

Perceptual Features of Unknown Foreign Languages as Revealed by Multi-dimensional Scaling

V. Stockmal, D. Muljani, and Z. Bond

Ohio University
Athens, Ohio

ABSTRACT

Adult listeners are able to discriminate between and often identify spoken samples of languages that are unknown to them. Two studies were designed to explore which perceptual properties inherent within the phonological structure of languages are salient to foreign language listeners.

In study one, fifteen subjects were asked to judge whether pairs of spoken foreign language sentences were selected from same or different languages and to explain how they had made the judgement. A multi-dimensional scaling (MDS) was conducted on the subject responses for the 'same language' condition. The resulting map revealed that responses could be characterized along two dimensions: phonologically based psychoacoustic properties and talker specific characteristics. The two dimensions define the distinctiveness of the languages and elicited different perceptual feature relationships in subjects.

In study two, this perceptual feature relationship was tested using similarity judgements. Thirty subjects rated similarity on a seven-point scale for the same set of sentence pairs that had been judged in study one. MDS analysis revealed that the 'different language' condition yielded a map in which the language proximities closely approximated those which had been derived by focusing on phonological properties. This finding suggests that since analysis of both 'different language' and 'same language' sentence pairs produced similar maps, perceived language similarity among foreign languages depends upon the listeners' salient organizational categories inherent within the phonological structure of language and the talker specific characteristics of voice quality and speech rate.

I. INTRODUCTION

Perception of spoken language is dependent on the initial sensory and perceptual analysis of the acoustic phonetic input, the essential properties of the language in which the utterance is spoken (12). Each language has a characteristic structure which makes that language sound different from other languages. The language structure is defined by its phoneme inventory, phonotactic rules, syllable construction and prosodic patterns. Languages differ in their phonological complexity but it is likely that both a minimum number of segments required to build an adequate vocabulary of distinct morphemes and a maximum number of segments which can be distinguished efficiently exists (10). From a database of 317 languages, Maddieson concluded that the typical size of segment inventory for a language is between 20 and 37 segments. Phonemes, syllables and even words may be common to more than one language.

Nevertheless, human listeners can easily recognize spoken non-native utterances as foreign and can often identify the language in which the utterance has been spoken (3). Presumably, identification is based on some exposure to that specific language. Human perceptual analysis is so acute that even having had no exposure to a language, naive listeners can distinguish it from other unfamiliar languages when the utterances are only two-seconds in length with an 80% accuracy rate (16).

When meaning is absent from an utterance, listeners may attempt to extract all the possible recognizable physical cues inherent within the language's structure which can then be categorized for analysis. Or, they may require only a few salient cues which can be compared to native language categories already established since speech perception abilities develop and become organized in conjunction with language acquisition (8). Lorch and Meara (9) studied how people listen to a foreign language. Naive listeners were able to describe salient sounds and prosodic features as well as how the foreign language differed from their native language. Some could identify a few "words" or "syllables". This gives an insight into the perceptual process but does not show how listeners distinguish foreign languages from one another.

Many studies have examined non-native discrimination of single properties of languages and found that discrimination is not uniform. Non-native phonemes are sometimes easily distinguished, sometimes not, even with training (17). It may be that some contrasts are psychoacoustically based (1), while others are assimilated to native phonemic categories (13) requiring different processing techniques for perceptual analysis. It is not known whether within a sentence some single segments remain more salient than others and become magnets for identification of a particular language. One might predict, however, that aggregates of segments dictated by phonotactic rules are more easily recognized for identification.

It has been suggested that listeners rely on prosodic information to help segment speech (5) perhaps, into aggregates of short sounds which may not require resolution into discrete components (2). Listeners seem to seek these aggregates; for example, words, or syllables even in non-meaningful speech. Because prosodic features have a tendency to co-occur; for example, duration may be a manifestation of stress, listeners' discussions of prosody are often vague. Although comparisons of prosodic features across several languages have not been based on human perception studies (6, 7), it is clear that temporal organization of an utterance is important to listeners.

Our research attempts to address the question of how perceptual categories cue naive listeners to the differences between unknown foreign languages. In this study we analyzed descriptive data from listener reports explaining how subjects judged whether the pair of

sentences within a stimulus were spoken in the same foreign language or different foreign languages. The analysis was conducted to determine

- which perceptual categories are organized to elicit unique language responses
- and if language patterns can be found based on perceptual relationships.

