A Synthetic Study of the Association of a Tone with a Voiceless Vowel

Kunitoshi TAKAHASHI

1. Introduction[1]

The high vowels /i, u/ in Japanese have long been paid much attention to in the literature (e.g. McCawley(1968), and Lovins (1976)) since they are apt to undergo such processes as devoicing or deletion, usually known as High Vowel Devoicing or High Vowel Deletion, respectively, which are interesting both descriptively and theoretically.[2] In this paper, we are concerned with experimentally investigating possibility of a tone to be associated with a voiceless vowel, /i/ in particular, in Tokyo Japanese.

1.1. Tone Patterns in the Tokyo Dialect

The tone pattern (also called accent type, accent pattern, tone melody, pitch contour, etc.) of a word in the Tokyo dialect is determined according to whether or not the word carries an accent, and if it does, on which mora the accent falls. Thus, words consisting of *n* morae can exhibit *n*+1 tone patterns. For example, four-mora words manifest five tone patterns as shown in Table 1.

Table 1 Tone Patterns of Four-Mora Words in the Tokyo Dialect

Tone Patterns		Exmaples	
1	**H**LLL	/siitake/	"mushroom"
2	L**H**LL	/irogami/	"color paper"
3	LH**H**L	/kagaribi/	"bonfire"
4	LHH**H**↓	/otooto/	"younger brother"
0	LHHH→	/sanpatu/	"haircut"

In Table 1, boldfaced H(igh) tones are meant to be accented. A word of tone pattern 0 does not have an accent, which is shown with a horizontal rightward arrow put at the end. Tone pattern *n* (*n* > 0) means that an accent falls on the *n*th mora of the word, making the immediately following mora L(ow) in pitch. For tone pattern 4, a downward arrow is put at the end to suggest this effect. In tone patterns 2, 3, 4, and 0, the initial mora is L and the second mora H in pitch, due to a constraint in the

Tokyo dialect, sometimes called Initial Lowering, which says that the initial mora must be L in pitch if the second is H. For more sophisticated descriptions, see e.g. Haraguchi (1991).

1.2. Problems

Lovins (1976, 118) claims that if a phonologically H-toned vowel is devoiced in Japanese, there is no phonetic evidence indicating H pitch. Truely, this statement holds for the time span of devoicing because when there is no vocal cord vibration, F_0, fundamental frequency, which determines the pitch, is absent. However, she also means by this that there is no phonetic influence caused by the phonologically H tone of the devoiced vowel on the surrounding morae.[3]

Haraguchi (1977), on the other hand, claims in his autosegmental analysis that a devoiced vowel is not a "tone-bearing" element (p.36), which is restricted to a voiced vowel or a syllabic nasal, and that it thus cannot carry a tone (p.8). The same is true of Haraguchi (1991).

Lovins' and Haraguchi's claims, however, are both counterintuitive to the ears of native speakers of the Tokyo dialect. The present investigation was to experimentally prove that this native intuition is correct. Cf. Fujisaki et al. (1974); Fujisaki and Sugito (1976).

As a byproduct, we examined whether or not tone patterns in the Tokyo dialect are categorically perceived, and, if so, to what extent they are so perceived.

2. Sonagraphic Observations

As a preliminary investigation, let us first examine some spectrograms (obtained by Series 700 Sound Spectrograph, Voice Identification, Inc.). Figure 1 is the sonagram of the word *kagaku* "science," which is of tone pattern 1 (thus HLL), and Figure 2 is the sonagram of the word *mogaku* "to struggle," which is of tone pattern 2 (thus LHL). What should be noticed in these figures is that a steep drop in pitch is observed starting at the beginning of the L-toned mora that immediately follows the H-toned mora.[4] Now campare these with Figures 3 and 4. Figure 3 shows the HLL word *sIkai* "The Dead Sea," and Figure 4, the LHL word *sIkai* "dentist," with the voicelessness of the first vowel in both words expressed with the capital letter *I*. Clearly the same or an identical pitch contour occurs in the corresponding figures in spite of the fact that the initial vowel is voiceless in Figures 3 and 4. What is significant is that Figure 1 is similar to Figure 3, and Figure 2 to Figure 4, with respect to the location of a steep drop in pitch. Cf. Sugito (1969); Hattori (1975).

Moreover, comparing Figures 5 and 6, the two possible tone patterns of *yasasIkatta* "gentle+PAST," i.e. LHLLLL and LHHLLL (the latter being preferred by younger generations), in both of which the third mora is voiceless, we find a clear difference with respect to the position of a steep drop. Although we do not see any steep drop on the antepenultimate mora in the former tone pattern, there is one in the latter. This fact, it seems, convincingly refutes Lovins' (1976, 118) claim that:

> Such a partial deletion 〔= nonexistence: K.T.〕 of the pitch contour might however result in some confusion if it occurred in a context like ... /LHHL/, with the third mora voiceless. Then it would be unclear, if this mora were perceived as pitchless, whether the accent really fell on the second or the third mora.

Given these sonagraphic facts concerning the position of the steep drop in pitch, it is plausible to conclude as a first approximation that a voiceless mora can be associated with a tone, and that there is enough acoustic influence of the phonologically H tone of a voiceless mora on the immediately following mora. In fact, the acoustic influence is comparabel to that of a voiced mora. The next question is whether this acoustic evidence is auditorily significant or not, and this is the question that we try to answer in the following sections on the basis of the results of our perceptual experiments.

3. Experiment I

It has been known that the pitch contour i.e. the F_0 coutour, among other factors, such as segment duration and intensity, is sufficient to cue perception of the tone pattern (cf. Fujisaki, Morikawa and Sugito (1976) and the references cited there). The purpose of the first experiment was to confirm that F_0 is a sufficient cue to differentiate tone patterns even if stimuli are four-mora nonsense words, which do not contain voiceless vowels.

3.1. Method for Experiment I

3.1.1. Subjects for Experiment I

Eight volunteers (6 males and 2 females) served as the subjects for Experiment I. Their average age was 25.4 years old, varying form 21 to 30. All of them were linguistically naive native speakers of the Tokyo dialect.

3.1.2. Apparatus for Experiment I

An OVE IIId speech synthesizer was used to generate the test stimuli for this experiment. The relevant controllable variables, inter alia, were AV (Amplitude of Voicing), AN (Amplitude of Nasals), F_0, F_1, F_2, F_3, N_1 (Nasal Formant 1) and B_1

(Bandwidth of F_1).

3.1.3. Stimuli for Experiment I

We synthesized a four-mora nonsense word *mamamama*, on which the different pitch contours, varying from tone pattern 2 i.e. LHLL to tone pattern 3 i.e. LHHL, were superposed to yield ten different stimuli. For the nasal part, N_1=ca. 260 Hz, F_1 =ca. 260 Hz, F_2=ca. 1100 Hz, and F_3=ca. 2800 Hz. For the vowel part, on the other hand, F_1=ca. 850 Hz, F_2=ca. 1250 Hz, and F_3=ca. 2900 Hz, with a 30-msec transition on both ends. Formant frequencies for vowels were obtained by considering data then available (e.g. Kohno and Takahashi (1983); Fujisaki, Morikawa and Sugito (1976); Han (1962)) and the data from our supplementary measurements. The details were decided by trial and error so as for the spectrograms of the synthesized sounds to sound natural or at least similar to those by native Tokyo speakers.

Figure 7 shows the rough F_0 contours of the ten stimuli. We considered pitch contours that Fujisaki et al. (1976) obtained by means of analysis-by-synthesis and our preexperimental exploration to get the contours in Figure 7.

The time at which a steep drop in pitch starts will be called the Drop-Starting Time (DST). The DST of each stimulus is shown in Table 2. The difference in DST between adjacent stimuli was fixed at 25 msec except that there was a 20-msec difference between stimulus 1 and stimulus 2. As can be seen in Figure 7, the durations of all the eight segments of *mamamama* were approximately, from the first, 55, 120,

Table 2 DST of Each Stimulus in Experiment I

Stimulus Number	DST (msec)	Segments
1	355	/a/
2	375	
	380	------
3	400	
4	425	/m/
5	450	
	465	------
6	475	
7	500	
8	525	/a/
9	550	
10	575	
	590	------

85, 20, 85, 120, 90 and 135 in msec. Thus, the entire stimulus length was 810 msec.

3.1.4. Procedure for Experiment I

For tone pattern identification, the subjects were presented with the randomized stimuli, in which each stimulus appeared ten times, through headphones, and instructed to judge the tone pattern of each stimulus as either tone pattern 2 or tone pattern 3, but not as others. The number of judgments per subject were thus just one hundred. There was a ca. 3.5-sec interval between stimuli, with a longer pause followed by a beep after every five stimuli, which enabled the subjects to know on which part of the experiment they were working.

To test discriminability, the SAME-DIFERENT(or AX) method was adopted, in which the stimuli are arranged pairwise, A and X being either the same or different (cf.Wood(1975)). The subject were required to indicate whether the members of a pair sounded the same or different. Discrimination measured were between each stimulus and those which were three steps away from each other on the stimulus scale. This discrimination distance was determined according to our pretest. This gave us seven pairs. Each obtained pair appeared ten times, five times in one order e.g. (3,6), five times in the permuted order e.g. (6,3), which is to counterbalance series effects or order effects. There were then seventy trials, each of which consisted of different stimuli. Call those pairs DIFFs. Also prepared were ten pairs that consisted of the same stimuli. Call these SAMEs. These pairs appeared seven times, which gave us seventy trials. All these pairs were presented to the subjects in random order through headphones, the number of trials for each subject being one hundred and forty. The interval between the two stimuli within a pair was set at ca. 0.8 sec and that between pairs, ca. 3.8 sec, with a longer pause followed by a beep after every five trials.

