Ann.Bull.RILP(1985)No.19, 85-93

THE AUDITORY TAU EFFECT FOR SPEECH SOUNDS

Sumi Shigeno*

ABSTRACT

The tau effect, an effect of time interval on perception of spatial separation or sensory difference, has been observed in various sense modalities including pitch bisection judgments in audition. However, systematic studies on the mechanisms of the auditory tau effect have been rather scarce. In the present study the perception of vowel quality was investigated, using the AXB method. The results indicate that the auditory tau effect depends upon the difference between phonetic and non-phonetic judgments and that the reverse trend of the auditory tau effect can be found in phonetic judgments of between-category speech sounds. These findings suggest that the interaction between time interval and vowel quality perception can be explained mainly by the integration of the proactive and retroactive context effects and that these effects occur at an early stage of auditory information processing.

INTRODUCTION

Many experiments have been conducted on time-space relations. For example, the tau effect is usually cited as evidence for the dependence of space perception on time. The tau effect was first identified by Helson (1930) in tactile perception. He found that, when two points are stimulated successively, the perceived distance between the points varies as a function of the temporal interval separating the two points. More specifically, these findings demonstrate that longer (shorter) temporal intervals result in judgments of greater (lesser) perceived spatial separation in relation to a standard when the two distances are physically equal. A few years later, Geldreigh (1934) demonstrated an analogous effect in visual perception. Furthermore, Bill & Teft (1969), using a method suggested by Geldreigh (1934), investigated the effects of variation in the temporal interval separating two visual stimuli on the perceived extent demarcated by these stimuli and confirmed the findings of Geldreigh (1934).

Meanwhile, Cohen, Hansel, & Sylvester (1954) found the tau effect in bisection judgments of pitch (the auditory tau effect). If a subject adjusts a tone so that it appears for him to be intermediate in pitch between two other tones, he makes tones which are presented closer together in time further apart in frequency. Cohen, Christensen, & Ono (1974) found that this auditory tau effect occurs monaurally as well as binaurally, and

*Kitasato University, Kitasato, Sagamihara-shi

they suggested a hierarchical order of perceptual processing into which temporal and/or spatial factors enter as an integral component. Christensen & Huang (1979) replicated this research with a more precise control of the time condition, and reported this findings as being in perfect accordance with those reported by Cohen et al. (1974). They offered an explanation for the auditory *tau* effect in terms of the deterioration and interference of memory for pitch.

It has been reported that a phenomenon similar to the auditory *tau* effect can be obtained in two-tone (or sound) context effects, where assimilation--a shift of category boundary in the opposite direction to the context stimulus/stimuli--is produced in the case of non-phonetic judgments, or contrast--a shift of category boundary toward the context stimulus/stimuli--is produced in the case of phonetic judgments (Shigeno & Fujisaki, 1979; 1980). It is, therefore, necessary to investigate the influence of the categorization of a stimulus upon the auditory *tau* effect. Since the essential factors causing the auditory *tau* effect can be considered to be associated with distinct short-term storage systems, two different sets of speech stimuli within-category and between-category, whose perceptual processes are different from each other, were used to clarify these relations.

METHOD

Subjects

The subjects were 19 undergraduate students of Kitasato University. None of them reported any history of hearing disorders. All were native speakers of Japanese.

Stimuli

Synthetic vowels were prepared using a PDP 11 computer (DEC). They were generated by a terminal-analog speech synthesizer and were normalized by the overall maximum amplitude. All the stimuli were read out at a sampling rate of 10 kHz with an accuracy of 10 bits; converted into the analog waveform; and recorded through a low-pass filter with a cutoff frequency of 4.5 kHz on a magnetic tape for the off-line experiment. All the stimuli had a steady duration of 200 msec, with rise and decay times of 20 msec.

Three kinds of vowels, /e/, /i/, and /I/, were employed. In Japanese, /e/ and /i/ are perceived to be typical vowels. On the other hand, /I/ sound is perceived to be ambiguous and different from both /e/ and /i/ in vowel quality, though /I/ is usually categorized into the same category with /i/. The first formant frequencies of /e/, /i/, and /I/ were 520 Hz, 270 Hz, and 370 Hz, respectively. The other stimulus parameters of these vowels are shown in Table 1.

Table 1 Stimulus parameters used in Experiment 2.