We then evaluate the validity of response patterns by analyzing similarity judgements made on the same set of languages by a different set of subjects.

2. METHOD

2.1. Materials

Materials for both experiments were the same. Two native speakers of Arabic, Chinese, Indonesian, Japanese, Russian and Spanish recorded full sentences of approximately five-second length. The stimulus tape was generated by recording the sentences in pairs after normalization to the same peak amplitudes. Inter-sentence silence interval was 0.5 second. Each language was paired with all the other languages and with itself. Sentence pairs in which the sentences were spoken in the same language were identical, that is, both the sentence and speaker were the same, or either the speaker or the sentence varied. The sixty-six pairs of sentences were then randomly ordered for presentation.

2.2. Subjects

The fifteen subjects in experiment 1 and the thirty subjects in experiment 2 were university students between the ages of eighteen and twenty. All were native speakers of American English with no known deficiency in speech, language or hearing. Although all subjects had studied some foreign language in high school, none has fluency in any language but English.

3. EXPERIMENT 1

3.1. Procedure

Subjects were asked to listen to each pair of sentences and decide if they were both spoken in the same language or not and record their answers on a score sheet. After each sentence pair, the subject provided an explanation of her decision strategy which was recorded by the experimenter. In order not to influence subject's reports, subjects were not told whether their answer was correct nor were they provided with linguistically technical terms for description.

3.2. Results and Discussion

A multi-dimensional scaling analysis (MDS) was performed on the descriptive data derived from the "same" condition only (when both pairs of sentences were the same language). Subjects' observations were categorized into 12 variables. Proximities were calculated as the frequency of the variables by language. The best solution for the dissimilarity coefficient matrix for the six languages was computed by ALSCAL (15) as a 2-dimensional Euclidean space: stress value 0.046, RSQ = 0.987.

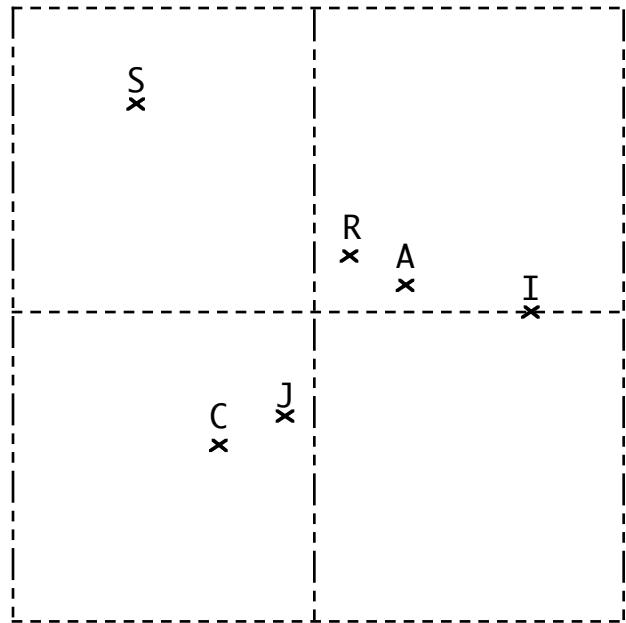


Figure 1: Two dimensional nonmetric solution for language proximity data based on subject discrimination judgements in pairs of same language sentences. The horizontal dimension represents talker specific characteristics. The vertical dimension represents psychoacoustic properties.

The first dimension was interpreted in terms of the psycho-acoustic property of pitch which emerged as a salient dimension. Languages which use large pitch excursions over short duration contrast with languages which do not. In this dimension, Chinese, a contour tone language and Japanese, which uses pitch similarly although not contrastively, cluster together. Confusion between these two languages was greater than between any other two languages in the study. When they were incorrectly identified as the same language, the reason most often expressed was "the intonation, stress, or accent is similar". Even when subjects gave the correct answer, they still maintained awareness of this same pitch pattern, remarking that they 'sound' the same. But awareness of segments was also evident by further clarification, for example, one sentence had more 'sh' or more nasal sounds. It was also clearly apparent that subjects were aware that there were differences in intonation between Chinese and Japanese which was often described as 'choppy'.