The indentification task was presented before the discrimination task for each subject. It took each subject about thirty minutes including a 2-minute break between the two tasks to finish both.

3.2. Results of Experiment I

3.2.1. Identification Task of Experiment I

The results of the identification task of Experiment I are shown in Figure 8, in which the mean identification functions for the tone patterns are presented. The abscissa is DST expressed in terms of the stimulus number, and the ordinate is the percentage identification of the stimuli as either tone pattern 2 or tone pattern 3. These data are similar to those obtained for stop consonants by Liberman et al. (1957) and for VOT by Wood (1975) except that they show near 90%, rather than near 100 %, identification toward the ends of the abscissa, and that they show a less abrupt boundary. These two differences, however, are attributable to the averaging of individual functions. That is, six out of eight individul casese showed near 100% identification toward the ends of the abscissa. Also, each subject had a tone pattern boundary at a different DST. Six out of eight had a steep boundary. Therefore, it can be said that our results here exhibit functions typical of the categorical mode.[5]

The boundary between tone patterns 2 and 3 was detected with the naked eye based on the functions appearing in Figure 8 as 6.5 on the stimulus scale (or ca. 488 msec in DST), which is in the middle of the third mora of the stimuli. It is shown by an upward arrow in the relevant figures.

The same data were also analyzed in the signal detection theory described in McNicol (1972) in terms of a discriminability parameter i.e. the sensitivity index *d′* to yield Figure 9, in which *d′* for each pair of adjacent stimuli is plotted.[6] Between the stimuli (1,2), (2,3), (3,4), (4,5), (8,9), and (9,10), *d′* may be considered a little above the chance performance, whereas between the stimuli (5,6), (6,7) and (7,8), it is comparatively high.

3.2.2. Discrimination Task of Experiment I

In Figure 10 are displayed the mean discrimination functions for SAMEs and DIFFs with three-step distance of Experiment I, the former being shown with closed circles connected with solid lines, and the latter with open circles connected with dotted lines. The curve for SAMEs is dented around the tone pattern boundary and shows near 90 % or more correct rejections (see note 6 for "correct rejections") within tone patterns. The curve for DIFFs is protruding upward with the top at the boundary. While the first two pairs give ca. 20% hits (see note 6 for "hits"), the pairs around the boundary give as high as ca. 90% hits.

Figure 11 shows the mean *d′* for each stimulus pair, in which the two stimuli are three steps away from each other. The value of *d′* is near the chance level for the pairs (1,4) and (2,5), goes up as DST increases until it reaches the highest value at the tone pattern boundary, and goes down after DST is later than that of the boundary.

The mean $\log_{10}\beta$ for each stimulus pair is in Figure 12, which is comparable to Figure 11.[7] When the stimulus pair is away from the tone pattern boundary, a bias toward SAME responses is observed, which is shown by the positive value of $\log_{10}\beta$. However, $\log_{10}\beta$ decreases toward the boundary, i.e., the value of $\log_{10}\beta$ is negative around it, which is an indication of a bias toward DIFFERENT responses.

4. Experiment II

The purpose of the second experiment was to investigate if F_0 is or is not a sufficient cue in cases where a voiceless vowel is involved, and if so, to what extent it is sufficient.

4.1. Method for Experiment II

4.1.1. Subjects for Experiment II

The subjects for the second experiment were eight volunteers (6 males and 2 females) with their average age 25.5, varying from 22 to 30. Again, as in Experiment I , all of them were linguistically naive native speakers of the Tokyo dialect.

4.1.2. Apparatus for Experiment II

The same equipment was used as in Experiment I . The variables controlled were AV, AN, F_0, F_1, F_2, F_3, N_1, B_1, AH (Amplitude of Aspiration), AC (Amplitude of Frication), K_1 (Fricative Formant 1), K_2 (Fricative Formant 2) and AK (Fricative Pole/Zero Ratio).

4.1.3. Stimuli for Experiment II

A four-mora nonsense word *masIkama* was first synthesized. Ten diffrerent pitch contours were superposed on it to obtain ten different stimuli, whose pitch contours vary from tone pattern 2 to tone pattern 3. For /m/, the values of variables were set as in Experiment I . For /a/, the values were also set as in Experiment I . For/s/, K_1 =ca. 4300 Hz, K_2=ca. 5400 Hz, and AK=0. For /i/, F_1=ca. 250 Hz, F_2=ca. 2300 Hz, and 3200 Hz. For /k/, K_1=ca. 1600 Hz, K_2=ca 4800 Hz and AK=31. The details were determined by trial and error just as in Experiment I .

We used approximately the same F_0 contours as in Experiment I .[8] There was a minor difference in stimulus duration: 820 msec in Experiment II and 810 msec in Experiment I . The difference of only 10 msec was unnoticeable and thus ignored for the purpose of our experiments. DST was set the same as in Experiment I , so that it would be easy to compare the results of Experiment I with those of Experiment II. As shown in Table 3, the ten stimuli used in Experiment II were numbered from 2 to 11. Stimulus 3 in Experiment I , for instance, corresponds to stimulus 3 in Experiment II in DST. The durations of all the segments in msec were, as shown in Figure 13, from the first, /m/=ca. 55, /a/ = ca. 120, /s/=ca. 110, /i/ = ca. 55, /k/ = ca. 25, /a/ = ca. 150, /m/ = ca. 90, and /2/ = ca. 145, with silence of ca. 70 in msec between /i/ and /k/. F_0 is absent during the time span of the second mora /si/ and the consonant of the third mora /k/ because they are voiceless, which is indicated by dotted lines in Figure 13.

Table 3 DST of Each Stimulus in Experiment II

Stimulus Number	DST (msec)	Segments
2	375	
3	400	/i/
	410	------
4	425	/k/
	435	------
5	450	
6	475	
7	500	/a/
8	525	
9	550	
10	575	
11	585	------
	600	/m/

4.1.4. Procedure for Experiment II

The procedure for Experiment II was the same as in Experiment I .

4.2. Results of Experiment II

4.2.1. Identification Task of Experiment II

Figure 14 shows the mean identification functions for the tone patterns in the identification task of Experiment II. The abscissa is DST expressed in terms of the stimulus number, and the ordinate is the percentage identification of the stimulus as either tone pattern 2 or tone pattern 3. These data do not look categorical as those of Experiment I . In other words, they are more of the coutinuous mode than the categorical mode. Again, as in Experiment I , this fact appears to result from averaging of individual functions. That is to say, three subjects had an abrupt boundary, which is comparable to the ones in Experiment I , and another subject, a considerably abrupt boundary. Two other subjects, on the other hand, had a gradual slope with no clear boundary, and two other subjects, near flat responses. Figure 15 shows the mean identification functions of the former four subjects, who we will call the categorical subjects. It is clear, in this figure, that the functions for those subjects are highly categorical.

The boundary between tone patterns 2 and 3 determined with the naked eye according to the functions in Figure 15 was, approximately, 6.5 on the stimulus scale (or 488 msec in DST), which is located in the midst of the third more. Although the functions in Figure 14 is less categorical, the boundary determined was almost the same i.e. 6.5 The value obtained in Experiment I was in accord with this value. The boundary is shown by an upward arrow in the relevant figures.

Figures 16 and 17 exhibit the same results analyzed in terms of the sensitivity index *d'* for the entire eight subjects and the four categorical subjects, respectively. We plotted *d'* for each pair of adjacent stimuli. Nevertheless, the values of *d'* for the stimulus pairs (2,3), (3,4), (4,5), (5,6) and (7,8) are a little above the chance performance level; whereas, the value for the pair (6,7) is considerably high. The pattern in Figure 16 is less obvious. The reason that the values for the pairs (8,9), (9,10) and (10,11) are higher (than expected for categorical perception) is not surprising. Namely, since the number of trials was not large enough, a small difference in probability density (or percentage) toward the end of a standard normal distribution is considred to have brought about a large value of *d'*. It appears that this kind of deviation (or unexpected pattern) can be ignored where there are not enough trials, since if there are more trials, it is likely that the pattern becomes more categorical.

4.2.2. Discrimination Task of Experiment II

In Figures 18 and 19, the mean discrimination functions for SAMEs and DIFFs with three-step distance of Experiment II are displayed for the entire subjects and the categorical subjects, respectively. The curve for SAMEs is slightly dented toward the tone pattern boundary. The cuve for DIFFs is protruding upward with the top around the boundary. These characteristics are more obvious for the categorical subjects. The curves for the categorical subjects are by and large similar to the corresponding curves of Experiment I .

Figure 20 represents the mean *d'* for each stimulus pair for the entire subjects, in which the two stimuli are three steps away from each other. Figure 21 is for the categorical subjects. The value of *d'* is near the chance level for the pairs (2,5) and (8,11), and is getting higher toward the tone boundary.

