RHYTHM AND REGION: SCALING THE PERCEPTUAL DIMENSIONS OF KOREAN

Verna Stockmal and Z. S. Bond
Ohio University, Athens OH 45701 USA

ABSTRACT

After 10 minutes of exposure to a spoken sample, listeners rated the similarity of Korean to rhythmically similar languages (Experiment 1) or to languages from the same geographical area (Experiment 2). Listener ratings were submitted to MDS analysis receiving two-dimensional solutions. When listeners were comparing Korean to rhythmically similar languages, the dimensions of the derived stimulus configuration could be interpreted as 'geographic area' and 'syllable type variety'. When listeners were comparing Korean to other languages from South East Asia, the derived stimulus configuration showed sensitivity to speaker affect, prosodic pattern and segment inventory.

1. INTRODUCTION

Listeners are able to identify foreign languages without understanding them on the basis of their 'sound' or 'acoustic signature' after relatively brief exposure. When making identification or discrimination judgments, listeners claim to be attending to both segmental and suprasegmental properties. Not surprisingly, listeners are strongly influenced by the languages they are asked about and by the nature of the task given them.

Only a few studies have specifically examined the learning responsible for language identification. Bond & Fokes [2] tested listener ability to identify two-second samples of foreign languages in a forced-choice format. All listeners identified the foreign languages at above chance rates, but confusion patterns implicated language exposure as an important determinant of success. For example, listeners from South America had difficulty identifying Asian languages. In an earlier study not specifically aimed at language identification, listeners spontaneously guessed the identity, family, or geographical region of some of the six languages presented to them [10]. Muthusamy et al. [12] reported that listeners showed some improvement in language identification performance after training and continuous feedback. Listeners identified familiar European languages better than less familiar languages.

In a multidimensional scaling analysis of subject responses, Nazzi [13] reported a similar finding, that listeners were sensitive to language familiarity and classified familiar languages by rhythmic pattern but had difficulty responding to unfamiliar languages. Stockmal, Muljani & Bond [16] investigated listener descriptions of foreign languages using multi-dimensional scaling. They found that listeners were sensitive to speaker voice characteristics or affect and prosodic pattern.

Bond, Stockmal & Muljani [4] examined how listeners' ability to identify a target language, Japanese, changed after exposure to different types of listening materials for different exposure times. Even brief (five minute) exposure to samples of the target language enabled listeners to identify it accurately when they were tested immediately after exposure. Confusion patterns indicated that Russian and Arabic were almost never misidentified as Japanese. Listeners confused Indonesian and Chinese with Japanese at different rates, depending on the length of exposure and materials. Apparently, listeners could reject Russian and Arabic on the basis of the phonetic properties associated with stress-based rhythm [1,5,6] and from speaker characteristics [7] associated with East Asian languages.

The study was replicated using Korean as the target language and presenting it for identification with competitors classified as employing the same rhythmic patterns or spoken in the same geographical region. Korean was chosen as the target because it is similar to Japanese at some levels [16] but American listeners, if they have been exposed to Korean at all, are unlikely to have even vague ideas about how Korean 'sounds'. Muthusamy et al. [12] reported that listeners had the most difficulty recognizing Korean and sometimes called a language sample Korean if they could not identify it as any other language.

Listeners who had no previous exposure to Korean had minimal expectations about its 'sound'. When listening to Korean and rhythmic competitors, they confused the Asian languages, particularly Japanese, with Korean more than European or African languages. After exposure to the target language, listener judgments shifted and they made more false alarms for the non-Asian languages and fewer for Japanese. For geographical area competitors, listeners without exposure had extreme difficulty identifying Korean, but after exposure, listener false alarms to Chinese and Japanese were reduced. As when Japanese was the target, listeners could identify novel samples of Korean at better than chance levels after exposure. Different types of competitors selectively focused listener attention and influenced confusion patterns [3].

The present study was conducted to extend our understanding of how language characteristics in the test materials interact with listeners' perceptual judgments. Listeners were asked to rate the similarity of Korean to rhythmically matched competitors and to competitors from the

same geographical area in order to investigate the perceptual dimensions underlying identification and discrimination judgments. By equating some characteristics of the competitor languages, listener attention would be focused on other characteristics which distinguish the 'sound' of languages.

2. EXPERIMENT 1

2.1. Purpose.

The first experiment investigated the perceptual dimensions which listeners use when they are asked to compare the target language, Korean, with other rhythmically similar languages. According to Dauer's [5, 6] cross-linguistic method of classifying rhythmic properties, Korean is syllable-timed.