At the other end of the dimension, Russian and Arabic are spatially closer to each other than either is to Indonesian. Russian and Arabic are considered stress-timed languages, but Indonesian is syllable-timed. When subjects describe either Arabic or Russian they rarely mention rhythm or intonation. The close rhythmic relationship with English, the native language of the listeners, may account for this phenomenon. Listeners did report they heard 'words' and word endings in these languages. If listeners believed they were able to segment an utterance of these unfamiliar languages into words or find syllables, familiar rhythm may have been the cue which contributed to the segmentation process.

Similarly, subjects used "word" to identify Indonesian but only in response to reduplicative words; e.g. lain-lainnya. Subjects often referred to Indonesian as Spanish and in one case "oriental Spanish" when comparing it to other languages. Not surprisingly, both languages are syllable-timed. However, when Indonesian and Spanish were paired, they were not confused. In fact, Spanish is a unique case in this study in that most subjects were able to identify it by name.

The second dimension was interpreted as talker specific characteristics since speaker voice and speech rate emerged as salient to listeners. Listeners are very conscious of voice qualities and several reported having difficulty with voice interference in the language decision process. Confusion patterns of the Indonesian and Russian indicated that subjects who perceived similar speaker voices called the languages the same. Also, some subjects who confused Chinese and Japanese claimed they heard an 'oriental voice' and decided the language pair was spoken in the same language.

Speech rate seemed particularly important. Listeners complained often that speakers spoke too fast. Speech rate contributes to clear speech (11) and both native and non-native listeners find fast speakers less intelligible (4) than speakers who speak more slowly. In addition, monotonous sections of speech have been reported to be perceived as having been spoken faster (14). Indonesian, considered a 'monotonous' language, was perceived to be spoken very fast and was, therefore, difficult to discriminate. Russian, also considered fast, was somewhat easier to discriminate, perhaps, because it is a stress-timed language. Interestingly, no subject considered the tone languages, Japanese and Chinese, to be spoken too fast to distinguish.

The structure underlying the two dimensional figure suggested by this analysis reveals that clusters within the first dimension are indicative of the complexity of languages. Perceived differences in linguistic rhythm may depend on the language's specific segmental variation (6). Naive listeners' reports defined languages with similar rhythm patterns by segmental variables. So far, language identification studies which attempt to solve the problem by studying prosody or segmentation have obtained comparable results with either prosodic or segmental algorithms. These studies may be looking at only one of the dimensions used by human listeners. Speaker characteristics may have to be incorporated into the process to successfully discriminate languages.

4. EXPERIMENT 2

4.1. Procedure

Subjects were instructed to listen to the sentence pairs and rate the similarity of the languages in each pair on a scale of one to seven (15). On this scale, a rating of one was given if subjects thought the languages were very dissimilar and a rating of seven if the languages were very similar.

4.2. Results and Discussion

An MDS analysis was performed on the data derived from the "different" language pairs only. Proximities were calculated from similarity ratings. The best solution for the dissimilarity coeffi-

cient matrix for the six languages was again a 2-dimensional Euclidean space: stress value 0.053, RSQ = 0.983.

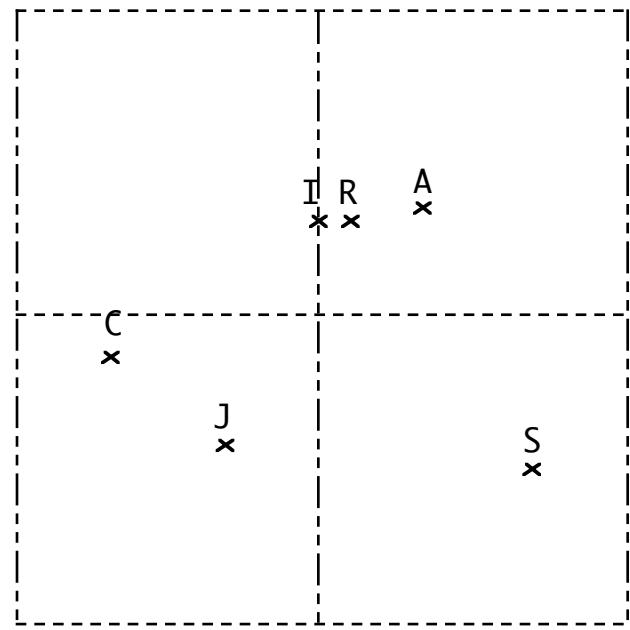


Figure 2: Two dimensional nonmetric solution for language proximity data based on similarity judgements of pairs of different language sentences. The horizontal dimension represents talker specific characteristics. The vertical dimension represents psychoacoustic properties.

