# A study on dynamical structures contributing to perception of central vowel in /VVV/ concatenated vowels.

Reiko Tako (010066)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2002

**Keywords:** time expansion, time contraction, noise replacement, SFTR(Spectral Feature Transition Rate).

## 1 Introduction

In continuous speech, phonemes are neutralized by incomplete articulation. The central vowel that isolated from /VVV/ concatenated vowel is often perceived as different vowels[1]. Despite of this problem, listeners are able to perceive such ambiguous phonemes as which speakers intend to utter.

For this perceptual procces, various sources of important information can be considered. In those sources, especially spectral transition is considered as important because spectral transition presents movement of articulator and almost of segments in speech is dynamical segments. Actually, the importance of spectral transition was investigated by perceptual experiments using natural speech[2][3].

However, information contained in spectral transition and its effects on speech perception are not clear. Therefore, explanations for mechanism of speech perception are still not completed.

Many preceding studies have investigated about spectral transition and shown interesting result. Lindblom and Studdert-Kennedy[4] hypothesized that listeners perceived vowel with perceptual compensation using the rate and the direction of the formant transition. Suzuki[5] suggested mutually complementary effect between amount and rate of formant transition in perception of vowels. From results of these studies, It is considered that the rates of spectral transitions have some effects on speech perception. However, these studies used synthetic stimuli. The synthetic stimuli imitated spectral transition and may not correctly present information included in complex spectral transition of the natural speech. Therefor, it is not clear whether these results are achieved in perception of natural speech.

On the other hand, Strange and her colleagues[2][3] used natural speech in perceptual experiment. They used mainly /CVC/ silent-centerd stimuli and suggested that time-varying information is necessary for accurate perception of ambiguous vowels. Furthermore, they suggested that perceptually relevant dynamical information is defined over

initial and final transition portions of /CVC/ syllables and this information is available. However, the structure of dynamical information detected perceptually and more detailed information is not yet investigated. The cause is that controlling of natural speech is difficult. The aim of this study is to investigate the effects of spectral transition modified in time domain on speech perception and the factors contained in the spectral transition. Perceptual experiments use stimuli of /VVV/ concatenated vowels which are expanded and contracted in time domain using STRAIGHT[6] and replaced the central vowel segment with noise. The effects of various dynamical structures of spectral transition portions are measured.

## 2 Experiment 1

The aim of this experiment is to measure the effects of various dynamical structures of spectral transition portions.

### 2.1 Stimuli

Perceptual experiments use stimuli of /VVV/ concatenated vowels that are expanded and contracted in time domain using STRAIGHT and replaced the central vowel segment with noise. Time expanding and contracting make control of dynamical information. Replacements with noise control segment lengths that contain dynamical information.

The /VVV/ concatenated vowels are /iai/ in /wariai/, /aia/ in /zaiaku/ and /ioi/ in /nioi/. The speech data are spoken by a male speaker 'mtm' in the ATR speech database.

In the experiment, lengths of /VVV/ are fixed and variables of stimuli are the rate of time expanding and contracting. As shown in figure 1, length of four segments which decided from three center points of vowel (○ circle) and two center points of transition (◎ dual circle) are fixed. Transition segment and vowel segment in each of segment 1-4 are decided and expanded or contacted in three rate patterns:

(1) Decided segments are not expanded and contracted in time domain

(2) Transition segments are expanded and vowel segments are contracted. Spectral transitions of stimuli change from transition segment to static segment toward central vowel segment.

(3) The pattern is opposite to (2). Spectral transitions of stimuli change from static segment to transition segment toward central vowel segment.

The stimuli are replaced central vowel segments by white noise. For stimuli of pattern (2) and (3), noise lengths are three patterns as the remained syllable lengths are long, half and short. For stimuli of pattern (1), all six patterns for pattern (2) and (3) are used.

All of 72 stimuli is modified at 20kHz sampling and 16bit quantization.

### 2.2 Subjects

Subjects were 3 Japanese students with no known hearing impairment.

