averaged across listening conditions.

Spectrograms of the MRT test words indicate that the least intelligible talker produced words shorter in duration, used the least differentiated vowel space and minimal cues for consonantal contrasts, and had the most varied vowel amplitude. Acoustic-phonetic characteristics of the speech of the talkers are described more fully in [1]. These characteristics are similar to those reported to distinguish deliberately clear from relatively casual speech [13,19].

### 4.3. Response Mode

Fig. 2 shows percent correct for MRT words and for spoken and written sentences. As expected, the non-native listeners' performance was significantly poorer than the native listeners.' Native listeners' scores were highest when they simply had to select the word they heard from the six choices provided by the MRT, 80%. When repeating and writing sentences, their scores were almost identical, 76% and 77%. These results contrast with those of the non-native listeners, who scored 66% on the MRT words, 55% on spoken sentences and 32% on written sentences. The non-native listeners not only performed poorer than the native listeners in terms of percent correct responses, but as the response mode became more demanding their relative performance deteriorated.

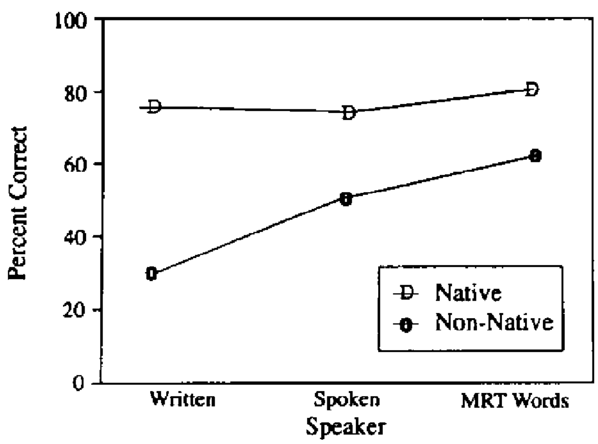


**Figure 2.** Percent correct for native and non-native listeners as a function of response mode averaged across listening conditions.

### 4.4. Listening Conditions

Average correct responses to all test materials are shown in Fig. 3 across all six listening conditions. As expected, the performance of the non-native listeners was invariably poorer than the performance of the native listeners. The intelligibility of speech in the 'clear' condition was significantly higher than speech at +3 dB SNR and at 0 dB SNR. The relative performance degradation due to added noise was significantly greater for the non-native listeners than for the native listeners. Passing the speech through a 300-3000 Hz bandpass filter before adding noise did not yield as large an effect. Band-pass filtering made no significant difference to the performance of the non-native listeners. For the native listeners, the addition of band-pass filtering resulted in significantly lower intelligibility scores when the speech was at a +3 dB SNR.

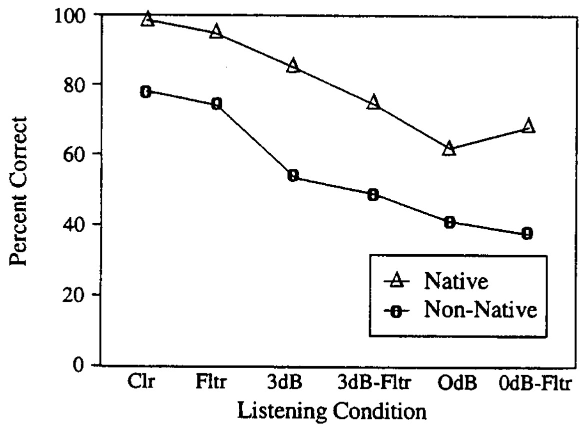


**Figure 3.** Percent correct for native and non-native listeners as a function of listening condition averaged across response mode.

# 5. DISCUSSION

The questions which motivated this research can be provided with answers, though sometimes the answers are tentative.

1. **Do non-native listeners from the Pacific rim encounter difficulties similar to those found for other groups?** As has been found for listeners from other language backgrounds, the non-native listeners showed difficulty identifying English words and understanding English sentences. While there were clear differences in their performance, in some respects the performance of both groups was similar. Both experience performance decrements as listening conditions deteriorated. Both also found the same talkers more or less intelligible.

2. **Are there talker specific differences in speech recognition scores?** Since both listener groups found the same talkers relatively more or less intelligible, both are probably using the same characteristics to identify speech. That is, the properties which make speech intelligible are not dependent on the linguistic background of the listeners. In all probability, native and non-native listeners find the same kinds of acoustic-phonetic information helpful, that is relatively long words, distinctive vowels, clearly articulated consonants, and so forth.

3. **What are the effects of signal to noise level on speech recognition?** Non-native listeners were much more affected by lower SNR's than native listeners, whether they were listening to words or sentences. This suggests that non-native listeners require much clearer acoustic-phonetic information than native listeners. Non-native listeners appear to have difficulty using partial or fragmentary phonetic information. Their performance is data-driven ('bottom-up') requiring full specification of the acoustic information relevant for selecting a particular word.

4. **Does the linguistic knowledge required for a response result in different speech recognition scores?** When dealing with more complex linguistic material, non-native listeners were at a greater disadvantage. The sentences used in the present study were designed to be grammatically simple and employed common lexical items. Non-native listener ability to understand sentences seemed slower and more effortful, as well as more dependent on full acoustic-phonetic information. It may be that, as Sajavaara [20] argues, complex tasks such as language processing require attention, but capacity is limited. As certain lower level tasks become automatized, more capacity is available for higher level sub-tasks. Perhaps the non-native listeners have not automatized the processing of acoustic information.