The mean $\log_{10}\beta$ for each stimulus pair is found in Figure 22 for the entire subjects, and in Figure 23 for the categorical subjects. Both curves are similar in shape, i.e., they are both concave. The curve for the entire subjects, on the one hand, does not manifest a bias toward DIFFRERNT responses, i.e., it does not have any negative values of $\log_{10}\beta$. The curve for the categorical subjects, on the other, shows a bias toward DIFFRERENT responses around the tone pattern boundary, which is indicated by the negative value of $\log_{10}\beta$, and a bias toward SAME responses where the stimulus pair is away from the boundary, which is indicated by the positive value of $\log_{10}\beta$.

5. Discussion

The results of Experiment I manifest a higher discriminability between tone patterns i.e. around the midst of the typical pitch contours of tone patterns 2 and 3 than the other pitch contours within each tone patterns, which we take to be evidence for existence of a boundary effect.[9] The results also buttresses our intuition that the F_0 contour alone is a sufficient cue to perception of the tone pattern in the case of a four-mora nonsense word.

The results from Experiment II are, however, not straightforward. Although the results from the categorical subjects are of the categorical mode, the results from the entire subjects (ultimately for the "noncategorical" subjects) exhibits characteristics typical of the continuous mode. The reason that we obtained such results can be that besides F_0, the noncategorical subjects use other cues, such as intensity, duration, etc., and that F_0 alone is not a sufficient cue for them to perceive tone patterns if the stimuli include voiceless sounds. If this line of reasoning is correct, then it is expected that if we provide enough cues, even the noncategorical subjects would show a categorical effect.

In any case, at least for the categorical subjects, a comparatively high discriminability at around the tone pattern boundary is observed, just as in Experiment I. Thus, it is plausible to conclude that at least as far as the categorical subjects are concerned, there is a boundary effect in perception of tone patterns whether voiceless sounds are involved or not, and that the acoustic influence caused by the phonologically H tone on a voiceless mora enables us to auditorily perceive the physically absent phonological tone of the voiceless mora. Also, Haraguchi's (1977) claim that a voiceless vowel cannot carry a tone has now experimentally been proven groundless. Hence, it needs to be abandoned or at least revised in one way or another.

Our data from Experiment I and Experiment II for the categorical subjects also lend support to Wood's (1975) claim that an additional perspective on the phoneme boundary effect is the systematic change in response bias from one toward SAME responses within categories to one toward DIFFERENT responses around the category boundary. If "phoneme boundary" and "category" in his claim are substituted by "tone pattern boundary" and "tone pattern," respectively, then the claim seems to hold for tone patterns. Following Wood, we propose this change in response bias as a diagnostic of existence of a boundary in general.

Note in passing that Miller et al. (1974) and Kuhl and Miller (1975) present data suggesting existence of a perceptual discontinuity for noise-buzz judgments in a temporal region, a discoutinuity similar to the voiced-voiceless VOT distincion in English. They query the role of phonetic categorization in the phoneme boundary effect for VOT. That is, as Wood(1975) states, it may be the case that "the phoneme boundary effect in VOT discrimination arises not as result of some specialized phonetic categorization process, but as a consequesnce of ... a more general property of auditory temporal sensitivity." Since our data are concerned with a suprasegmental feature, pitch, rather than phonemes, they render support to Wood's claim that a boundary effect in general arises as a consequence of a general property of auditory temporal sensitivity.

6. Conclusion

In this paper, we have challenged Haraguchi's (1977) treatment of the relation between a voiceless sound and a tone. His claim, which is reminiscent of Lovins' (1976) claim that there is no phonetic evidence of indication of a phonologically H tone when associated with a voiceless vowel, is that a voiceless vowel cannot be associated with a tone since voiceless vowels are, in his analysis, not tone-bearing elements. Pursuing his claim, nevertheless, would bring us to phonological counterexamples, to which his analysis as such gives illicit tone patterns.[10] Moreover, a series of perceptual experiments we conducted has revealed, contrary to Haraguchi's and Lovins' claims, that at least for a class of subjects we named the categorical subjects, a phonetic cue observable on spectrograms i.e. F_0 can serve as an auditory cue for perception of tone patterns even when voiceless vowels are involved. Hence, a proof that a voiceless vowel can carry a tone.

As a byproduct of the experiments, we have seen that at least for the categorical subjects, tone patterns belong to the categorical mode rather than the continuous mode irrespective of whether the stimuli involve a voiceless vowel or not.

We have examined the effect of only F_0 as a potential cue for the perception of a ponologically H tone. It is a suficient indicator for some, but not all, of the subjects; this in itself is unexpected and incoherent with Haraguchi's (1977) and Lovins' (1976) analysis. Moreover, there may be other ways than F_0 by which a H tone is cued, and, if so, once such other phonetic cues have been identified, we would probably find the noncategorical subjects responding just as the categorical subjects do.

Our research has shown, in sum, that it is possible for phonetic findings to support or refute phonological proposals.

Notes

1. I am very grateful to Keiko Otuka and Yashy Tohsaku who judged the naturalness of the synthesized sounds and encouraged me in many ways. Many thanks also go to Jeff Hardy who taught me plenty of things about the operation of the synthesizer.

Without Jeff Elman's advice, comments, and assistance, this work could not have been finished. Finally, I wish to thank all the subjects for cooperation. Of course, all remaining errors and inadequacies are due only to me.

The accent dictionaries used were Hirayama (1960), Kindaichi (1958), and Nihon Hoosoo Kyookai (1966).

2. Roughly, high vowels get devoiced between voiceless consonants in Japanese. For details, see Haraguchi (1984).

See Sawashima (1969) for a glottographic study of devoiced vowels; Sawashima (1971), Sawashima et al. (1971), Sawashima and Miyazaki (1973) for fiberscopic studies; Weitzman et al. (1976) for the difference between devoiced and whispered vowels.

3. Lovins, furthermore, purports to say alongside that when a phonologically L-toned vowel is devoiced, there is some evidence of indication of L pitch. This claim does not appear to be supported. Cf. Fujimura (1971).

4. It has been suggested e.g. by Simada and Hirose (1971), and Simada, Hirose, Sawashima and Fujimura (1971) that the relaxation of the cricothyroid muscle and the contraction of the sternothyroid muscle are responsible for pitch drop in speech. If any of these kinds of EMG activities is correlated with phonological pitch changes, we expect that we can determine the phonological tone of a devoiced vowel by using it. Cf. Sawashima Kakita and Hiki (1973); Hirose et al. (1970).

5. It is widely known that there exist two modes in the perceptual process of speech i.e. the categorical mode and the continuous mode (cf. Eimas (1963)). In the former, as in the case of stop consonants (cf. Liberman et al. (1957)), discriminability of a pair of sounds is enhanced around phoneme boundaries, which is called the "phoneme boundary effect"; while, in the latter mode, as in the case of isolated vowels (cf. Fry et al. (1962)), discriminability is considerably uniform across the whole continuum. There is, however, doubt on absolute dichotomy of perceptual modes. As for the perceptual mode of isolated vowels, which is usually reported to be continuous, Fujisaki et al. (1969; 1971) show that it can be regarded to a certain degree as categorical. Then we would have to say that such-and-such sounds are more categorical than continuous or vice versa.

6. According to McNicol (1972, 56), d', which is "the value of the signal distribution mean, measured in S.D. units of the noise distribution, when the noise distribution mean is equal to zero and both distributions are Gaussian and have S.D. =1," is calculated as follows:

(i) $d' = z(D/S) - z(D/D)$

in which $z(D/S)$ and $z(D/D)$ are the z-scores corresponding to $P(D/S)$, and $P(D/D)$, respectively. $P(D/S)$, which is called a "false alarm," is the probability density of responding DIFFERENT for a stimulus pair SAME, and $P(D/D)$, which is called a "hit," is that of responding DIFFERENT for a stimulus pair DIFF. The sensitivity index d' is equal to zero at chance level and higher when discrimination is more accurate. Similarly $P(S/D)$ $(=1-P(D/D))$ is called a "miss," and $P(S/S)$ $(=1-P(D/S))$ a "correct rejection."

7. The criterion value of the likelihood ratio β as a response bias parameter is calculated according to the following formulas:

(i) $\log_{10}\beta=\log_{10}(y_D/y_S)$

where y_D is the height of the DIFF distribution and y_S is the height of the SAME distribution at the tone pattern boundary. The logarithmic scale value of β to the base 10 was used because it enables us to campare values with ease (proportionally). So, $\log_{10}\beta$ is equal to zero when there exists no response bias, negative when the bias is toward DIFFERENT responses, and positive when the bias is toward SAME responses.

8. Actual F_0 contours for *masIkama* by a human speaker is much more complex. In general, if there are nonsonorants in a word, its F_0 contour tends to be more complex and/or less continuous.

9. Wood (1975) found that irrespective of whether forced-choice paradigms, in which response biases are eliminated experimentally, or signal detection methodology, in which they are eliminated analytically, are used, there is evidence for the phoneme boundary effect between voiced and voiceless stop consonants

10. Under Haraguchi's (1977; 1991) theory, the tone pattern of *atUkuwa* "thick+CONTRAST" and *tatIkawa*, a place name in Tokyo, which actually is LHLL, is incorrectly predicted to be LLHL. All our subjects, however, reject the latter as an illegitimate tone pattern in Tokyo Japanese, and also it seems that the pattern is unattested. cf. McCawley (1977).

References

Eimas, P. D. (1963) "The relation between identification and discrimination along speech and non-speech continua," *Language and Speech* 6. 206-217.

Fry, D. B., A. S. Abramson, P. D. Eimas and A. M. Liberman (1962) "The identification and discrimination of synthetic vowels," *Language and speech* 5. 171-189.