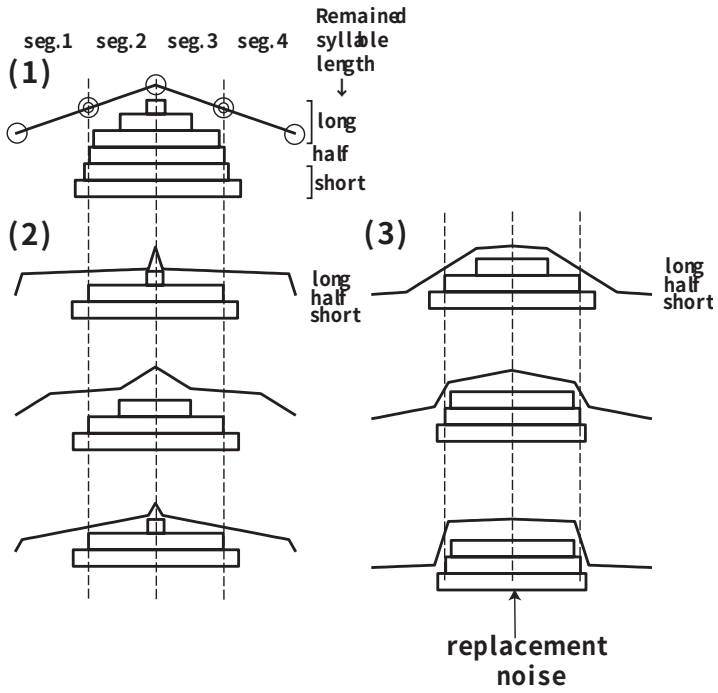


Figure 1: Schema of stimuli in experiment 1.
These are shown by schematic formant transitions.

### 2.3 Procedure

In one session, subjects listened all of 72 stimuli presented randomly and answered the syllables as they could identify using keyboard of notePC. Subjects were allowed to repeat listening one stimulus. If subjects perceived noise as some portions contained in syllables, they could represent the noise portions with 'x' and so on. Each subject listened to 5 sessions in sound proof room. All of stimuli are randomly presented to biannual through headphone in about 55dBSPL.

### 2.4 Exp.1 results and discussion

#### 2.4.1 Existence of Spectral Feature Transition Rate (SFTR)

Perceptual performances show that central vowels are perceptual or not. To study this tendency, the stimuli are analyzed into Line Spectral Frequency (LSF) and Spectral Feature Transition Rate (SFTR). Then, the performances of the subjects show that existence of maximum points of Spectral Feature Transition Rate(SFTR) is necessary for central vowel perception.

#### 2.4.2 Results of pattern (2)

In the case of expanding of spectral transition segment, the stimulus contains SFTR. Therefor, subjects can perceive central vowels. However, as the lengths of the static

segments that continue from the transition segment to the replaced noise are varied, perceived results of the central vowel are also varied. If the length is relatively long, subjects perceive central vowels as phoneme maybe shown by acoustical feature of static portions. On the other hand, if the segment length is relatively short, subjects perceive central vowel as same phoneme as original. As figure 2 shows, the performances are shown by subjects 1 and 3. The performances of subjects suggest those transition rates and lengths of static segments may have some relation.

### 2.4.3 Results of pattern (3)

In the case of contracting of spectral transition segment, the stimuli contain SFTR. However, when the maximum value of the SFTR became too large by contracting of spectral transition segment, the center vowel is not perceived.

This finding leads that the subjects cannot get information for central vowel perception from the rapid transitions.