   That limited capacity is responsible for the poorer performance of the non-natives is further indicated by the decrease in performance in the written sentence response mode. In this condition, listeners have to remember linguistic material as well as understand it. For native listeners, memory proved to be a small additional burden. Non-native listeners had much less capacity to deal with the language and were almost overwhelmed by the demands of the additional task.

5. **What are the effects of band-pass filtering on speech recognition scores?** The effects of band pass filtering on the intelligibility of speech in the present study are less clear. The native listeners performed significantly better at moderate SNR's (+3 dB) when the speech signal was unfiltered. In other conditions, band pass filtering the speech had no significant effect on intelligibility. The non-native listeners showed no significant effect of band pass filtering at any of the listening conditions, presumably because of their relative inability to utilize the partial phonetic information available in the higher frequencies of speech.

Overall the results of this study support earlier findings that as listening conditions deteriorate the difference in performance between the two groups increases. In addition our results indicate that when the test words are embedded in sentences or the listener is required to remember speech, the differences existing between native and non-native listeners become even greater. These results suggest that extra care may be needed to assure successful communication in classroom instruction, military or civilian aviation or any number of situations where the language being used is not the native language of some of the participants.

# 6. ACKNOWLEDGEMENTS

Without the co-operation of the students and staff of the Ohio Program of Intensive English at Ohio University, this project would have been impossible. We thank all of them for their assistance. We would also like to acknowledge with gratitude the work of Miki Ueda, a research assistant of unlimited energy and resourcefulness. This project was supported in part by AFOSR.

## 7. REFERENCES

1. Bergman, M. *Aging and the Perception of Speech.* University Park Press, Baltimore, 1980.

2. Bond, Z. S. and Moore, T. J. Intelligibility of speech produced in noise and while wearing an oxygen mask. Paper presented at the Acoustical Society of America, May, Syracuse, NY, 1989.

3. Bond, Z. S. and Moore, T. J. The acoustic-phonetic characteristics of inadvertantly clear speech. *Speech Commun.* 14, 1994, 325-337.

4. Borchgrevink, H. M. Second language speech comprehension in noise—A hazard to aviation safety. AGARD Proc., Aural *Communication in Aviation,* AGARD-CP-311, 1981, 15.1-15.5.

5. Buus, S., Florentine, M., Scharf, B., and Canevet, G. Native French listeners' perception of American English in noise. *Inter-noise 86: Proc. Int. Conf. Noise Control Engineering,* 1986, 895-898.

6. Conrad, L. (1989). The effects of time-compressed speech on native and EFL listening comprehension, *Studies Second Lang. Acq.* 11, 1989, 1-16.

7. Egan, J. P. Articulation testing methods II, Office of Scientific Research and Development Report No. 3802, 1944.

8. Florentine, M. Non-native listeners' perception of American-English in noise. I*nter-noise 85: Proc. Int. Conf. Noise Control Engineering,* 1021-1024, 1985.

9. Gaies, S. J, Gradman, H. L., and Spolsky, B. Toward the measurement of functional proficiency: Contextualization of the noise test, *TESOL Quart.,* 11, 1977, 51-57.

10. Gat, I. B. and Keith, R. W. An effect of linguistic experience: Auditory word discrimination by native and non-native speakers of English, *Audiology,* 17, 1978, 339-345.

11. Greene, B. G. Perception of synthetic speech by non-native speakers of English. Proc. *Human Factors Society,* 30th Annual Meeting. 1340-1343, 1986.

12. House, A. S., Williams, C. W., Hecker, M. H. L., and Kryter, K. D. Articulation testing methods: Consonantal differentiation with a closed response set. *J. Acoust. Soc. Amer.,* 37, 1965, 158-166.

13. Lindblom, B. Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling,* Kluwer, Dordrecht pp 403-439, 1990.

14. Logan, J. S. and Ozawa, K. Using native and non-native speakers of English to evaluate the segmental intelligibility of digitally coded speech. *Proc. Speech-Tech '89,* 106-108, 1989.

15. Mack, M. Sentence Processing by non-native speakers of English: Evidence from the perception of natural and computer-generated anomalous L2 sentences, *Jl. Neuroling.*, 3, 1988, 293-316.

16. Mack, M. and Tierney, J. The intelligibility of natural and vocoded semantically anomalous sentences: A comparative analysis of English monolinguals and German-English bilinguals. *Massachusetts Institute of Technology Lincoln Laboratory Technical Report 792,* 1987.

17. Mack, M., Tierney, J., and Boyle, M. E. T. The intelligibility of natural and LPC-vocoded words and sentences to native and non-native speakers of English. *Massachusetts Institute of Technology Lincoln Laboratory Technical Report 869,* 1990.

18. Nabelek, A. K. and Donahue, A. M. Perception of consonants in reverberation by native and non-native listeners, *Jl. Acoust. Soc. Amer.,* 75, 1984, 632-634.

19. Picheny, M. A., Durlach, N. I. and Braida, L. D. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech, *Jl. Sp. Hear. Res.,* 29, 1986, 434-446.

20. Sajavaara, K. The monitor model and monitoring in foreign language speech communication. In R. C. Gingras (Ed.): *Second Language Acquisition and Foreign Language Teaching,* Center for Applied Linguistics, Arlington, VA, pp. 51-67, 1978.

21. Spolsky, B., Sigurd, B., Sato, M., Walker, E., and Arterburn, C. Preliminary studies in the development of techniques for testing overall second language proficiency, *Lang. Learn.,* 18, 1968, 79-101.

22. Takata, Y. and Nabelek, A. K. English consonant recognition in noise and in reverberation by Japanese and American listeners, *J. Acoust. Soc. Amer.,* 88, 1990, 663-666.