First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

EFFECTS OF DIFFERENT STANDARDS ON THE WITHIN-CATEGORY DISCRIMINATION OF SYNTHESIZED /ABA/ SEQUENCES : COMPARISON BETWEEN JAPANESE AND SPANISH

Minoru Tsuzaki * *Jorge A. Gurlekian* * **

* ATR Auditory and Visual Perception Research Laboratories
Soraku-gun, Seika-cho, Kyoto, 619-02 Japan

** Universidad de Buenos Aires- CONICET
Buenos Aires, (1453) CC-53 Argentina

ABSTRACT

To investigate the effects of the internal phonetic prototype on auditory discrimination, two groups of subjects with different phonological systems were tested for auditory discrimination of two different synthesized speech continua. One group was Japanese, and the other was Spanish. Each group received an AX discrimination task both with a typical Japanese /b/and a typical Spanish /b/ as the standard stimulus. The estimated dispersion width for a physically identical stimulus pair was smaller when the standard was a typical realization of the phoneme /b/ for each group than when it was atypical. Thus, even in a discrimination task within a category, the difference in the phonological system results in some behavioral differences. These behavioral differences are discussed in relation to a long term memory factor, or an internal prototype which may be involved in auditory information processing of speech sounds.

INTRODUCTION

The role of phonemic categories in speech perception has been one of the central issues in speech science and the phenomena associated with categorical perception have been investigated actively. Although the original assumption of a unique and discrete recording of speech stimuli, which disables the discrimination within phomene category [1], has been criticized empirically, it is still important to investigate the effects of acquired and/or inherent categories on speech perception.[2]

Repp and Liberman suggested that we should take account of the prototypes as well as the boundaries [3], and some researchers have investigated the role of "category prototypes" using the paradigm of selective adaptation [4, 5]. Although these studies focused on the prototype, the effects themselves were measured primarily in terms of the boundary shifts, which were not related with the phenomena around the category prototype but with the phenomena around category boundary. In order to discuss the psychological validity of the prototype, we should deal with the phenomena around the category center rather than the category boundary.

The purpose of the present paper was to investigate the effects of the different standard stimuli on within-category discrimination. The fundamental task of the subjects was to compare two synthesized VCV sequences with three response categories. The task can be classified as the AX discrimination paradigm, in which a fixed standard stimulus A is followed by a comparison stimulus X varying at each trial. The Japanese- and Spanish-native listeners responded to each of two experimental sets with the different standard stimuli, one of which was close to the typical realization of the phoneme class in Japanese and the other was close to that in Spanish.

Guirao and Luis reported that Spanish-native listeners could correctly identify 89% of Japanese CV syllables and claimed that phonetic systems between Japanese and Spanish are similar [6]. There are, however, cases that the typical realization of the common phoneme is classified in the other phonetic categories between Japanese and Spanish. One example is intervocalic /b/, which is classified, by phoneticians, as the plosive [b] in Japanese and as the approximant [ß] in Spanish.

METHOD

Subjects

Two groups of subjects participated in the experiment. The Japanese Group consisted of 12 Japanese native listeners who were undergraduates of Doshisha University. The Spanish group consisted of 12 Spanish native listeners who came from various South American countries — Argentine, Mexico, Colombia, and Peru.

Stimuli

Using the first four formants, two types of the stimulus continua were synthesized with a Klatt parallel formant synthesizer installed on a Masscomp workstation (MC5600). One was called the Japanese continuum, and the other the Spanish continuum. In each continuum, the approximate stimulus for the typical realization of the /aba/ in each language was set as the standard stimulus, and the comparison stimuli differed from the standard stimulus in two direction. One direction was called "negative" and the other was called "positive". To obtain such stimuli, the following parameters were selected to change: (1) the formant amplitudes of the inter-vowel part; (2) the transition time of the formant frequency between the vowel and the inter-vowel part; (3) the transition time of the formant energy between the vowel and the inter-vowel part. The other parameters