2.2. Method

2.2.1. Materials. Three female native speakers of the same dialect of Korean recorded prose passages read at a normal speaking rate. One recorded passage was used to familiarize listeners with Korean. The other two passages were employed in the test materials.

A listening test of 30 five-second spoken samples from five different languages was constructed. Ten samples were Korean. No identifiable lexical items occurred in both the exposure and the test materials, making it impossible for listeners to select the language simply by recognizing words.

The remaining test items consisted of five samples from each of four other rhythmically similar languages. The syllable-timed competitors to which listeners compared Korean consisted of languages from Asia, Europe and Africa including Japanese, Latvian, Ombawa (or Mbawa) and Tagalog. By selecting languages which employ similar rhythmic patterns, listener attention would be focused on other salient properties which serve to distinguish spoken languages. The Korean and competitor samples were presented to listeners in random order.

2.2.2. Listeners. Forty-one undergraduate monolingual college students with self-reported normal speech and hearing participated in the study in return for a small amount of course credit.

2.2.3. Procedure. After familiarization with the task, listeners first heard a recorded passage of spoken Korean for approximately 10 minutes. Then they judged the similarity of each language sample to the target, Korean, on a 7-point scale. Listeners assigned a rating of 7 to an item if they considered it to be very similar or identical to the target language, Korean. They assigned a rating of 1 if they considered the sample to be very different from the target.

2.3. Results.

Listeners judged novel samples of Korean to be most similar to the target Korean, with a mean score of 4.5. That listeners gave the highest rating to Korean indicates that they were responding to some of the salient acoustic-phonetic characteristics of the language. Essentially, the listeners were identifying Korean.

Japanese was judged as also very similar to Korean, receiving a mean rating of 4.1. The other three competitor languages received somewhat lower ratings: Tagalog 3.3, Latvian 3.0, and Ombawa 2.3.

The listener ratings were converted to a proximity matrix and submitted to multi-dimensional scaling analysis (MDS). The assumption underlying MDS procedures is that judgments can be viewed as representative of perceptual distances and that the derived spatial configuration reveals dimensions relevant to listeners [14].

The optimal MDS solution was a two-dimensional configuration, RSQ = 1.0, Stress = 0.0. The first dimension separated Asian languages from the European language, Latvian, and the African language, Ombawa. The second dimension was interpreted as representing phonotactic properties, specifically the variety of syllable types, which are more complex in Latvian than in Japanese or Korean. The isolated location of Ombawa may result from judgments based on strong vowel nasalization throughout syllables which is characteristic of the language [9]. The MDS configuration is given as Fig. 1.

That Asian languages were distinguished from the other two is consistent with the suggestion made by Esling and Wong [7], among others, that talker voice quality or 'voice settings', a term used to describe the auditory impression made by a certain mechanical or habitual setting of the speech organs over a stretch of speech, may function not only to characterize a speaker but also a language, or a speaking style associated with a particular geographically or socially defined variety of a language.

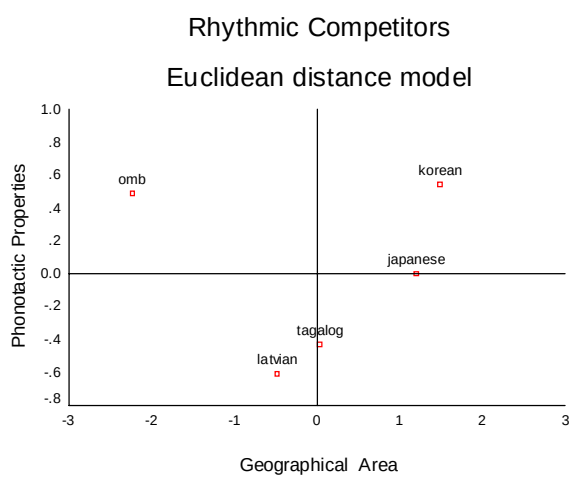

Fig. 1. Two-dimensional ALSCAL solution for rhythmic competitors.

3. EXPERIMENT 2

3.1. Purpose.

The second experiment investigated the perceptual dimensions which listeners employ when they are asked to compare the target language, Korean, with other languages which are spoken in the same geographical region, East Asia. The languages selected as competitors included Chinese, Indonesian, Japanese and Tagalog, controlling for characteristic regional speaking styles.

