An Instance-Based Model of Japanese Speech Recognition by Native and Non-Native Listeners

Kiyoko Yoneyama and Keith Johnson

Department of Linguistics, Ohio State University, 222 Oxley Hall, Columbus, OH 43210

Abstract: Categorical perception based on segmental duration is observed in Japanese. In contrast, non-native Japanese listeners find it difficult to acquire this contrast and it is acquired only after extensive exposure to Japanese. These three observations are explored in an instance-based model of phonological learning and word recognition. Results showed categorical perception of a nasal duration continuum in a model trained on Japanese utterances and non-categorical perception in a model trained on American-English utterances. 'Bilingual' models had to have at least as many Japanese words as English words to exhibit categorical perception.

INTRODUCTION

Many researchers have shown that categorical perception of segmental duration is observed in Japanese. For example, koma 'top' and komma 'comma' are Japanese words that are perceived as a two-mora, and a three-mora word, respectively. The only difference between the two is the duration of the bilabial nasal segment. In the case of koma, the bilabial nasal is relatively shorter and it is perceived as a single segment. In the case of komma, the duration is relatively longer and it is perceived as a geminate consonant. The first part of the geminate consonant is placed in the coda of the following syllable and is perceived as an extra mora (1). It is also reported that non-native Japanese learners have difficulty acquiring this contrast (2). According to these previous findings, we hypothesize that categorical perception arises from enough exposure to instances of category members in the native language.

The purpose of this paper is to investigate this hypothesis by conducting simulations of three groups of language users: Japanese, American-English, and Japanese and English bilinguals. Before going on to discuss simulations of the process by which linguistic structure emerges from specific instances in memory, we will briefly explain major features of the model.

THE MODEL

Our instance-based model of phonological learning and word recognition is based on three assumptions. First, speech is recognized by reference to stored instances (exemplars) (3). Second, these exemplars have no internal structure, rather they are unanalyzed auditory representations. And third, they are word-sized chunks, as a result of primitive auditory scene analysis where isolated word productions form the basis for word recognition in running speech. In this model, each 23 ms frame of speech is processed by an auditory model, vector quantized, and compared with the set of remembered sequences in an exemplar covering map (4). Word node activation is the product of similarity to the covering map location and the learned associations between that location and the word node (see (5) for details).

SIMULATIONS

The first two simulations investigated whether categorical perception arises from instances of category members in Japanese. The first simulation used two Japanese words (koma 'top' and komma, 'comma') spoken in a list-reading style by one female native speaker of Japanese. The model was trained on the first 18 of 24 repetitions of each word. The training algorithm led to 100% correct recognition of the remaining six repetitions of the words. Afterwards, the model was tested with a continuum from koma to komma, which was created by manipulation of a bilabial nasal duration in komma from 10% to 170% of the original duration (Tokens 1 to 17 in Figures 1 and 2). The second simulation used two English words (coma and comb) — the two most frequent responses when American-English students were asked to identify both koma and komma in a pilot experiment. The ratio of coma and comb was approximately 74.3% to 25.7% in the pilot data. The training algorithm led to 100% correct recognition of coma and comb. The model replicated the American-English mean ratio of coma and comb responses to tokens in the koma-komma continuum (74.5% to 25.5%). Figure 1 shows the model's response probabilities for
the tokens in the koma-komma continuum for the monolingual Japanese and monolingual English models. Categorical perception of nasal duration is observed only in the Japanese simulation. This might suggest that enough instances in Japanese enable people to acquire this contrast. If this is the case, non-native listeners can acquire this contrast after acquiring enough Japanese words. The last simulation explored this possibility.

The last simulation dealt with American-English listeners who learned Japanese as a second-language. We manipulated the proportion of Japanese and American-English exemplars stored in the model's memory. Figure 2 shows the model's response probabilities for koma in the koma-komma continuum. When there were fewer Japanese instances than American-English instances, categorical perception of the continuum did not occur (12:36; 12 Japanese instances: 36 English instances and 24:36). However, when the number of instances is the same in the two languages, categorical perception is observed (36:36). Moreover, when the proportion of American-English instances is smaller than that of Japanese ones, the same phenomenon is observed (36:24 and 36:12). These results suggest two possible interpretations. First, storing enough Japanese instances of category members is necessary for non-native listeners to acquire a meaningful contrast. Second, suppressing knowledge of L1 (for example, through a language 'set') enhances a non-native listeners' ability to adopt a more appropriate listening strategy in L2.

**CONCLUSION**

We investigated whether categorical perception of a segmental duration contrast arises from language experience. Simulations of native and non-native listeners of Japanese confirmed that this is true of the Japanese nasal duration distinction. Simulations of 'bilinguals' showed that second-language learners of Japanese can acquire this categorical perception when they acquire enough instances in Japanese — or when native language knowledge is suppressed.

**ACKNOWLEDGMENTS**

We thank Allison Blodgett, Laurie Maynell and Kaori Miyaishi for their help with this study. This study was supported by NIH FIRST Award R27-DC01645-06, granted to the second author.

**REFERENCES**