Figure 2 finds the six languages clustered into configurations similar to that of Figure 1. However, in Figure 2 the space has been rotated somewhat. This is reasonable since the first set of judgements was based on 'same' language pairs and this set of judgements on 'different' language pairs. In this experiment, Indonesian and Russian have closer proximity than Arabic and Russian. Perception of speech rate may have been more acute since subjects had only 20 seconds between pairs in which to make their decision. Spanish is once again unique. Chinese and Japanese remain in close proximity with each other. Chinese is still the most easily differentiated language. This analysis validates the analysis derived from phonological properties since subjects' actual perception of language similarity compared very closely with language similarity derived from perception of salient phonological properties organized to give unique characteristics to each language.

5. CONCLUSION

This finding suggests that since analysis of both 'different language' and 'same language' sentence pairs produced similar maps, human listeners perceive distinctiveness among foreign languages by analyzing salient organizational categories inherent within the phonological structure of the language as well as talker specific characteristics. Further studies should try to determine how much the phonological features and how much the talker characteristics contribute to the discrimination and identification of unfamiliar foreign languages.

6. REFERENCES

1. Best, C., McRoberts, G. and Sithole, N. "Examination of perceptual reorganization for non-native speech contrasts: Zulu click perception by English-speaking adults and infants," *J. Exp. Psychol: Human Perception Performance*, Vol. 14, 1988, p 345-360.
2. Bond, Z. "On the specification of input units in speech perception." *Brain and Language*, Vol. 3, 1976, p 72-87.
3. Bond, Z. and Fokes, J. "Identifying foreign languages," *Proc. XII Con. Phon. Sciences*, 1991, p 198-201.
4. Bond, Z. and Moore, T. "A note on the acoustic-phonetic characteristics of inadvertently clear speech," *Speech Communication*, Vol. 14, 1994, p 325-337.
5. Cutler, A. and Mehler, J. "The periodicity bias," *Jour. of Phonetics*, Vol. 21, 1993, p 103-108.
6. Dauer, R.M. "Stress-timing and syllable-timing reanalyzed," *Jour. of Phonetics*, Vol. 11, 1983, p 51-62.
7. Eady, S.J. "Differences in the FO patterns of speech: Tone language versus stress language," *Language and Speech*, Vol. 25, 1982, p 29-42.
8. Jusczyk, P.W. "from general to language-specific capacities: The WRAPSA model of how speech perception develops," *Journ. of Phonetics*, Vol. 21, 1993, p 3-28.
9. Lorch, M and Meara, P. "How people listen to languages they don't know," *Language Sciences*, 11(4), 1989, p 343-353.
10. Maddieson, I. *Patterns of Sound*, Cambridge University Press, Cambridge, 1984.
11. Picheny, M., Durlach, N. and Braida, L. "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *Jour. Speech and Hearing Research*, Vol. 29, 1986, p 434-446.
12. Pisoni, D.B. "Speech perception: Some new directions in research and theory," *Jour. Acous. Soc. Amer.* Vol. 78, 1985, p 381-388.
13. Polka, L. "Cross-language speech perception in adults: honemic, phonetic and acoustic contributions," *Jour. Acous. Soc. Amer.* Vol. 89, 1991, p 2961-2978.
14. Rietveld, A. and Gussenboven, C. "Perceived speech rate and intonation," *Jour. of Phonetics*, Vol. 15, 1987, p 273-285.
15. Schiffman, S., Reynolds, M.L., and Young, F. *Introduction to Multidimensional Scaling*, Academic Press, Orlando, 1981.
16. Stockmal, V., Muljani, D. and Bond, Z. "Can children identify samples of foreign languages as same or different?" *Language Sciences*, Vol 16 (2), 1994, p 237-252.
17. Werker, J. "The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics*, Vol. 7, 1986, p 141-156.