Fujimura, O. (1971) "Hatsuon no katei [The process of pronunciation]," *Onsei Gengo Igaku* 2:1. 11-21.

Fujisaki, H. and T. Kawashima (1969) "On the modes and mechanisms of speech perception," *Annual Report of the Engineering Research Institute* 28. 67-73, Faculty of Engineering, University of Tokyo.

Fujisaki, H. and T. Kawashima (1971) "A quantitative model for the mechanisms of discriminating speech sounds," *Seventh International Congress on Acoustics* 3. 433-436.

Fujisaki, H., Y. Mitsuya and M. Sugito (1974) "Analysis, synthesis, and perception of accent types of two-mora words in Tokyo and Osaka dialects," *Transactions of the Committee of Speech Research*, S 73-51, Acoustic Society of Japan.

Fujisaki, H., H. Morikawa and M. Sugito (1976) "Temporal organization of articulatory and phonatory controls in realization of Japanese word accent," *Annual Bulletin of the Resaerch Institute of Logopedics and Phoniatrics* 10. 177-190, University of Tokyo.

Fujisaki, H. and M. Sugito (1976) "Acoustic and perceptual analysis of two-mora word accent types in the Osaka dialect," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 10. 157-171, University of Tokyo.

Han, M. S. (1962) *Japanese Phonology: An Analysis Based upon Sound Sectrograms*, Tokyo: Kenkyusha.

Haraguchi, S. (1977) *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*, Tokyo: Kaitakusha.

Haraguchi, S. (1984) "Some tonal and Segmental effects of vowel height in Japanese," In M. Aronoff and R. T. Oehrle eds. *Language Sound Structure* 145-156, Cambridge, Mass.: MIT Press.

Haraguchi, S. (1991) *A Theory of Stress and Accent*, Dordrecht, Holland : Foris.

Hattori, S. (1975) "Gengo onsei no mimi niyoru kansatsu to kikai niyoru jikken [Observation of speech by ear and experiments on it by machine]," *Onsei Gengo Igaku* 16, 6-15.

Hirayama, T. (1960) *Zenkoku Akusento Jiten* [All-Japan Accent Dictionary], Tokyo: Tookyoodoo.

Hirose, H., Z. Simada, and O. Fujimura (1970) "An electromyographic study of the activithy of the laryngeal muscles during speech utterances," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 4. 9-25, University of Tokyo.

Kindaichi. K. (1958) *Meikai Nihongo Akusento Jiten* [Clear Japanese Accent Dictionary], Tokyo: Sanseidoo.

Kohno, T. and K. Takahashi (1983) "On the acoustic properties of the intervocalic /h/ in English and Japanese," *Otsuma Joshi Daigaku Bungakubu Kiyoo* 15. 79-84.

Kuhl, P. K. and J. D. Miller (1975) "Speech perception by the chinchilla : voiced-voiceless distinction in alveolar plosive consonants," *Science* 190. 69-72.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith (1957) "The discrimination of speech sounds within and across phoneme boundaries," *Journal of Experimental Psychology* 54. 358-368.

Lovins, J. B. (1976) "Pitch accent and vowel devoicing in Japanese : a preliminary study," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 10.113-125, University of Tokyo.

McCawley, J. D. (1968) *The Phonological Component of a Grammar of Japanese*, The Hague : Mouton.

McCawley, J. D. (1977) "Accent in Japanese," in L. M. Hyman ed. *Studies in Stress and Accent*, *Southern California Occasional Papers in Linguistics* 4. 261-302, University of Southern California.

McNicol, D. (1972) *A Primer of Signal Detection Theory*, London: George Allen and Unwin Ltd.

Miller, J. D., R. W. Pastore, C.C. Wier, W. J. Kelley and R. J. Dooling (1974) "Discrimination and labeling of noise-buzz sequences with varying noise lead times," *Journal of the Acoustic Society of America* 55. 390.

Nihon Hoosoo Kyookai ed. (1966) *Nihongo Hatsuon Akusento Jiten* [Japanese Pronunciation Accent Dictionary], Tokyo: Nihon Hoosoo Shuppan Kyookai.

Sawashima, M. (1969) "Devoiced syllables in Japanese : a preliminary study by photoelectric glottography," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 3. 35-41, University of Tokyo.

Sawashima, M. (1971) "Devoicing of vowels," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 5. 7-13, University of Tokyo.

Sawashima, M., H. Hirose, T. Ushijima, and O. Fujimura (1971) "Devoicing of vowels," *Seventh International Congress on Acoustics* 3. 109-112.

Sawashima, M, Y. Kakita, and S. Hiki (1973) "Activity of the extrinsic laryngeal muscles in relation to Japanese word accent," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 7. 19-25, University of Tokyo.

Simada, Z. and H. Hirose (1971) "Physiological correlates of Japanese accent patterns," *Annual Bulletin of the Research Bulletin of Logopedics and Phoniatrics* 5. 41-49, University of Tokyo.

Simada, Z., H. Hirose, M. Sawashima and O. Fujimura (1971) "An EMG study of Japanese accent patterns," *Seventh International Congress on Acoustics* 3. 121-124.

Sugito, M. (1969) "A study on voiceless vowels in Tokyo and Osaka dialects," *Study of Sounds* 14. 249-264, Phonetic Society of Japan.

Weitzman, R. S, M. Sawashima, H. Hirose and T. Ushijima (1976) "Devoiced and whispered vowels in Japanese," *Annual Bulletin of the Research Bulletin of Logopedics and Phoniatrics* 10. 61-79, University of Tokyo.

Wood, C. C. (1975) "Discriminability, response bias, and phoneme categories in voice onset time discrimination," ms., Walter Reed Army Institute of Resaerch.

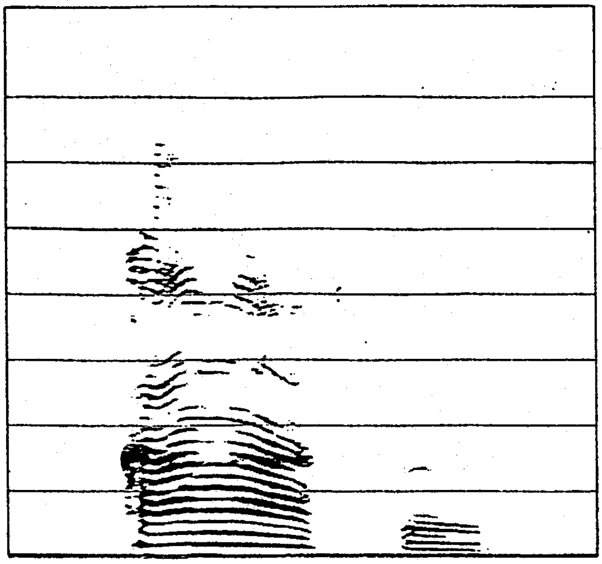

Figure 1. Spectrograph of *kagaku* (HLL) "science."

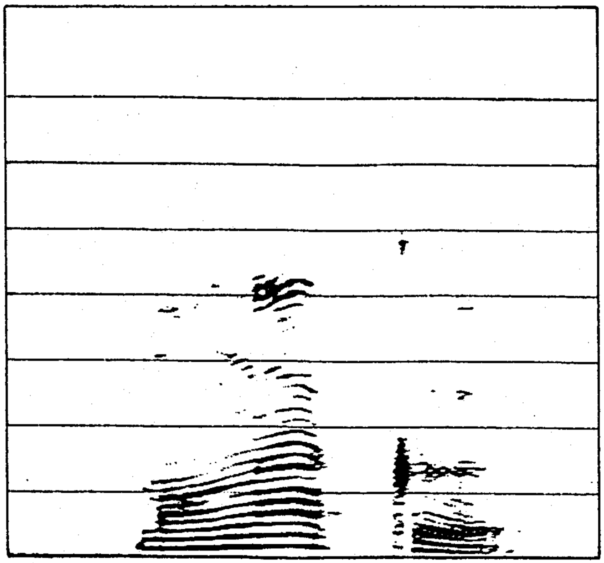

Figure 2. Spectrograph of *mogaku* (LHL) "to struggle."

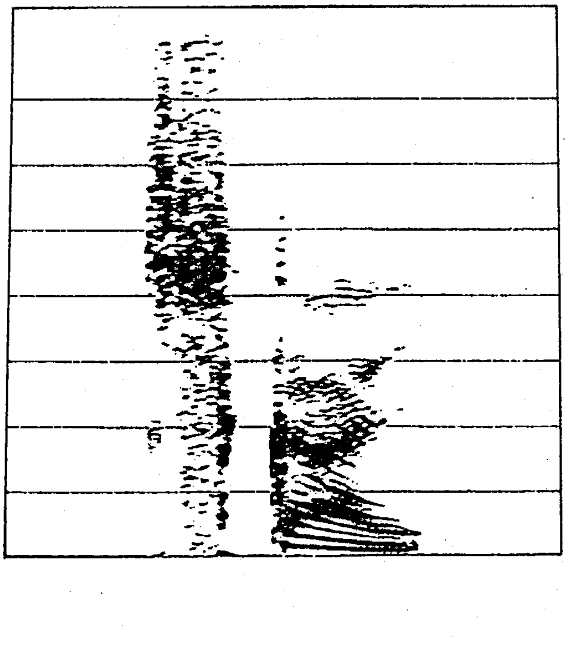

Figure 3. Spectrograph of *sIkai* (HLL) "The Dead Sea."

