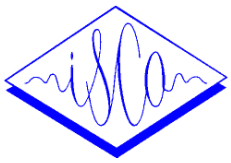

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

VOWEL-CONTINGENT ANCHORING EFFECTS ON THE PERCEPTION OF STOP CONSONANTS

Sumi Shigeno

Kitasato University
Sagamihara-shi, Kanagawa, 228 Japan

ABSTRACT

The vowel-contingent effect in the anchoring paradigm was examined. Two experiments were conducted to investigate (1) the vowel-contingent effect in the anchoring effect upon the perception of voiced stop consonants and (2) the duration of vowel-contingent anchoring effects. Anchors were presented to fifteen subjects as a part of the stimulus series in Experiment 1 and to five subjects as a preceding stimulus to one of the stimulus series in Experiment 2. Two conditions of interstimulus interval between anchor and target were employed: 1.0 s and 3.0 s. The vowel-contingent anchoring effect was obtained in both of the presentation conditions. More effetcs were obtained when the acoustical similarity between the anchor and stimulus series increased. It was also found that the vowel-contingent anchoring effect decreased when the ISI was 3.0 s. The present results suggest that vowel-contingent anchoring effects occur at the auditory level irrespective of their phonetic code.

I. INTRODUCTION

It has been widely recognized that the perception of consonants in a consonant-vowel (CV) syllable is influenced by its context in selective adaptation[1], anchoring[2], and the auditory tau effect paradigms[3]. These context effects were also found to be influenced by the vowel environment. Cooper (1974), by carrying out selective adaptation experiments, indicated that the effect of voicing adapatation was contingent on vowel quality and called it contingent adapation effects[4]. According to his method, subjects listened to two alternating adaptors, which were the voiced syllable [da] and voiceless syllable [ti], and were then tested on both a [bi]-[pi] series and a [ba]-[pa] series. The [bi]-[pi] series showed a reliable category boundary shift toward voiceless [pi] (because the voiceless adaptor [ti] and the stimulus series shared the vowel [i] in common), while the [ba]-[pa] series showed a shift in the opposite direction, toward voiced [ba] (because of the voiced adaptor [da]). These results were interpreted as reflecting a level of analysis of the voicing feature which was vowel dependent.

The results of vowel-contingent selective adaptation raised the question whether adaptation occurs entirely at an auditory level of processing or at both an auditory and at more abstract phonetic levels. Regarding the perception of place of articulation, when the adapting syllable was a member of the test series, there was a significant adaptation effect, although when the adapting syllable was not from the test series, no adaptation effect was found [5]. These results led to the one-level model where spectral commonality or overlap between the adapting and test syllables is required for selective adaptation to occur, and seems to indicate that feature detection takes place at an auditory, rather than phonetic, level of processing.

Phoneme boundary shifts on the place dimension, on the other hand, were obtained even though the adaptor and the stimulus series were signaled by different cues to place information such as that about bursts and formant transitions. These results, suggest that the adaptors might affect the phonetic level as well as the auditory level[6]. It has also been reported that when the spectral overlap between the adapting and test syllables do not share any frequency components in common, the selective adaptation effect occurs. Sawusch (1977), using two sets of synthetic [bae] and [dae] syllables, showed that when the formant frequencies for one set (termed the "high" series) were scaled upward logarithmically from the formant frequencies of the other ("low") series, both high- and low-syllable adaptors were found to cause a shift in the low-series category boundary, and the low adaptors caused about twice as much adaptation as their high-series adaptors. These results seem to provide support for a two-level interpretation[7], although it is further hypothesized that processing at the second level may still be partly frequency or vowel dependent[8].

Since the vowel-contingent selective adaptation has been, so far, examined in the cases in which the adaptor and test series are acoustically same or different, typical vowels have been used in CV syllables. If a vowel-contingent effect takes place at a spectral-specific auditory level of processing, when there is more acoustical similarity with vowels between the

adapting and the test stimuli, there will be more adaptation effects, irrespective of the phonetic identity of the vowels.

As the two paradigms--selective adaptation and anchoring--were found to be equivalent in principle[2], the vowel-contingent effect might be obtained in an anchoring paradigm. Furthermore, it can be said that the acoustical similarity between anchor and test syllables might determine the amount of vowel-contingent effect. When there is more acoustical similarity between the anchor and stimulus series, there will be more anchoring effects.

In the present study, anchors were presented in two presentation methods--one is the method where anchor is presented as a member of a stimulus series (in Experiment 1) and the other is where anchor is paired with each member of a stimulus series (in Experiment 2). In the first experiment, whether the amount of contingent effects were defined by the acoustical similarity between anchors and stimulus series were also investigated. In the second experiment the duration of vowel-contingent effect and some temporal charateristics of feature extraction mechanisms in the paired presentation were examined.