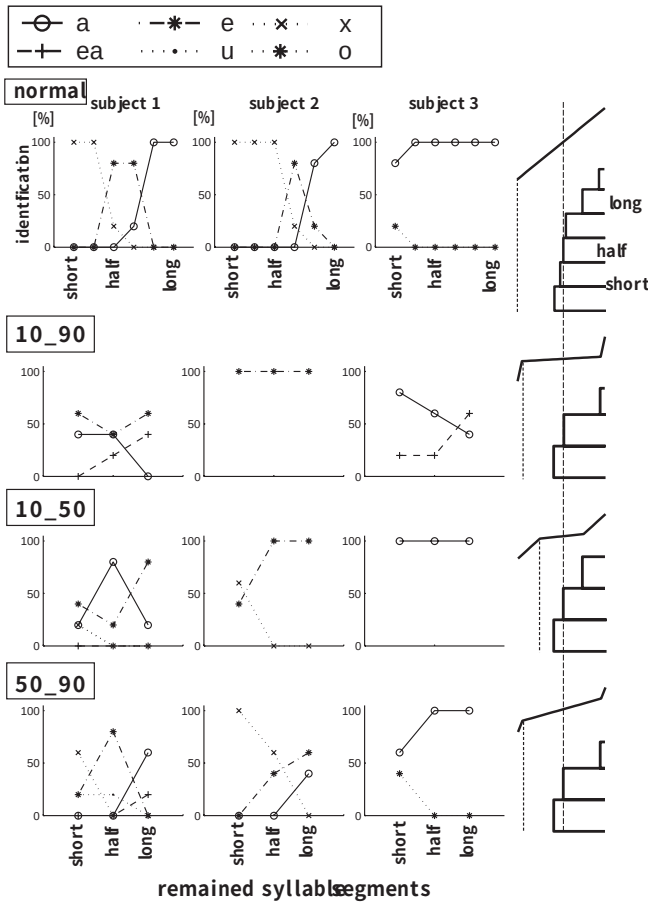


Figure 2: Perceptual performance of /iai/ modified as normal and pattern (2).

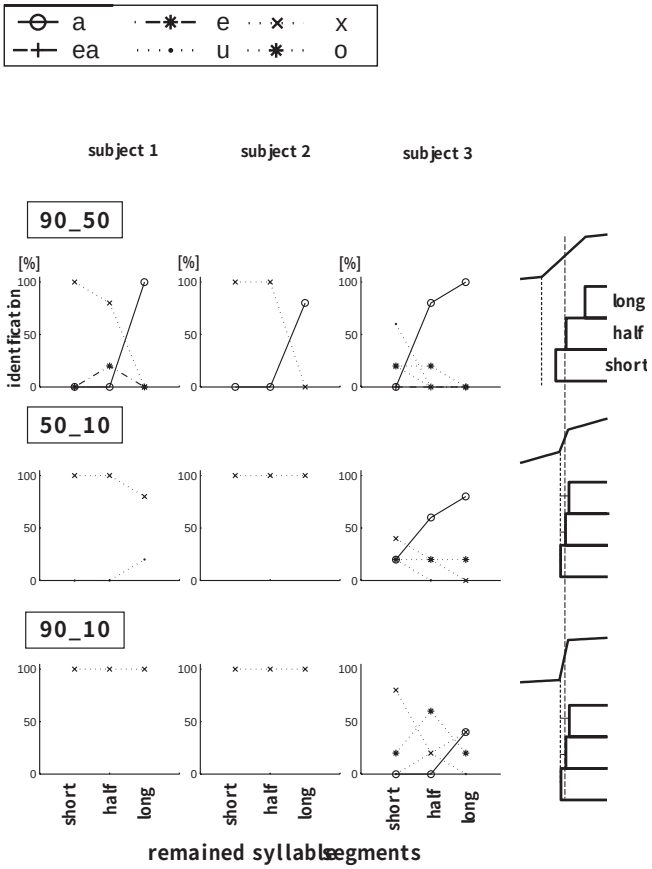


Figure 3: Perceptual performance of /iai/ modified as pattern (3).

# 3 Experiment 2

The aim of this experiment is to investigate the influence of spectral transition of stimuli that modified as pattern (2) and the factors are contained in the spectral transition. The influences of some relation in transition rates and lengths of static segments are investigated in the experiment.

## 3.1 Vowel boundary

In preceding experiment, phonemes presented by each spectral feature of some frames in /VVV/ concatenated vowels and boundaries of the phonemes are measured. Stimuli used for preceding experiments are isolated vowels synthesized using STRAIGHT. Each of stimuli has spectral feature of one frame in /VVV/ concatenated vowels.

Perceptual performance shows those spectral features in static segments of stimuli for experiment 2 present phonemes different from vowels /VVV/.

### 3.2 Stimuli

Stimuli used for experiment 2 are similar to pattern (2) in experiment 1. The original /VVV/ concatenated vowels are same as those used for experiment 1 and preceding experiment, and modified in two types as shown in figure 4. In type A, the centers of transitions in original /VVV/ (◎ shown as dual circle in figure 1) locate at center in transition segment and static segment of stimuli. In type B, the centers of original transitions locate at center of static segments of stimuli. The rates of time expanding (5 times) and contracting ((1) 1, (2) 0.75, (3) 0.5 and (4) 0.25 times) are same in both types.