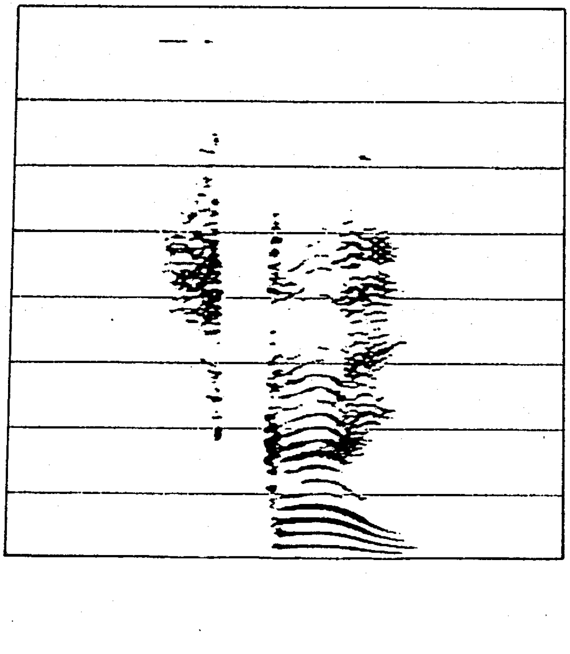

Figure 4. Spectrograph of *sIkai* (LHL) "dentist."

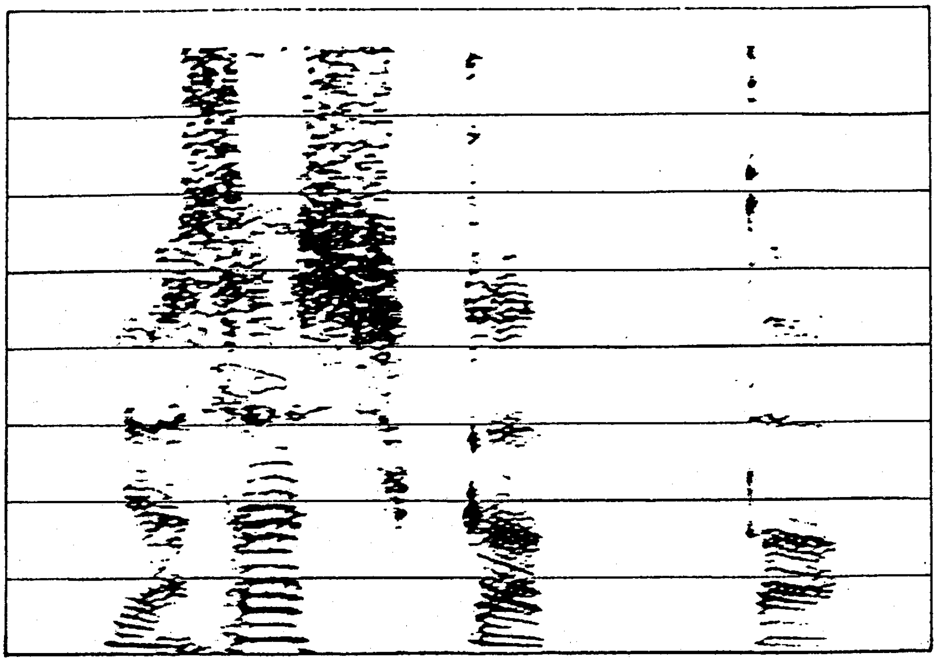

Figure 5. Spectrograph of *yasasIkatta* (LHLLLL) "gentle + PAST."

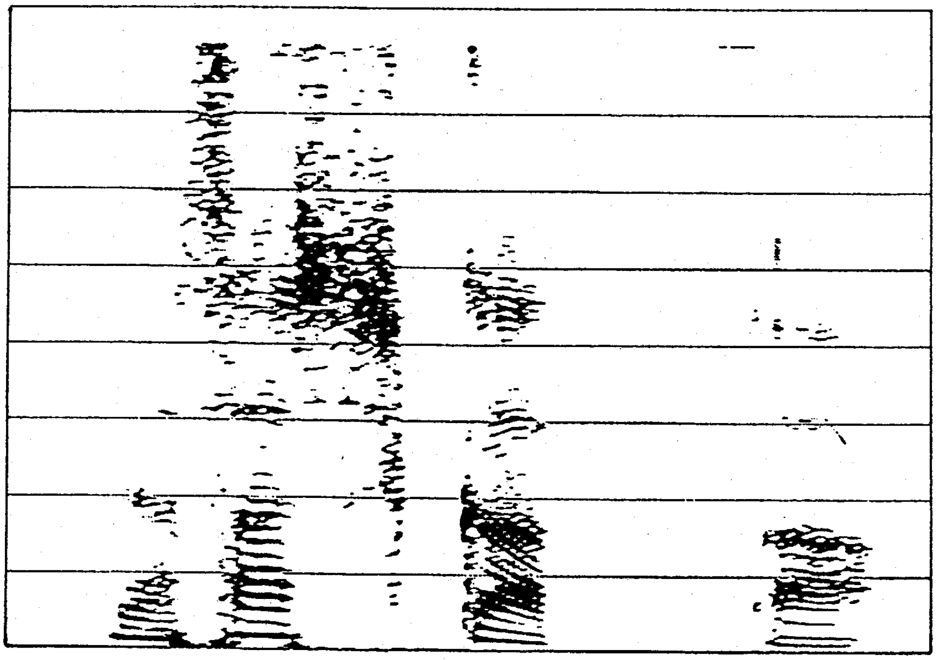

Figure 6. Spectrograph of *yasasIkatta* (LHHLLL) "gentle + PAST."

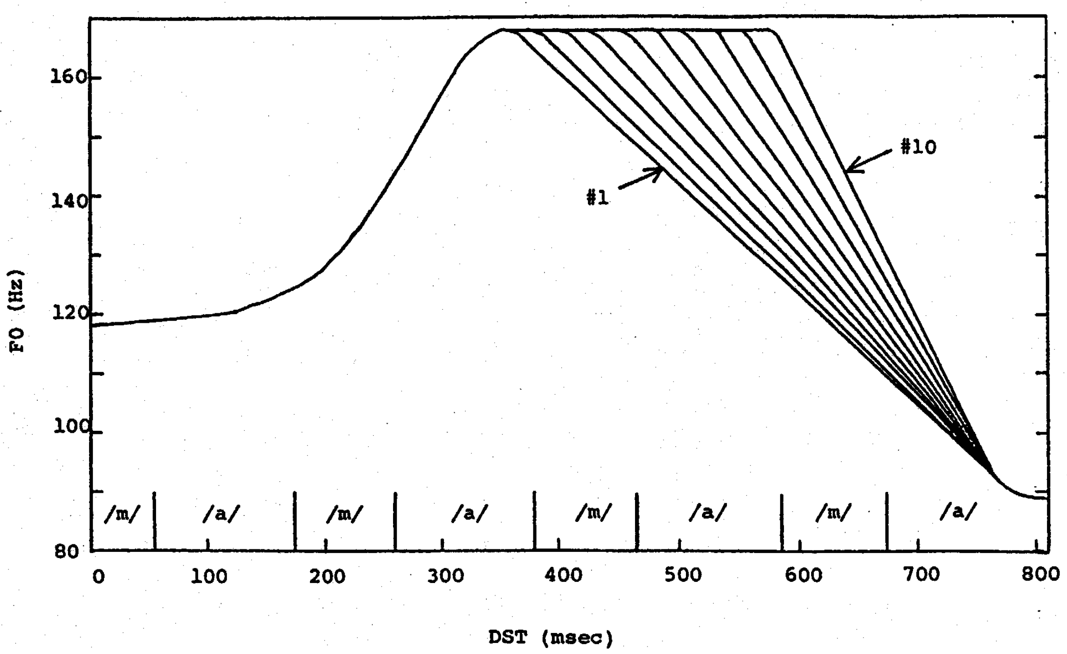

Figure 7. Rough F_0 contours of the stimuli in Experiment I.

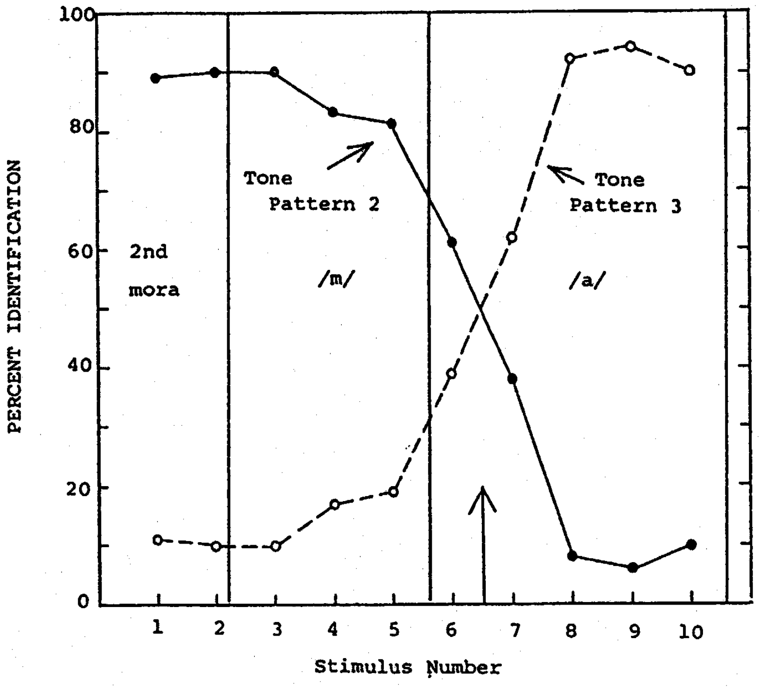

Figure 8. Mean identification functions for DST in Experiment I. Each plotted value is based on 80 judgments.