Table. 1 Common parameters for synthesis

LENGTH :	500.0 ms
SAMPLING FREQUENCY :	10.0 kHz
FRAME RATE :	5.0 ms
WINDOW LENGTH :	0.0 ms
F0 FRAME RATE :	5.0 ms
F0 WINDOW LENGTH :	0.1000000 ms
FIRST GLOTTAL RESONATOR FREQUENCY :	0 Hz
FIRST GLOTTAL BANDWIDTH :	50 Hz
SECOND GLOTTAL BANDWIDTH :	100 Hz
GLOTTAL ZERO FREQUENCY :	6000 Hz
GLOTTAL ZERO BANDWIDTH :	1000 Hz
F1 & A1 OF VOWEL PART:	800 Hz (80 dB)
F2 & A2 OF VOWEL PART:	1200 Hz (80 dB)
F3 & A3 OF VOWEL PART:	2443 Hz (70 dB)
F4 & A4 OF VOWEL PART:	3300 Hz (50 dB)
F1 OF INTER-VOWEL PART:	244 Hz
F2 OF INTER-VOWEL PART:	864 Hz
F3 OF INTER-VOWEL PART:	2443 Hz
F4 OF INTER-VOWEL PART:	3300 Hz
F1 BANDWIDTH:	50 Hz
F2 BANDWIDTH:	70 Hz
F3 BANDWIDTH:	110 Hz
F4 BANDWIDTH:	250 Hz
END POINT OF STEADY STATE OF F1 :	160 ms
RESTART POINT OF STEADY STATE OF F1 :	275 ms
END POINT OF STEADY STATE OF F2 :	150 ms
RESTART POINT OF STEADY STATE OF F2 :	285 ms
END POINT OF STEADY STATE OF A1-A4 :	170 ms
RESTART POINT OF STEADY STATE OF A1-A4 :	265 ms

were common to all the stimuli as shown in Table. 1. Below, the amplitude and frequency of each formant will be abbreviated to A1, A2, A3, A4, F1, F2, F3, and F4, respectively.

The standard stimulus in the Japanese continuum was synthesized to get an approximate stimulus for the Japanese /aba/ ([aba]). The A1, A2, A3 and A4 in the inter-vowel part were 60, 0, 0, 0 dB, respectively. The transition times for F1 and F2 were 10 and 20 ms, respectively. The transition time of the A1 was 6 ms, and those of the A2, A3 and A4 were all 8 ms. The comparison stimuli in the negative direction of the Japanese continuum were obtained by decreasing the A1 of inter-vowel part in 5-dB steps. The comparison stimuli in the positive direction were obtained by increasing the transition times of F1, F2, A1, A2, A3, and A4 in 6-ms steps. The relation among the stimuli in the Japanese continuum are schematically depicted in Fig. 1.

The standard stimulus in the Spanish continuum was synthesized to get an approximate stimulus for the Spanish /aba/ ([aßa]). The A1, A2, A3, and A4 in the inter-vowel part were 60, 55, 45, 35 dB, respectively. The transition times for F1 and F2 were 34, 44 ms, respectively. The transition time for all formant amplitudes was 24 ms. The comparison stimuli in the negative direction of the Spanish continuum were obtained by decreasing simultaneously the formant amplitudes except for A1 and all the transition times. The step size of amplitude decrement was 2.5 dB, and that of transition time decrement was 6 ms. The comparison stimuli in the positive direction were obtained by increasing simultaneously the formant amplitudes except for A1 and all the transition times. The step size of amplitude increment was 2.5 dB, and that of the transition time decrement was 6 ms. The relation among the stimuli in the Spanish continuum are shown schematically in Fig. 2.

The duration of the synthesized /aba/ sequence was 500 ms. In each continuum, the standard stimulus was paired with each of the nine comparison stimuli consisting of the four negative stimuli, the physically identical stimulus to the standard, and the four positive stimuli. The standard stimulus always preceded the comparison stimulus at an inter-stimulus interval (ISI) of 250 ms.