II. EXPERIMENT 1

2.1 Method

2.1.1 Stimuli The stimuli in this experiment were a series of synthetic consonant-vowel (CV) syllables ranging perceptually from [be] to [de]. All the stimuli were prepared using VAX850 computer (DEC). They were read out at a sampling rate of 10 kHz with an accuracy of 12 bits. The series consisted of eleven CV syllables for identification test and of nine syllables (from Stimulus 2 to Stimulus 10) for identification with the anchor test. They varied in the direction and extent of F2 and F3 transition. Taking the results from a preliminary identification test into account, the equal steps of F2 and F3 transition onset frequencies were determined. The first formant frequency for the first 80-110 ms of each syllable varied from 200 Hz to 490 Hz. The formant transitions began at 90-95 ms in the case of F2 and F3. The stimulus values, including transition durations, are displayed in Table 1. Each syllable was 300 ms in duration.

Table 1. Onset frequencies and transition durations for the two formants of the [be]-[de] continuum in Experiment 1. All syllables were 300 ms in duration.

Stimulus number	F2 Onset frequency in Hz	F2 Transition duration in ms	F3 Onset frequency in Hz	F3 Transition duration in ms
1 [be]	1200	20	2400	20
2	1270	20	2440	20
3	1340	25	2480	25
4	1410	30	2520	30
5	1480	35	2560	30
6	1550	35	2600	30
7	1620	35	2640	35
8	1690	40	2680	35
9	1760	45	2720	40
10	1830	50	2760	45
11 [de]	1900	50	2800	45

Four anchors were used. They were a typical [be] sound in Japanese, an ambiguous sound [be]' (intermediate but a little nearer to [be] sound on the [be]-[bi] continuum), an ambiguous sound [bi]' (intermediate but a little nearer to [bi] sound), and a typical [bi] sound in Japanese. In the case of [bi]-anchor, the steady-state Japanese vowel [i] which followed the transitive portion, was 250 Hz for F1, 2300 Hz for F2, and 3100 Hz for F3. The first formant frequency for the first 80-110 ms varied from 200 Hz to 250 Hz. The second and third formant frequencies for the first 90-130 ms varied from 1200 Hz to 2300 Hz and from 2400 Hz to 3100 Hz, respectively. Fundamental frequency of all the stimuli was fixed at 130 Hz.

Two kinds of identification tape were prepared. One was for an identification test of series stimuli. The eleven syllables were recorded at equal intervals. The other was for an identification test of series stimuli with the anchor. The anchoring stimulus was recorded six times as often as the other nine series stimuli; ten random permutations of fifteen stimuli and ten extra stimuli (the initial and the last five stimuli were excluded from the data, considering the serial position effect) yielded a total of 160 stimuli.

2.1.2 Procedure One identification tape and four identification with anchor tapes were given to the subjects twice, separately. Then, twenty judgements per each stimulus was obtained from each subject. The interval for identification was 3 s. Figure 1 shows the method of presen-

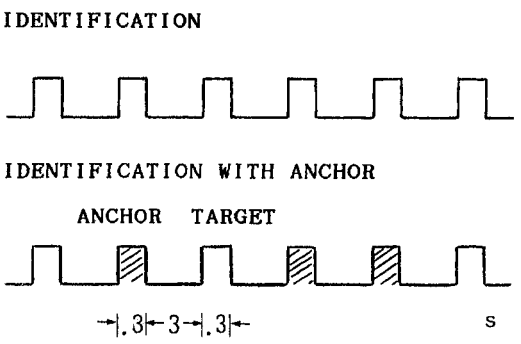

Fig. 1 Presentation of stimuli in an anchoring paradigm. Anchor is presented as a member of a stimulus series.

tation of stimuli. All the stimuli were presented to the subjects through loudspeaker about 75 dB SPL in a quiet room.

2.1.3 Subjects Fifteen normal hearring undergraduate students participated in Experiment 1. All were native speakers of Japanese and had no experience in hearing synthetic speech. They were divided into two groups. One group consisted of nine subjects who were required to identify the anchor and the test stimuli as beginning with [b] or [d]. The other group, consisting of the six remaining subjects, were required to identify both the anchor and test stimuli such as [be] or [de].

2.2 Results and Discussion

The percentage of responses indicating that the target began with [d] or [de] for each identification test for each subject was approximated by a cumulative normal distribution whose mean and standard deviation were obtained by a least square method. The mean indicates the subjective middle stimulus number on the [be]-[de] continuum. The standard deviation serves as an index for the accuracy of the judgements.

