AN EXPLORATORY STUDY ONTO THE ROLE OF PHONEMIC CATEGORIES IN SPEECH PERCEPTION

Yun Zheng[1], Christian Giguère[2]
[1]Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6 zyun98@yahoo.com
[2]School of Rehabilitation Sciences, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5 cgiguere@uottawa.ca

1. INTRODUCTION

Stevens (1989; 2002) proposed that phonemic categories play a fundamental role in speech perception. According to this view, there is a process for specialized speech perception mechanisms to extract prototypical categories from variable acoustic inputs, and discard the variability to unmask the underlying phonetic or featural prototypes. One important support for this model is said to come from studies on categorical perception. For example, Cienfuegos et al. (1999) presented evidence that the deficit in categorizing acoustic properties of speech sounds may hinder the analysis of phonemes, and lead to the difficulty in language processing in schizophrenia.

Nonetheless, the assumed role of categorical perception in speech perception has been increasingly challenged in recent years. One major objection is related to the definition of categorical perception itself: true categorical perception is supposed to occur when discrimination is completely predicted by categorization, but this has never been demonstrated experimentally (McMurray et al., 2003). Schouten et al. (2003) pointed out that most discrimination tasks in categorical perception studies are subject to bias effects: the nature of the task usually compels subjects to use a labeling strategy. Subjects may prefer one response type to another, or they may refer the stimuli to an internal criterion of their own. When response bias is eliminated, categorical perception seems to fail. Furthermore, some studies have shown that listeners preserve sensitivity to fine-grained acoustic variation within phonetic categories in patterns of lexical activation (McMurry et al., 2002). Though small acoustic differences in Voice Onset Time (VOT) have minimal effects on phoneme identification, they have significant effects on lexical access.

Apart from the problems with categorical perception itself, the exact relationship between categorical perception and speech perception should be further studied. One intuitive inference from Stevens' model would be that speech perception is realized on the basis of the perception for individual phonemic categories. We will test such an inference in the present study.

2. METHOD

Three English words (IN, SIN, TIN) were read by a young native female English speaker, and digitally recorded using the VisiPitch IV system (Kay Elemetrics Corp.). The stimuli were further edited using the software Cool Edit Pro. Five categorically well-defined tokens (S = /S/ from SIN, T = /T/ from TIN, A1 = IN, A2 = /IN/ from SIN, and A3 = /IN/ from TIN) were extracted from the original stimuli. Six synthetic words were then constructed from these five tokens (A2, A3, SA1, SA3, TA1, TA2).

A 6x2 factorial design was adopted for the experiments. Factor one (Words) has six levels (the six synthetic words), and Factor two (Hearing Conditions) has two levels (normal-hearing and hearing-impaired). In Part one, the perception of the six synthetic words with the three natural words was compared under normal hearing conditions. An ABX task format was employed to reduce bias effects. The first two words (A, B) were a pair consisting of one natural word and one synthetic word, and X was always the natural word. The stimuli were delivered through high quality earphones. Subjects were prompted to report whether X was the same word as either A or B after each presentation. The serial position at which the natural word occurred within each task series was also controlled. Thus a total of 36 (6 x 3 x 2) task series were presented.

In Part two, the three natural words and the six synthetic words were processed through a hearing loss simulation system (University of Ottawa, 2004) to simulate the aging-induced hearing loss corresponding to a 65 year-old man (normalized data according to ISO 1999). The hearing thresholds were around 20dB for lower frequencies (125 Hz to 1 kHz), but significantly increased to about 80dB at higher frequencies (6 kHz to 8 kHz). The same ABX task format was employed in Part two, the 36 series of stimuli differed only from those in Part one due to the simulated hearing loss. As a result, some phonemes (e.g., /S/) of certain words were acoustically suppressed.

Preliminary data is available for 4 university students, all native English speakers with normal hearing. Presentation of the stimuli was controlled by a customized computer program (Visual C++). The order of presenting the 36 task series within each part was randomly determined by the computer program. The order of completing Part one and two was also counterbalanced between subjects.

3. RESULTS

In most cases, native English speakers perceived the synthetic words to be distinguishable from the natural words, even when a synthetic word and a natural word consisted of tokens from the same categories (e.g., SIN and SA1). Specifically, they had little difficult in perceiving the difference between IN and all synthetic words ($F(5,84) = 1.260$, $p = 0.289$), also the difference between SIN and all synthetic words (($F(5,84) = 0.60$, $p = 0.70$). But subjects seemed to be confused by TIN and some of the synthetic words ($F(5,84) = 2.576$, $p = 0.032$).