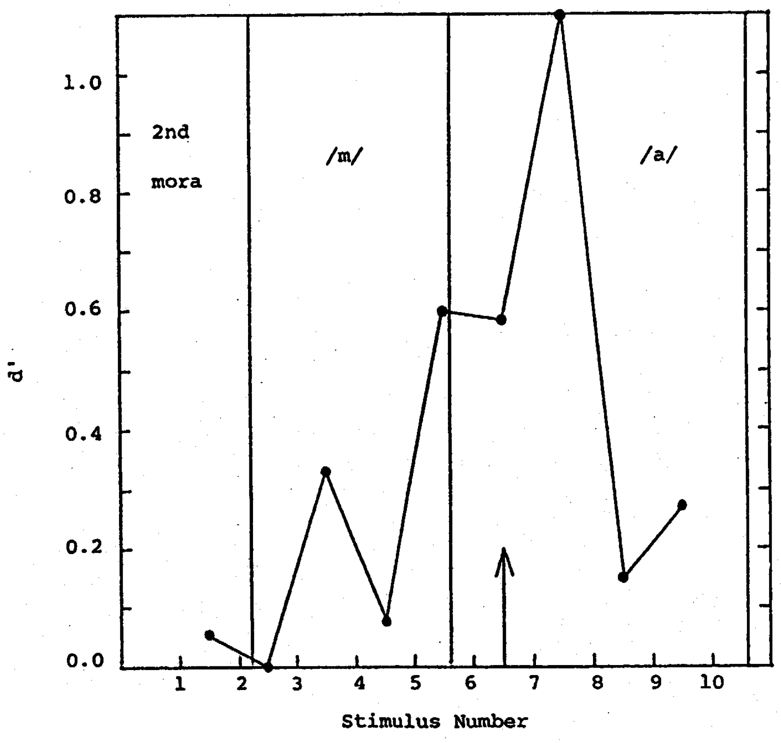

Figure 9. Mean discriminability (*d'*) as a function of DST in the identification task of Experiment I. Each value is based on 180 judgments.

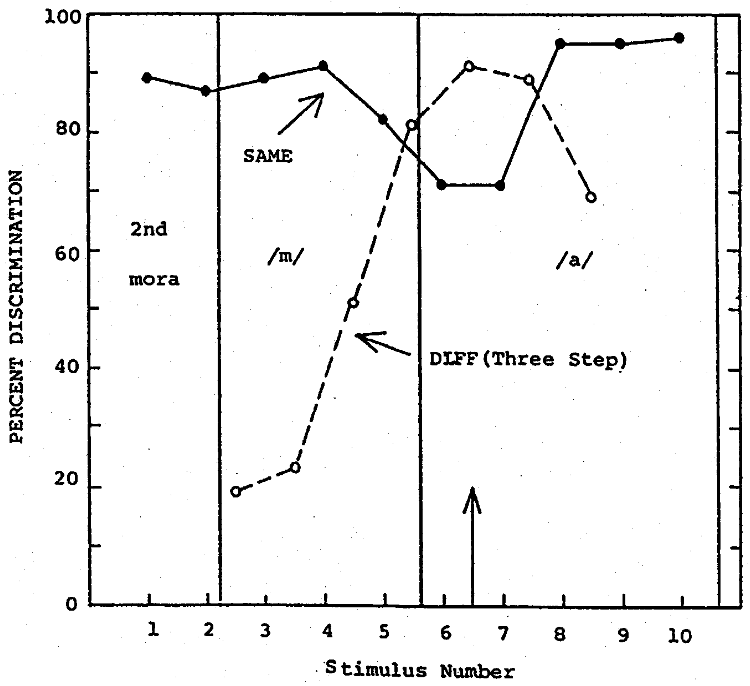

Figure 10. Mean discrimination functions for SAMEs and DIFFs in Experiment I. Each value is based on 56 judgments for SAMEs and 80 judgments for DIFFs.

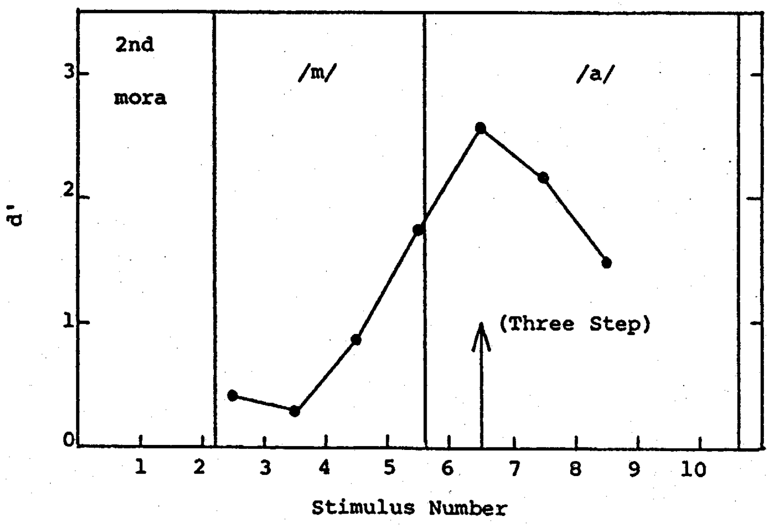

Figure 11. Mean discriminability (*d'*) as a function of DST in the discrimination task of Experiment I. Each value is based on 80 SAMEs and 112 DIFFs.

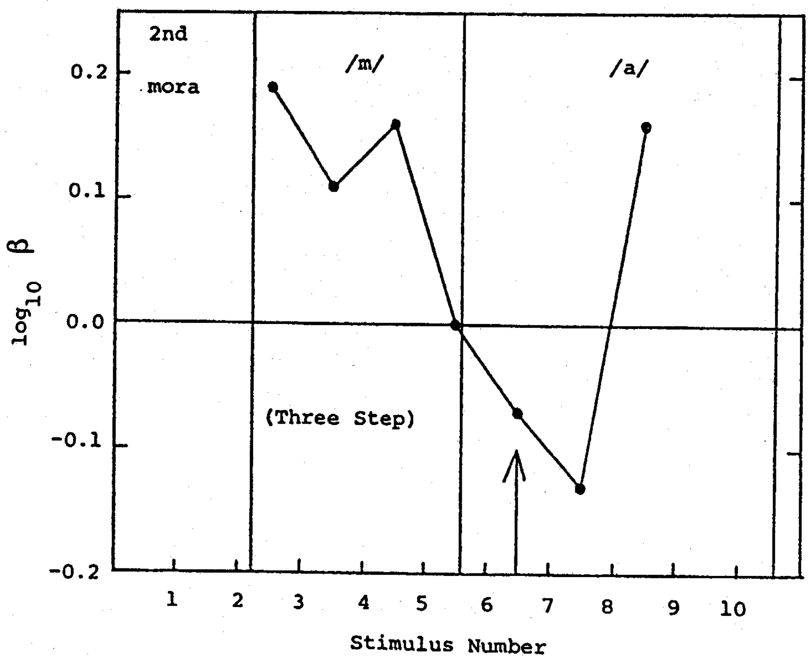

Figure 12. Mean response bias ($\log_{10}\beta$) as a function of DST in Experiment I. Each value is based on 80 SAMEs and 112 DIFFs.

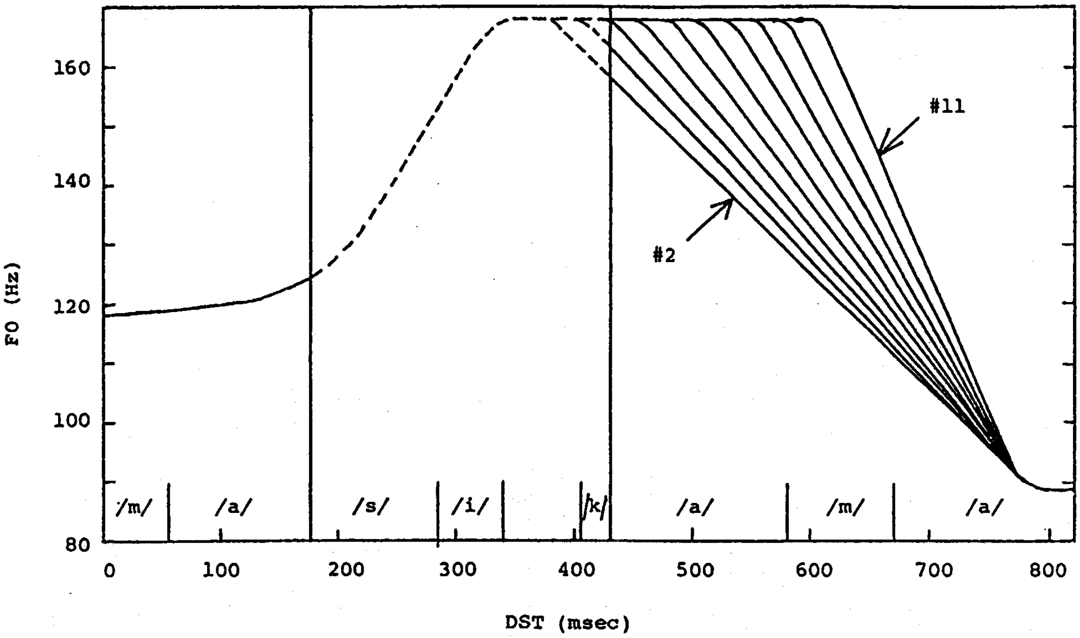

Figure 13. Rough F_0 contours of the stimuli in Experiment II.