Figure 2 shows the pooled data for the percentage of responses indicating that the stimulus begins with [d] or [de] in the two group conditions. Since there were no differences between the responses of the two conditions, the responses of the two groups were put together. The identification curve shifted greater in the case of the typical [be]-anchor than in any other anchor. The [bi]-anchor moved the identification curve slightly. The [be]'- and [bi]'-anchors also moved the identification curve so as to produce contrast, but the shift was not found to be significant. Thus, the [be]-anchor was the most effective. The difference in the mean stimulus number between the non-anchor identification and the identification with the [be]-anchor conditions in the group who were not required to judge the anchor, was found to be significant by two-tailed t-test [$t(8)=2.425$, $p<0.05$]. There is also great difference between the results from the [be]-anchor and the [bi]-anchor by the one-tailed t-test [$t(8)=4.873$, $p<0.01$]. There were no significant differences between standard deviations in all conditions.

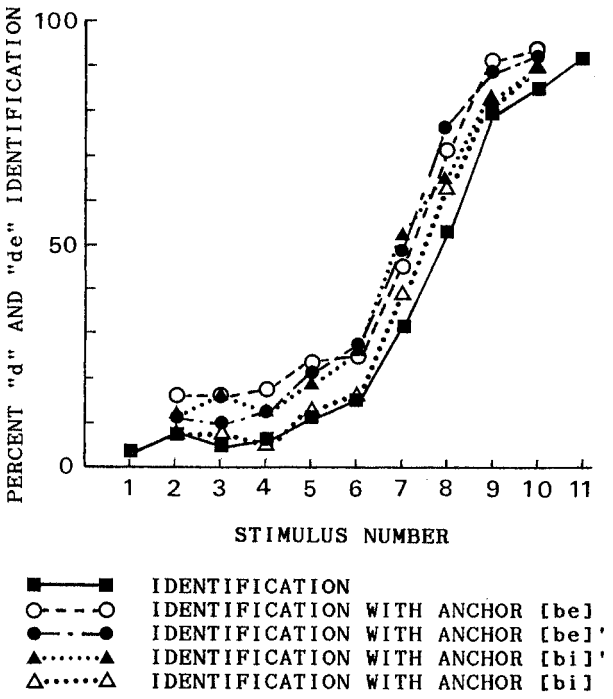

Fig. 2 The percentage of "d" and "de" identification responses obtained from the two conditions of anchor in Experiment 1.

Table 2 shows the subjective middle number and the percentage of identification of anchors obtained from the pooled results in the group who were required to judge the anchors. In the case of [bi]'-anchor, the anchor was perceived to belong to [bi]-category in as large a percentage of the cases as it was perceived to be what it actually was--a typical [bi]-anchor--but the shift caused by [bi]'-anchor seemed to produce a little more contrastive effect than that caused by the real [bi]-anchor, although no statistical significance could be obtained. It can be said that an anchoring effect should be produced in a vowel-contingent manner, more specifically it should be defined at the auditory level regardless of the phonetic codes.

The present data revealed that when the similarity of anchor to target is greater, there are more anchoring effects. The vowel-contingent effect in the anchoring paradigm as well as in the selective adaptation seems to occur at the auditory level. Some problems still remain unexamined in the vowel-contingent anchoring effect. Whether the vowel-contingent effect can be observed when the anchor and stimulus series are presented in a pair has not yet been examined. Another important problem is that of the temporal characteristics of the vowel-contingent effect, that is, the duration of contingent effect. Experiment 2 examined these points, using the anchors on the [be]-[ba] continuum.

Table 2. The effect of the identity of anchor upon the vowel contingent effect.

		Boundary between [be]-[de]	Identification of anchor (%) [be]	[bi]	Others
Identification		8.63	--	--	--
Anchoring	[be]	7.53	98.1	0.0	1.9[de]
	[be]'	7.67	80.8	18.5	0.7[de]
	[bi]'	7.59	0.4	99.5	0.1[de]
	[bi]	7.79	0.3	99.4	0.3[di]

III. EXPERIMENT 2

3.1 Method

3.1.1 Stimuli The stimulus series consisted of nine CV syllables, each with a duration of 250 ms. The steady-state portion appropriate to the Japanese vowel [e] followed the transitive portion in all syllables. The tapes were prepared under the same method as in Experiment 1.