Procedure

Subjects were required to compare the standard and comparison stimuli and to respond with 3 response classes: "negative"; "equal"; or "positive". When they responded

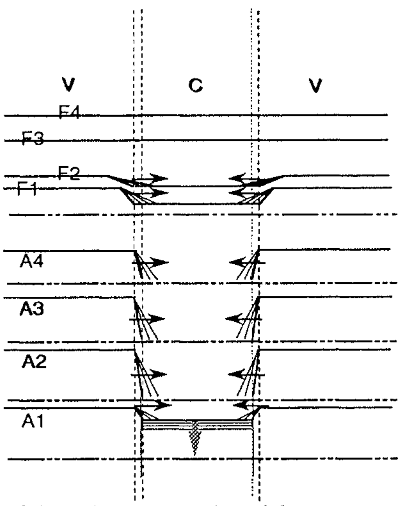

Figure 1. Schematic representation of the parameter settings of the stimuli in the Japanese continuum. The solid arrows depict the positive direction. The grayed arrow depicts the negative direction.

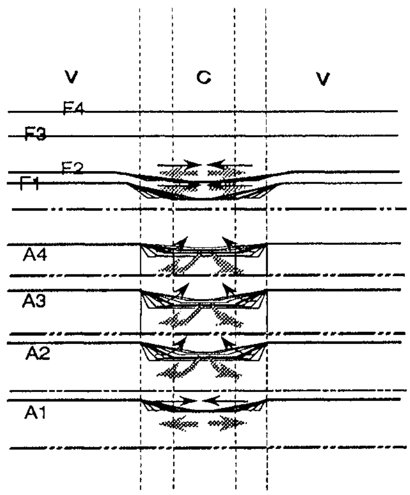

Figure 2. Schematic representation of the parameter settings of the stimuli in the Japanese continuum. The solid arrows depict the positive direction. The grayed arrow depicts the negative direction.

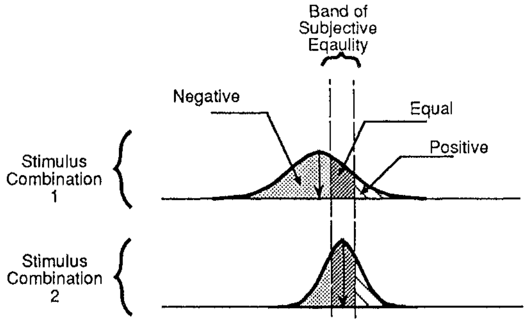

Figure 3. The relation between discriminal dispersions and three response classes.

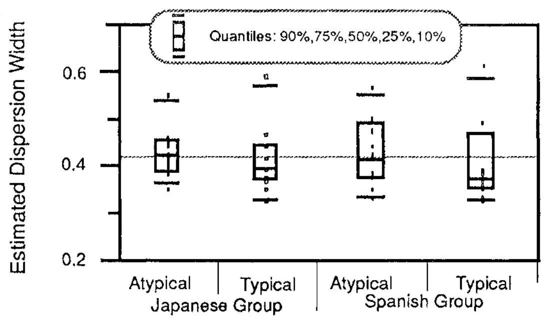

Figure 4. Box plots of the estimated dispersion width for the discriminal process of the physically identical combination. For the Japanese group, the atypical standard is [aßa]; the typical standard is [aba]. For the Spanish group, the atypical standard is [aba]; the typical standard is [aßa].

with either "negative" or "positive", they were also required to give confidence ratings by using an integer scale from 1 to 5.

Each subject received four blocks of 90 trials (360 trials in total). In each block, the stimulus continuum was fixed to either the Japanese or the Spanish one, and nine combinations of the standard and comparison stimuli were repeated 10 times each in random order. Half of the subjects received the first two blocks with the Japanese continuum and the latter two blocks with the Spanish continuum. The other half received in the reverse order.

Before they were tested in each continuum, subjects were presented with samples of the nine stimulus combinations. They also received 18 practice trials. If the results of the practice trials indicated that a certain subject had confused the negative and positive directions, he (or she) was given the samples and the practice trials again.

The randomized stimulus sequences were created on a workstation (Masscomp MC 5600) and a D/A converter (Pavec MD 8000MkII). They were low-pass filtered with a cut-off frequency of 5 kHz, digitally recorded on a DAT recorder (SONY DTC-1000ES), and then presented binaurally to the subjects through headphones (STAX SR-A Pro driven by STAX SRM-1/2 Pro) in a soundproofed room. Subject responses were recorded with a personal computer (Macintosh II) and MIDI control box (YAMAHA MCS-2).