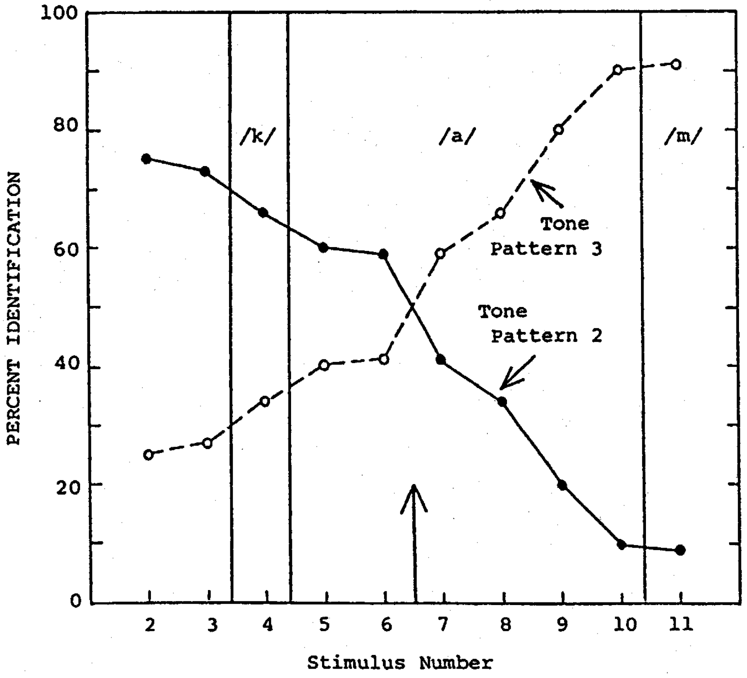

Figure 14. Mean identification functions for DST of the entire subjects in Experiment II. Each value is based on 80 judgments.

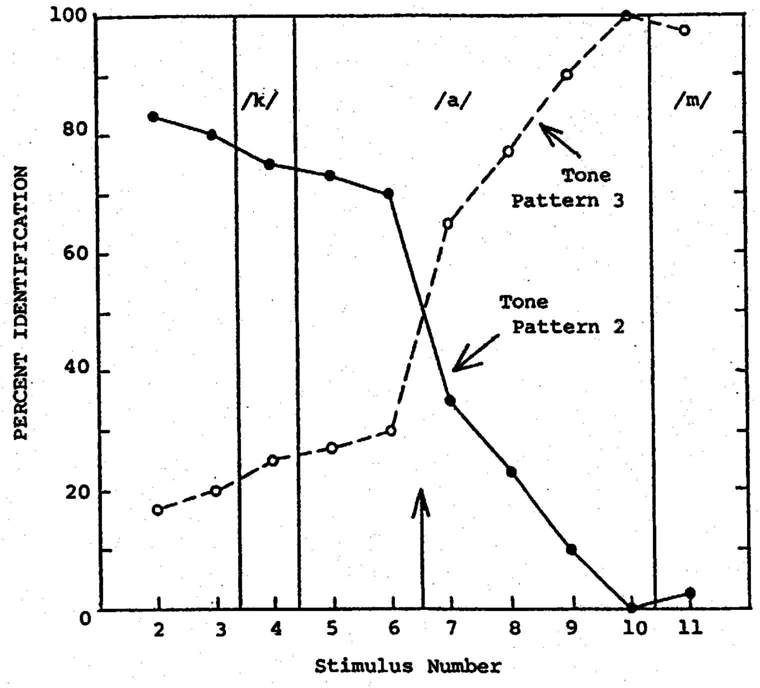

Figure 15. Mean identification functions for DST of the categorical subjects in Experiment II. Each value is based on 40 judgments.

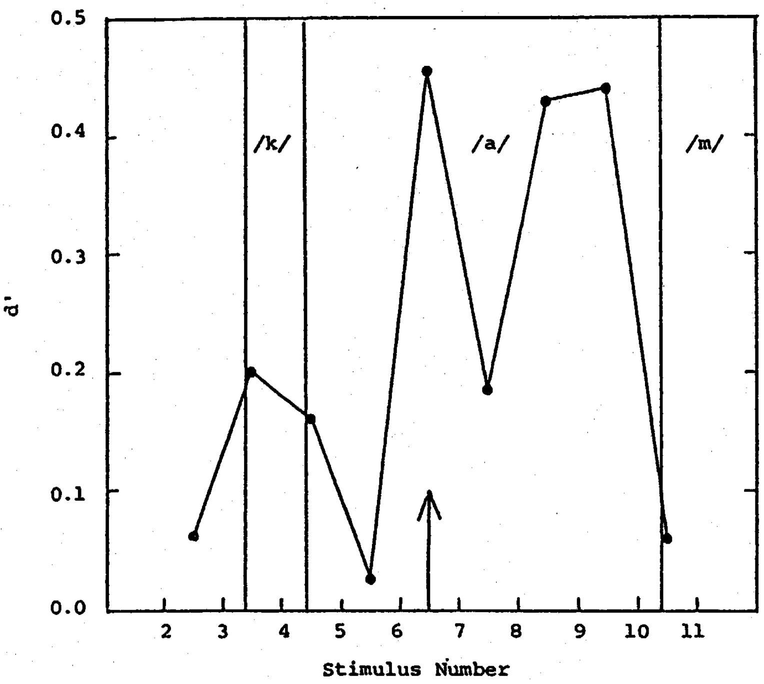

Figure 16. Mean discriminability (d') as a function of DST in the identification task of the entire subjects in Experiment II. Each value is based on 180 judgments.

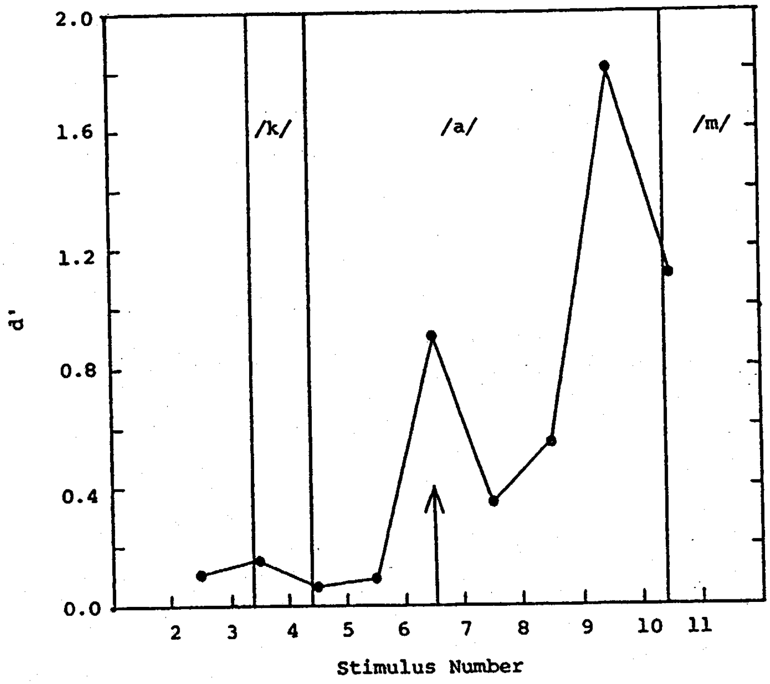

Figure 17. Mean discriminability (d') as a function of DST in the identification task of the categorical subjects in Experiment II. Each value is based on 90 judgments.

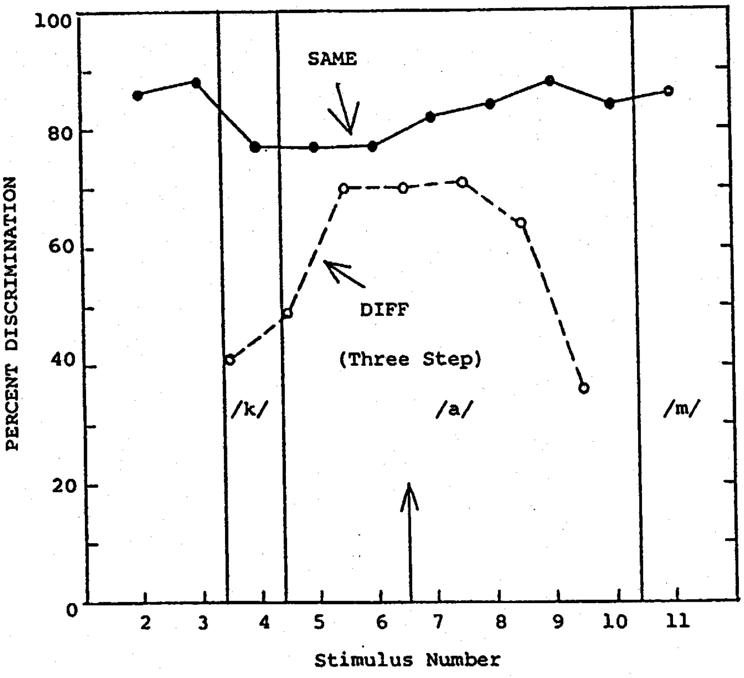

Figure 18. Mean discrimination functions for SAMEs and DIFFs of the entire subjects in Experiment II. Each value is based on 56 judgments for SAMEs and 80 judgments for DIFFs.