Stimulus Number		Formant frequencies (Hz) F_1	F_2	F_3
		/i/-/e/	/e/-/i/	
/i/	1	300	2500	2800
	2	350	2360	2760
	3	370	2310	2740
	4	390	2260	2720
	5	410	2210	2700
	6	430	2160	2680
	7	450	2110	2660
	8	470	2060	2640
/e/	9	520	1920	2600
		/i/-/I/	/I/-/i/	
/i/	10	270	2140	2780
	11	300	2140	2780
	12	310	2140	2780
	13	320	2140	2780
	14	330	2140	2780
	15	340	2140	2780
	16	350	2140	2780
	17	360	2140	2780
/I/	18	390	2140	2780

One stimulus set consisted of three sounds. When the first stimulus was /e/ and the third stimulus was /i/, one of seven stimuli on the /e/-/i/ stimulus continuum was selected as the second stimulus so that the first formant frequencies (F_1) of all seven formed a continuum between 470 Hz and 350 Hz with a step size of 20 Hz. When the first stimulus was /i/ and the third stimulus was /e/, on the other hand, the F_1s of the second stimulus formed a continuum between 350 Hz and 470 Hz with a step size of 20 Hz in F_1. The second and the third formant frequencies (F_2, F_3) of these stimuli were varied linearly on the /e/-/i/ or /i/-/e/ continuum to give stronger identity to these vowels. In the cases of the /I/-/i/ and /i/-/I/ continua , the F_1s of the second stimulus varied from 360 Hz to 300 Hz and from 300 Hz to 360 Hz, respectively. The step size was 10 Hz.

The fundamental, fourth, and fifth formant frequencies (F_0, F_4, F_5) of these vowels were held constant at 140 Hz, 3500 Hz, and 4500 Hz, respectively. The bandwidths of F_1, F_2, F_3, F_4, and F_5 were also held constant at 60 Hz, 100 Hz, 120 Hz, 175 Hz, and 280 Hz, respectively.

Procedure

The subjects were divided into four groups according to the four conditions of the stimulus continuum for /e/-/i/, /i/-/e/, /I/-/i/ and /i/-/I/. Each group consisted of six subjects. Five of the subjects participated in two conditions.

The AXB method was employed. The subjects were required to judge whether X was similar to A or B. The temporal positions of the target sound (X), represented as the time ratio $t_1/(t_1 + t_2)$, were varied under the same five conditions as in Experiment 1. The total duration of $(t_1 + t_2)$ remained constant at 1.0 sec.

The subjects were twice given the same series of 150 trials. The initial five and the last five trials of each series were excluded from the data. The total number of trials was 300 for each subject [{(5 temporal positions of the target sound) x (7 target stimuli) x 4 + 10 extra trials} x 2]. Eight judgments per each target sound were obtained from each subject for the five temporal conditions.

The recorded stimuli were played back on a tape recorder (National, RU-2006u) and were presented to the subjects in a quiet room through a loudspeaker (National, WS320) at a comfortable listening level.

RESULTS AND DISCUSSION

The probability that the target stimulus would be judged as similar to either one of two stimuli as function of the F_1 can be generally approximated by a cumulative normal distribution. Therefore, the responses that the target stimulus was similar to the third stimulus for five temporal conditions for each subject were approximated by a cumulative normal distribution whose mean and standard deviation were obtained by fitting a normal ogive to each subject's data with the normal graphic process outlined by Guilford (1954). The mean indicates the subjective middle F_1 between the first and the third sounds and theoretically corresponds to the mean adjusted F_1 that was obtained in Christensen et al. (1979) by the method of adjustment, while the standard deviation serves as an index for the accuracy of the judgments.

The means and standard deviations obtained by a normal graphic process for each subject are given in Table 2, where the mean represents the phonetic boundary between /e/ and /i/ or between /I/ and /i/. The differences in these phonetic boundaries due to the time ratio $t_1/(t_1 + t_2)$ were found to be significant by a t-test between the time ratio .05 and .95 (for the /e/-/i/ continuum $t(5) = 4.89$, $p<.01$; for the /i/-/e/ continuum $t(5) = 2.69$, $p<.05$). Likewise, the differences in the phonetic boundaries of the /I/-/i/ or the /i/-/I/ continuum between the time ratios .05 and .95 was found to be significant by a t-test (for the /I/-/i/ continuum $t(5) = 7.133$, $p<.01$; for the /i/-/I/ continuum $t(5) = 3.85$, $p<.05$).

The means obtained from the pooled performances of all the subjects correspond to the subjective category boundary and are plotted as a function of the temporal position of the target sound in Fig. 1. The upper graphs show the results for the /e/-/i/ and /i/-/e/ stimulus continua, while the lower graphs show the results for the /I/-/i/ and /i/-/I/ stimulus continua.