3.2. Method.

3.2.1. Materials. The same Korean talkers were used as in the first experiment. A listening tape was constructed identical to the one used in the first experiment except that the competitor languages were from East Asia.

3.2.2. Listeners. Twenty undergraduate monolingual college students with self-reported normal speech and hearing participated in the study in return for a small amount of course credit. The listeners were selected from the same population as in the first experiment.

3.2.3. Procedure. The procedure was identical to the procedure employed in the first experiment.

3.3. Results.

As in the first experiment, listeners judged novel samples of Korean to be most similar to the target Korean, mean similarity rating of 4.46. This rating is almost identical to the rating Korean obtained in the first experiment. In the context of Asian languages, Indonesian, at 3.8, and Chinese, at 3.6, were rated as most similar to Korean. Japanese and Tagalog both received a rating of 2.8, a lower rating than these two languages received when the competitors included languages from Europe and Africa.

The listener ratings were converted to a proximity matrix and submitted to multi-dimensional scaling analysis (MDS). The solution was a two-dimensional configuration, RSQ = .99, Stress = .003. The MDS configuration is given as Fig. 2.

The interpretation of the first dimension was not entirely clear. The dimension separated Korean from the other four languages and clustered Indonesian with Chinese, and Japanese with Tagalog. It is possible that listeners were reacting jointly to speaker and language characteristics in that both the Chinese and Indonesian speakers produced the language samples in a matter-of-fact style while the Tagalog speaker produced a much more dramatically read passage.

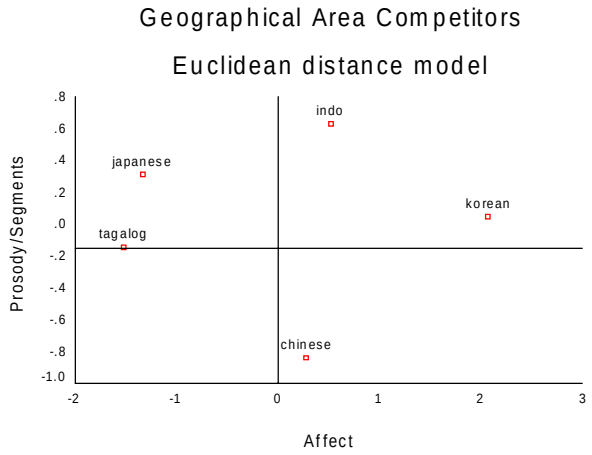

Fig. 2. Two-dimensional ALSCAL solution for regional competitors.

The second dimension separates Chinese from the other four languages. In making these judgments, listeners were probably reacting both to the pitch excursions which in Chinese reflect lexical tone and to the consonant and vowel inventory which shows more diversity in Chinese than in the other four languages [11]. In addition, it is possible that Mandarin Chinese employs stress-based rhythm, the only Asian language in the test to do so [8].

4. CONCLUSION

In the two experiments, listeners were asked to judge the similarity of spoken samples of a target language, Korean, to samples of other languages. Even though the task required listeners to make judgments about quite abstract phonetic properties, they apparently could do the task. Furthermore, the selection of languages serving as competitiors appears to have focused listener attention on different properties characteristic of the 'sound' of languages. For example, in the first experiment, listeners were sensitive to voice quality and speaking style characteristics they considered Asian, particularly noticeable in the closeness of mean ratings of Japanese and Korean. Consequently, the horizontal axis was identified as the geographical area of the target language. In the second experiment, when voice quality characteristic differences were only minimally available, listeners seemed to refine their judgments and adjusted their focus to the affective or paralinguistic properties of speaking style. This in turn re-defined the horizontal axis as affect.

In experiment 1, the vertical dimension emerged as phonotactics or syllable structure. This dimension could not be considered gradient, from more to less complexity. Rather, it appears to represent overall 'sound' contrast. Ombawa, characterized by strong nasality throughout syllables, is obviously different from all the other languages. Latvian, which employs complex syllable onsets, also appears different. However, Latvian shares with Korean a long-short vowel contrast and some overalp in fricative and affricate inventory, which may account for its position within the language map. It should be noted that in the identification experiment [3], listeners confused Latvian with Korean more after exposure than before exposure. Tagalog, with a three-vowel inventory and simple syllable patterns, is most distant from Korean within the Asian language group. In experiment 2, Chinese which employs a diverse consonant inventory and different rhythm from the other languages occupies a unique position. It

appears that listeners make simple judgments based on rhythmic properties when these are available.