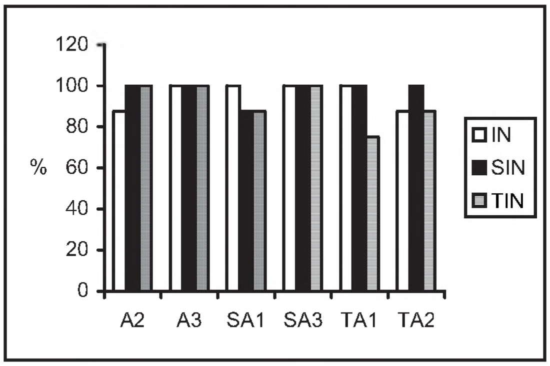

Fig. 1. The percentage of correctly identifying a natural word from a word-pair consisting of a natural word and a synthetic word under normal hearing conditions

When hearing loss was involved, the acoustic cue for the /S/ was largely suppressed. Yet this suppression had little effect on the performance of identifying the natural words from synthetic words, either for IN ($F(1,84) = 0.70$, $p = 0.405$), or SIN ($F(1, 84) = 0.333$, $p = 0.565$), or TIN ($F(1,84) = 0$, $p = 1.0$).

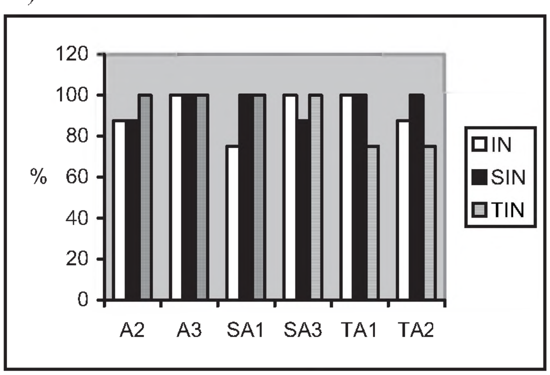

Fig. 2. The percentage of correctly identifying a natural word from a word-pair consisting of a natural word and a synthetic word when hearing loss was involved

No interactions between synthetic words and hearing conditions were significant. Interestingly, the suppression of /S/ by hearing loss made the acoustic waveforms for SA1 – IN, A2 – SIN hardly distinguishable. However, it resulted in little perceptual disturbance for the subjects. A paired t test showed that the effect by this acoustic suppression was not significant, $t(15) = 1.861$, and $p = 0.083$.

4. DISCUSSION

Critics to Stevens' model had raised doubts about some problems associated with categorical perception. But the exact relationship between categorical perception and speech perception had barely been discussed. In the present study we tested whether speech perception is realized on the basis of the perception for individual phonemic categories in speech sounds.

In our experiments, all the six synthetic words consisted of well-defined phonemic categories (extracted from natural words). However, the synthetic words were perceived differently from the natural words, even when both consisted of tokens from the same categories (e.g., SIN and SA1). With the suppression of certain categories due to simulated hearing loss, a synthetic word and a natural word (though consisting of different categories) might also become acoustically similar (e.g., SIN and A2), yet subjects had little difficulty perceiving the difference. These results seemed to suggest that speech perception does not depend on the perception for single categories. This is consistent with findings from studies on co-articulation effects, whereby speech sounds are not produced as isolated gestures, but are superimposed in a context-dependent fashion (see Fowler, 2005). Thus, speech perception is more likely based on perception for co-articulated categories rather than single phonemic categories. For example, when information about one category was unavailable, listeners could rely on information from another category that co-articulated with it. This may have implications for hearing aid processing.

REFERENCES

Cienfuegos, A., March, L., Shelley, A., and Javitt, D.C. (1999), Impaired categorical perception of synthetic speech sounds in schizophrenia, Biol. Psychiatry, 45, 82 – 88.

Fowler, C.A. (2005), Parsing coarticulated speech in perception: effects of coarticulation resistance, Journal of Phonetics, 33, 199–213.

McMurray, B., Tanenhaus, M. K., Aslin, R.N., Spivey, M.J. (2003), Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access, Journal of Psycholinguistic Research, 32(1), 77 - 97.

McMurray, B., Tanenhaus, M.K., Aslin, R.N. (2002), Gradient effects of within-category phonetic variation on lexical access, Cognition, 86, B33 – B42.

Schouten B., Gerrits, E., Hessen, van Hessen, A. (2003), The end of categorical perception as we know it, Speech Communication, 41, 71 – 80.

Stevens, K.N. (2002), Toward a model for lexical access based on acoustic landmarks and distinctive features, Journal of the Acoustical Society of America, 111(4), 1872 – 1891.

Stevens, K.N. (1989), On the Quantal Nature of Speech. Journal of Phonetics, 17(3), 3 – 45.