In the case of the /e/-/i/ continuum, the percentage of times that the target sound was judged similar to the first sound (/e/) decreased as the target stimulus was put temporally closer to the first stimulus, while the percentage that the target stimulus was judged similar to the third stimulus (/i/) decreased as the target stimulus was put temporally closer to the third stimulus. Likewise, in the case of the /i/-/e/ continuum, the percentage that the target stimulus was judged similar to the first stimulus (/i/) decreased as the target stimulus was put temporally closer to the first stimulus, while the percentage that the target stimulus was judged similar to the third stimulus (/e/) decreased as the target stimulus came temporally closer to the third stimulus. These results reflect the fact that the phonetic boundary between /e/ and /i/ moved toward the temporally closer sound, and, thus, the subject tended to hear the target stimulus as more similar to the farther sound. The two upper graphs of Fig. 1 show these downward and upward linear trends, respectively.

In the case of the /I/-/i/ continuum, on the other hand, the percentage that the target stimulus was perceived as similar to the first sound (/I/) got greater as the target stimulus came temporally closer to the first stimulus, while the percentage that the target stimulus was perceived as similar to the third

Table 2 The subjective middle frequency (or F_1) and the standard deviations for each subject.

Subject Number	Value of $t_1/(t_1+t_2)$ .05 Mean	SD	.25 Mean	SD	.50 Mean	SD	.75 Mean	SD	.95 Mean	SD
					/e/-/i/					
7	430	4	392	24	398	8	398	7	386	16
8	418	10	395	31	384	20	372	12	364	14
9	409	15	390	24	381	17	369	4	351	22
10	390	18	369	15	371	5	370	6	360	12
11	421	6	391	15	382	19	360	7	363	19
12	478	12	382	6	371	22	390	11	360	9
$\bar{X}$	424		387		381		377		364	
					/i/-/e/					
9	429	4	426	26	435	28	439	33	454	30
13	437	8	405	6	404	25	409	14	429	4
14	409	4	410	4	420	13	412	17	418	12
15	423	16	401	6	420	6	450	23	436	14
16	403	33	410	31	420	56	430	54	420	58
17	399	32	402	19	413	54	409	4	420	6
$\bar{X}$	415		409		419		425		430	
					/I/-/i/					
7	314	17	328	9	340	10	332	11	360	2
11	311	9	311	9	318	7	319	2	326	6
18	305	6	304	6	308	11	326	20	340	15
19	292	9	293	10	301	17	323	17	339	13
20	294	9	293	18	303	18	314	10	335	16
21	303	22	314	21	326	14	335	28	349	10
$\bar{X}$	303		307		316		325		342	
					/i/-/I/					
10	335	17	341	11	346	16	340	12	330	14
22	351	19	348	16	346	15	328	12	324	14
23	346	11	351	12	328	11	338	7	332	24
24	353	15	355	12	344	9	337	12	331	14
25	349	14	337	6	327	5	333	7	327	10
26	335	8	339	4	343	13	337	17	332	19
$\bar{X}$	345		345		341		336		329	

stimulus (/i/) got greater as the target stimulus came temporally closer to the third stimulus. Likewise, in the case of the /i/-/I/ continuum, the percentage that the target stimulus was judged similar to the first stimulus (/i/) got greater as the target stimulus came temporally closer to the first stimulus, while the percentage that the target stimulus was judged similar to the third stimulus (/I/) got greater as the target stimulus came temporally closer to the third stimulus. These results reflect the fact that the perceived category boundary between /I/ and /i/ moved toward the farther stimulus, and, thus, the subject tended to hear the target stimulus as more similar to the closer stimulus. The two lower graphs of Fig. 1 shows these upward and downward linear trends, respectively.