RESULTS

Raw response categories to each stimulus combination consisted of the integers from -5 to +5. The sign indicates the direction in which a comparison stimulus was judged to differ from a standard stimulus. The number indicates the degree of confidence or the distance between the standard and the comparison stimuli, and the response category "zero" means that two stimuli were judged to be equal. We can assume that the relative frequencies of the negative responses, zero responses, and positive responses correspond to three areas of the discriminal process along the psychological continuum as shown in Fig. 3. Each stimulus combination can be presented as a dispersion located along the psychological continuum which has its own center and dispersion width. Thus, we can estimate the relative position and width of the dispersion from the proportions of three response categories.

In traditional psychophysics where the difference threshold or the point of subjective equality (PSE) is of main interest, the center of the discriminal process is important. In the present study, however, the width of the discriminal process is important, because it can reflect a variance of mental representation corresponding to a given stimulus environment.

If you follow the model depicted in Fig. 3, increase of the equal response will fundamentally lead to smaller estimation of the dispersion width. We must, however, consider response biases. Some subjects may give the equal responses very loosely in order to maximize the probability of equal response to the physically identical combination, but others may give them very strictly in order to minimized the probability of equal responses to the physically unequal combination. Such difference among response biases should not be represented as the width of the discriminal process itself, but should be represented as the width of the band of subjective equality. The separation of the response bias factor was tried following the estimation method used in the signal detection theory.

For example, we can estimate the proportions of the "equal" response class with the most strict criterion for equality by regarding only the "zero" category as the "equal" class. We can estimate the proportion with the second most strict criterion by regarding "-1", "zero", and "+1" categories as the "equal" class. Thus, the T-scores which correspond to the edge positions of the equality bands with several criteria can be obtained for each stimulus combination. On the assumption that the discriminal process is distributed normally with a standard deviation of "1", the band widths of subjective equality are given by these T-scores. Actually, the standard deviation of the discriminal process might vary among the stimulus combinations, and the band widths with the same criterion should be equal across the stimulus combinations. Thus, the dispersion width of each discriminal process can be estimated to minimize the variation among the band widths by the method of least squares.

The dispersion width of the discriminal process was estimated for each subject and each stimulus combination in each continuum. Figure 4 is the box plot of the estimated dispersion width for the physically equal combination. There

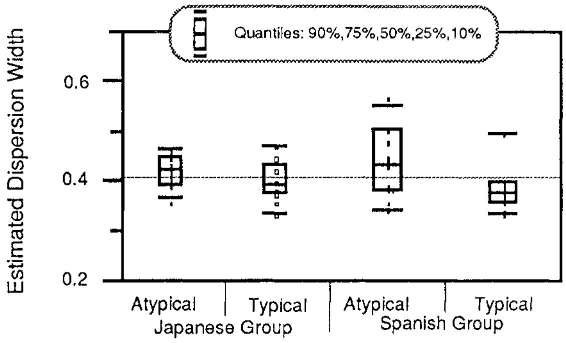

Figure 5. Box plots of the estimated dispersion width for the discriminal process of the physically identical combination after the removal of two subjects' data who showed deviation from the general tendency.

was a tendency for the typical standard ([aba] for the Japanese group; [aßa] for the Spanish group) to have smaller median scores than the atypical standard ([aßa] for the Japanese group; [aba] for the Spanish group). The 2 factorial ANOVA was applied with standard type and subject group as the two factors. The results, however, did not show any significant main effect.

The data for two subjects, however, deviated from the general tendency. Figure 5 shows the box plot after the removal of these two subjects' data. The results of ANOVA applied to the reduced data set with these two subjects removed showed a significant main effect of standard type (Typical vs. Atypical) [$F(1, 38) = 4.266$, $p < .05$].