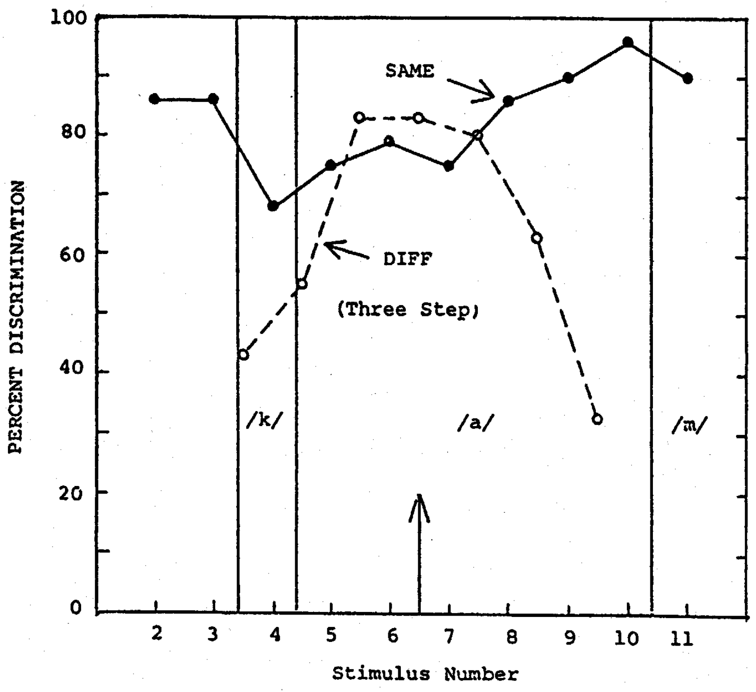

Figure 19. Mean discrimination functions for SAMEs and DIFFs of the categorical subjects in Experiment II. Each value is based on 28 judgments for SAMEs and 40 judgments for DIFFs.

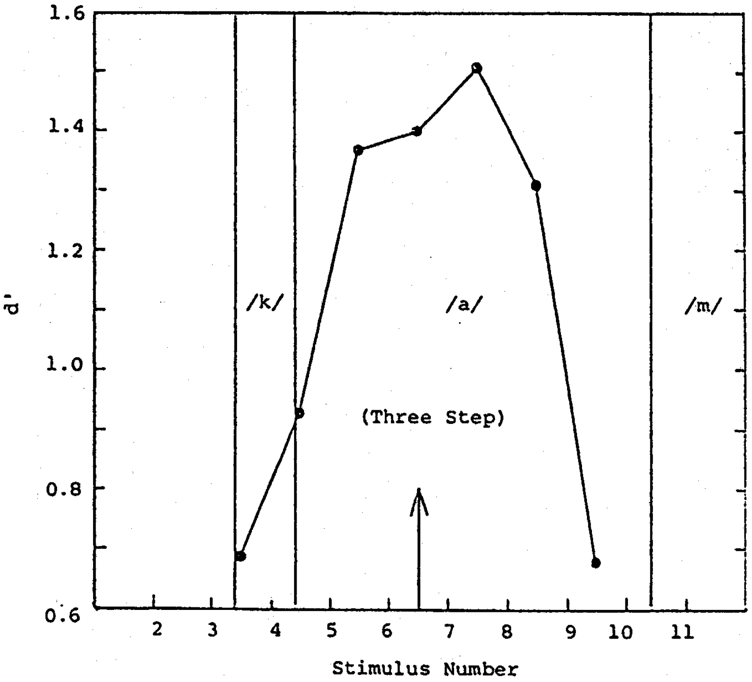

Figure 20. Mean discriminability (*d'*) as a function of DST in the discrimination task of the entire subjects in Experiment II. Each value is based on 80 SAMEs and 112 DIFFs.

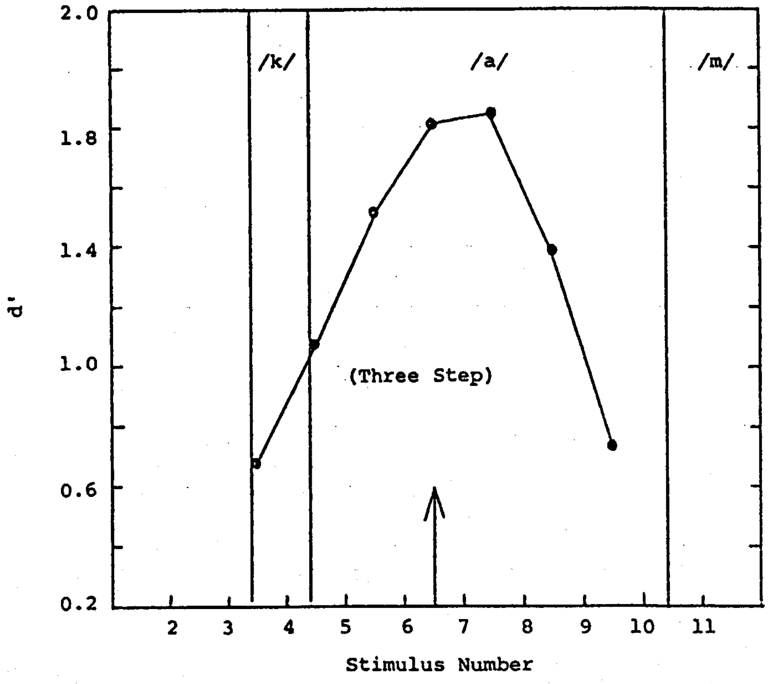

Figure 21. Mean discriminability (d') as a function of DST in the discrimination task of the categorical subjects in Experiment II. Each value is based on 40 SAMEs and 56 DIFFs.

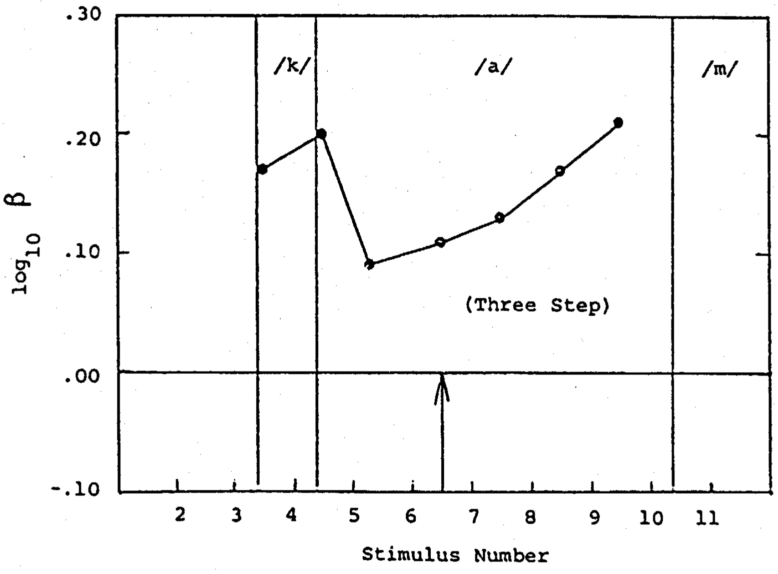

Figure 22. Mean response bias ($\log_{10} \beta$) as a function of DST of the entire subjects in Experiment II. Each value is based on 80 SAMEs and 112 DIFFs.

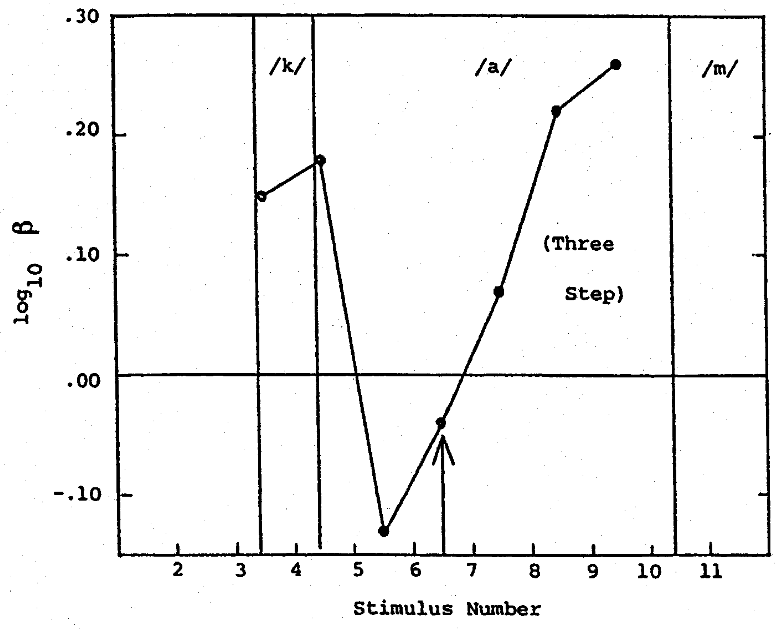

Figure 23. Mean response bias ($\log_{10} \beta$) as a function of DST of the categorical subjects in Experiment II. Each value is based on 40 SAMEs and 56 DIFFs.