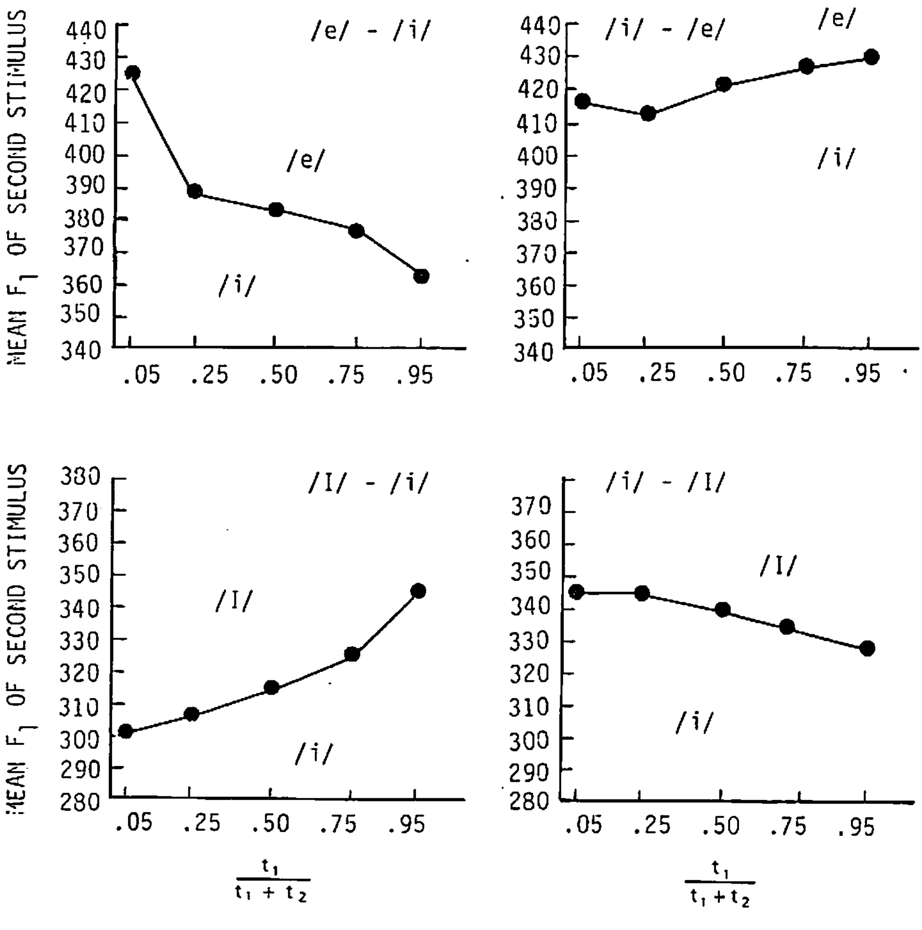

Fig. 1 The subjective middle F_1 of the second sound as a function of its temporal position.

It is thus possible to say that in speech perception the auditory tau effect doesn't occur when the stimuli are between different phonetic categories such as /e/ and /i/, though it occurs as in pure tones when the stimuli are within the same category as are /I/ and /i/. Furthermore, in the case of the /I/-/i/ or the /i/-/I/ continuum, the shift of the boundary toward the farther stimulus can be interpreted as an assimilative shift. In the case of the /e/-/i/ or the /i/-/e/ continuum, on the other hand, the shift of the boundary toward the closer stimulus can be interpreted as a contrastive phenomenon. These result suggest that the stimulus attributes (speech or non-speech) are not an essential factor causing the auditory tau effect, but that whether the stimuli are between different categories or within the same category is the important factor. Both the pairs of pure tones used in Experiment 1 and /I/ and /i/ used in this experiment belong to the same categories and produced an auditory tau effect, or assimilative shifts.

As stated above, in Japanese /I/ and /i/ are identified as belonging to the same category. If one of the major differences between the /e/-/i/ (or the /i/-/e/) continuum and the /I/-/i/ (or the /i/-/I/) continuum is assumed to be based on the difference in the amount of available auditory memory, then, some form of auditory memory may take a part in the auditory tau effect. Several auditory information processing models which hypothesize two kinds of short-term memory including auditory short-term memory, can be distinguished from each other according to their magnitude of influence upon the determination of the category boundary (e.g. Shigeno and Fujisaki, 1979).

The results may, then, bear on an interpretation that the auditory tau effect can be based on the same process that underlies the so-called context effect, and that the auditory tau effect is a phenomenon that occurs at a relatively primary stage before categorization.

REFERENCES

Bell, J. C. & Teft, L. W. (1969); Space-time relations: effects of time on perceived visual extent. Journal of Experimental Psychology, 81, 196-199.

Christensen, I. P. & Huang, Y. L. (1979); The auditory tau effect and memory for pitch. Perception & Psychophysics, 26, 489-494.

Cohen, J., Christensen, I. P. & Ono, A. (1974); Influence of temporal intervals on comparative judgments of pitch: a study of subjective reality. Tohoku Psychological Folia, 33, 76-87.

Cohen, J., Hansel, C. E. M. & Sylvester, J. D. (1954); Interdependence of temporal auditory judgments. Nature, 174, 642.

Geldreigh, E. W. (1934); A lecture-room demonstration of the visual tau-effect. American Journal of Psychology, 46, 483-485.

Guilford, J. P. (1954); Psychometric methods. New York: McGraw-Hill.
Helson, H. (1930); The tau effect--an example of psychological relativity. Science, 71, 536-537.
Shigeno, S. & Fujisaki, H. (1979); Effect of a preceding anchor upon the categorical judgment of speech and non-speech stimuli. The Japanese Psychological Research, 21, 165-173.
Shigeno, S. & Fujisaki, H. (1980); A comparison of forward versus backward context effect in phonetic and non-phonetic judgments on synthetic vowels. The Japanese Journal of Psychology, 51, 287-290. (In Japanese with English summary)