DISCUSSION

Both the Japanese and Spanish groups showed a smaller median value of the dispersion width for the typical standard than for the atypical standard. The results indicate that the mental representation for the typical realization of a phomene class is more stable than that for the atypical realization. Because the typical standard for the Japanese group is physically equal to the atypical standard for the Spanish group and vice versa, the difference between the typical and the atypical standards cannot be explained by any physical factor. We must consider difference of schemata with which the human perceptual system represent the external stimuli.

The motor theory may provide one candidate for such a schema [1]. If the perceptual system refers to the motor commands which produce the sound akin to the input, the typical sound will be represented almost always in reference to the same motor command. Thus, the probability that the physically identical stimuli are judged to be perceptually equal to each other will be greater. On the other hand, the atypical sound, in some cases, will be represented in reference to motor commands associated with the production of the sound belonging to a different phoneme class. Thus, the probability that the physically identical stimuli are judged to be perceptually equal to each other will be less.

Massaro insisted, however, that the level of such motor commands asserted in the categorical model is too discrete to explain some experimental results by referring to that the frequency distribution of rating responses for a /b/ - /p/ continuum did not show the two peaks as the categorical model predicted [7]. In fact, according to their verbal reports after the experiments, all the subjects agreed that both of the standard stimuli sounded like /aba/, and noticed that the standard stimuli were different for the Japanese and Spanish continua. These reports indicate that subjects can categorize two perceptually different sounds into the same phonemic category. If the level of the motor commands is assumed to correspond to the level of the phonemic categories as by Liberman et al [1]., the motor theory cannot explain the results. One can still hold the motor theory by assuming that the level of the motor commands correspond to the level of the phonetic categories rather than the phonemic categories. In this case, however, we must take account of the degree of learning or other factors which can reflect differentiation among the phonetic categories.

In the present experiment, the fixed standard AX discrimination task with a relatively short ISI (250 ms) was adopted, in order to minimize the memory load and to enhance the perceptual mode of the process. The results suggest that an effect of the acquired phonemic system might be observed even in the perceptual process. Long term memory must be involved to store information available through the acquisition of a phonemic system. Although such a factor of the long term memory has been considered in the dual process model proposed by Fujisaki and Kawashima [8] and Pisoni [9], the labeling process and the auditory process are modelled as fundamentally discrete and independent processes. Information in terms of the "phonetic prototypes" mentioned in the present paper does not correspond to the labeling process and is assumed to interact with the purely auditory information at the stage of construction of auditory representations.

REFERRENCES

[1] A.M.Liberman, K.S.Harris, H.S.Hoffman, & B.C.Griffith, "The discrimination of speech sounds within and across phoneme boundaries," *J. Exp. Psychol.*, Vol.54, 358-368, 1957.

[2] R.E.Pastore, "Categorical perception: some psychophysical models," In S. Harnad (Ed.) , *Categorical Perception: The Groundwork of Cognition.* Cambridge Univ. Prs., N. Y. 1987

[3] B.H.Repp & A.M.Liberman, "Phonetic category boundaries are flexible," In S. Harnad (Ed.) , *Categorical Perception: The Groundwork of Cognition.* Cambridge Univ. Prs., N. Y., 1987

[4] A.G.Samuel, "Phonetic prototypes," *Perception & Psychophysics*, Vol. 31, 307-314, 1982.

[5] J.J.Miller & T.Baer, "Some effects of speaking rate on the production of /b/ and /w/," *J. Acoust. Soc. Amer.*, Vol. 73, 1751-1755, 1983.

[6] M.Guirao & C.R.Luis, "Identification of Japanese syllables by Spanish-speaking listeners," *J. Acoust. Soc. Jap.*, Vol.3, 21-26, 1982.

[7] D.W.Massaro, "Categorical partition: A fuzzy-logical model of categorization behavior," In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* Cambridge Univ. Prs., N. Y., 1987

[8] H.Fujisaki, & T.Kawashima, "Some experiments on speech perception and a model for the perceptual mechanism," *Ann. Report Engin. Res. Inst.*, Vol. 29, 207-214, Tokyo: Univ. Tokyo, Facult. Engin., 1970.

[9] D.B.Pisoni, "Auditory short-term memory and vowel percption," *Mem. & Cogn.*, Vol.3, 7-18, 1975.