Four anchors were selected from the [be]-[ba] continuum. They were a typical [be] in Japanese, an ambiguous sound [be]' (intermediate but a little nearer to [be]), an ambiguous sound [ba]' (a little nearer to [ba]), and a typical [ba] in Japanese. In the case of [ba]-anchor, the steady-state Japanese vowel [a] which followed the transitive portion, was 780Hz for F1, 1200Hz for F2, and 2700Hz for F3. The first formant frequency for the first 30-60ms of each syllable varied from 300Hz to 780Hz. The transitions began at 30-40ms in F2 and F3. The fundamental frequency of all the stimuli was fixed at 130Hz.

The inter-stimulus interval (ISI) between anchor and target was 1.0 s or 3.0 s. Ten random permutations which consisted of nine anchor and target pairs and ten extra pairs yielded a total of 100 pairs. Eight tapes (4 anchors X 2 ISIs) were prepared.

3.1.2 Procedure The subjects were first given the tape for the identification of stimulus series, then given the tape for identification with anchor. The presentation of stimuli is shown in Fig. 3. In the identification of stimulus series the subjects were asked to judge each target. In the identification with anchor, they were required to judge the target, but only to listen to the anchor. Ten judgements on each of the stimulus series were obtained from each subject in each of the anchoring conditions. The stimuli were presented under the same method as in Experiment 1. The interval for identification was 4 s. Five subjects participated in this experiment.

3.2 Results and Discussion

The figures in Table 3 represent the differences (the mean stimulus number in identification of stimulus series minus that in identification with anchor). The plus sign represents that the percentage of [de] increased and the category boundary was moved toward the anchor, that is, contrast occurred.

When ISI was 1.0s, a contrast effect was produced by the [be]-anchor. The other three anchors moved the boundary a little. This result in paired presentation replicates the results obtained in Experiment 1, where the anchor was presented as a member of a stimulus series. When ISI was 3.0s, the shift of the boundary could hardly be obtained by any anchor. It can be said that the anchors are no longer effective and therefore the vowel-contingent effect can not be obtained when the ISI is longer than at least 3.0s.

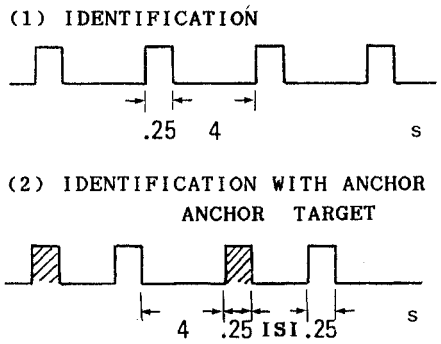

Fig. 3 Presentation of stimuli in an anchoring paradigm in Experiment 2. Anchor is presented as being paired with series stimulus. ISI is either 1.0 s or 3.0 s.

IV. CONCLUSION

The vowel contingent effect seems to occur at the auditory level and receive no influence from the phonetic code at the more abstract phonetic level. The data obtained here are not sufficient to draw any conclusion, but one possibility is that the perceptual mechanism for extracting some features of consonants in CV syllables, has some characteristics that are very short-lived and frequency-specific.

Table 3. The shift of category boundary due to adding anchor to the stimulus series.

ISI (s)	Anchor [be]	[be]'	[ba]'	[ba]
1.0	+1.21	-1.00	-0.41	+0.14
3.0	0.00	-0.75	-0.60	-0.13

REFERENCES

[1]Eimas,P.D. & Corbit,J.D."Selective adaptation of linguistic feature detectors for speech," Cogn.Psychol.4,99-109,1973. [2]Simon,H.J. & Studdert-Kennedy,M. "Selective anchoring and adaptation of phonetic and nonphonetic continua,"J.Acoust.Soc.Am. 64,1338-1357,1978. [3]Shigeno,S."The auduitory tau and kappa effects for voiced stop consonants," Jap.Psychol.Res.29,71-80,1987. [4]Cooper,W.E."Cognitive feature analysis in speech perception," Percept. Psychophys.16,201-204,1974. [5]Bailey,P.J. "Perceptual adaptation in speech: Some properties of detectors for acoustical cues to phonetic distinctions," Unpublished Ph.D.dissertation,Univ. of Cambridge,1975. [6]Diel,R."The effect of selective adaptation on the identification of speech sounds," Percept.Psychophys.17,48-52,1975. [7]Sawusch,J.R. "Peripheral and central processes in selective adaptation of place of articulation in stop consonants," J. Acoust.Soc.AM.62,738-750,1977. [8]Sawusch, J.R. & Pisoni,D.B. "Simple and contingent adaptation effects for place of articulation in stop consonants," Percept.Psychophys,23,125-131,1978.