The dimensions derived in these two experiments have been found to be salient to listeners in previous studies dealing with identification or discrimination among languages. In this study, changing competitors changed salience to listeners and languages were reconfigured within the perceptual space. Japanese, for example, was closest to Korean in experiment 1 but rather distant when competing with Asian languages in experiment 2.

There is one caution in interpreting the experiments. Listeners were exposed to Korean for only 10 minutes. When they were making judgments about competitors, they heard these languages for only 5 seconds. It is possible that 10 minutes may present a biased sample for judging the sound of Korean. It is quite likely that 5 seconds does not completely represent the 'sound' of a language. Possibly, the Korean sample used in experiment 1 contained a sufficient number of fricatives and affricates to make it sound similar to Latvian and also placed Latvian closer to Tagalog on the map than one would expect. Perhaps a different sample of Korean would have lead to somewhat less proximity between these two languages.

It is not known how long a sample needs to be to represent the 'sound' of a language to listeners. It may be that languages differ in their distinctiveness and that some can be identified from considerably shorther segments than others. For naive listeners, the maximal practical exposure time to a language they do not know is quite limited, no more than 10 or 12 minutes. At longer exposure times, they simply become inattentive.

One useful technique in further studies exploring listener responses to the 'sound' of foreign languages would be to select competitor languages to control more specifically for other potentially salient characteristics. Possibilities include stress-based rhythm, vowel inventory, syllable structure, and so forth. A complementary approach would be to ask listeners to rate samples of languages on a selection of properties. Listener ratings could then be converted to configurations by means of MDS analysis to investigate the dimensionality underlying judgments.

ACKNOWLEDGEMENTS

We want to acknowledge the assistance of our talkers. Without their help in providing language samples, this project would not have been possible. We also want to thank Coco and Feena for useful discussion.

REFERENCES

[1] Bertinetto, P. M.. 1989. Reflections on the dichotomy ‘stress’ vs. ‘syllable-timing.’ *Revue de Phonetique Appliquee*, 91/93, 99-130.

[2] Bond, Z. S. and Fokes, J.1991. Identifying foreign languages. In *Proceedings of the XII International Congress of Phonetic Sciences*, Aix-in-Provence, 198-201.

[3] Bond, Z. S. and Stockmal, V. 1999. Selecting samples of spoken Korean from rhythmic and regional competitors. Acoustical Society of America, Berlin.

[4] Bond, Z. S., Stockmal, V. and Muljani, D. 1998. Learning to identify a foreign language. *Language Sciences*, 20**,** 353-367.

[5] Dauer, R., 1983. Stress timing and syllable timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

[6] Dauer, R. 1987. Phonetic and Phonological components of language rhythm. In *Proceedings of the XI International Congress of Phonetic Sciences*, Tallinn,. 447-450.

[7] Esling, J. H. and Wong, R. F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL Quarterly*, 17, 89-95.

[8] Hung, Feng-Sheng. 1996. Prosody and the acquisition of grammatical morphemes in Chinese languages. Indiana University Linguistics Club: Bloomington, IN.

[9] Ladefoged, P. 1968. *A Phonetic Study of West African Languages*. London: Cambridge University Press.

[10] Lorch, M. P. and Meara, P. 1989. How people listen to languages they don’t know *Language Sciences*, 11, 343-353.

[11] Maddieson, I. 1984. *Patterns of Sound*. New York: Cambridge University Press.

[12] Muthusamy, Y. K., Jain, N. and Cole, R. A. 1994. Perceptual benchmarks for automatic language identification. In *Proceedings of IEEE International Conference of Acoustic, Speech and Signal Processing*, 94 I, 333-336.

[13] Nazzi, T. 1997. *Du rhythm dans l'acquisition et les traitement de la parole*. Ph.D. thesis. Ecole des Hautes Etudes en Sciences Sociale, Paris.

[14] Schiffman, S., Reynolds, M. L. and Young, F. 1981. *Introduction to Multidimensional Scaling*. New York: Academic Press.

[15] Stockmal, V., Muljani, D. and Bond, Z. S. 1996. Perceptual features of unknown foreign languages as revealed by multi-dimensioanal scaling. In *Proceedings of International Conference on Spoken Language Processing*, Philadelphia.

[16] Vendetti, J. J., Jun, S.A. and Beckman, M. E. 1996. Prosodic cues to syntactic and other linguistic structure in Japanese, Korean and English. In. J. L. Morgan and K. Demuth (eds.) *Signal to Syntax*, Manwah, N.J.: Lawrence Erlbaum.