In the experiment, each vowel length of /VVV/ are variable. The lengths of replacement noise are fixed as same length of segment 2-3 and the end of static segments followed by the replacement noise are truncated. The truncations control segment lengths that contain dynamical information (examples are shown in figure 5). The static segments are truncated by 10 % step from 100 % to 0 % of total length.

### 3.3 Subjects

Subjects were 3 Japanese students with no known hearing impairment. The subject 1 and 3 had served in experiment 1.

### 3.4 Procedure

All of stimuli was divided into two sessions as first half and second half. Each subject listened to 16 (2*8) sessions in sound proof room. Other procedures were same as those described in experiment 1.

### 3.5 Exp.2 results and discussion

Perceptual performances of the subject 2 and 3 are shows similar tendencies. As shown in figure 6, if the transition rate is not fast, the length of static segments at which perceived central vowel varies are long. The contrary is also realized.

On the other hand, performances of the subject 1 are different from that of the other subjects. As shown in figure 7, if the transition rate is not fast, the length of static segments at which perceived central vowel varies are short. The contrary is also realized.

These performances show that there is different influence of static segment lengths and that there is the typical relationship between transition rates and lengths of static segments, when subjects can perceive the central vowel in /VVV/ stimuli as original phoneme.

In preceding experiment, perceived vowels which presented by spectral feature in these static segments are not different by each subject and the vowels are medial phoneme of original phoneme in /VVV/. These results show that the different influences of static segment lengths are not caused by difference in vowel boundaries of each subject. Therefor, the transition segment and static segment are effect to central vowel perception, having typical relationship.

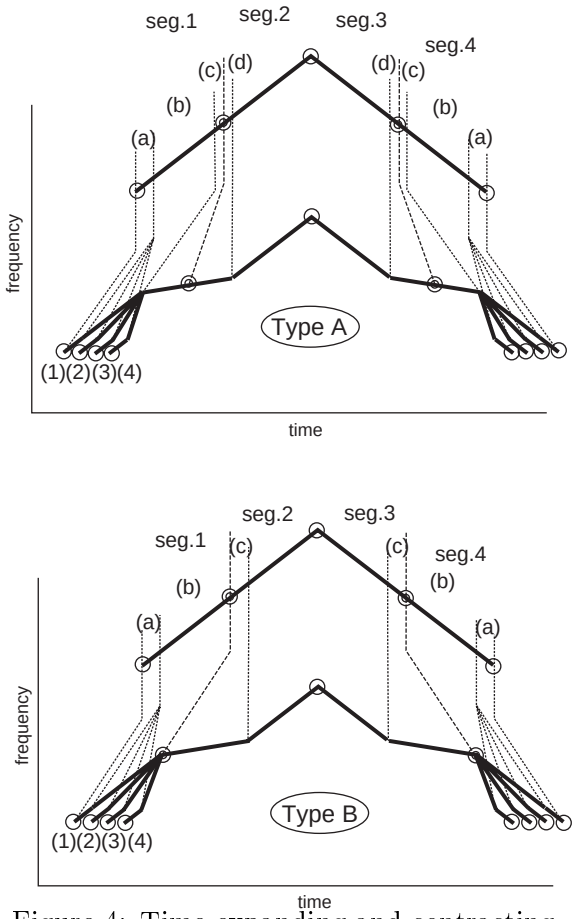


Figure 4: Time expanding and contracting.

(a):unmodified segment. (b):segment for expanding. (c) and (d):segment for expanding.

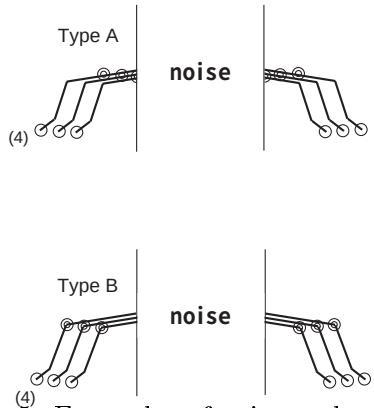


Figure 5: Examples of noise replacement.

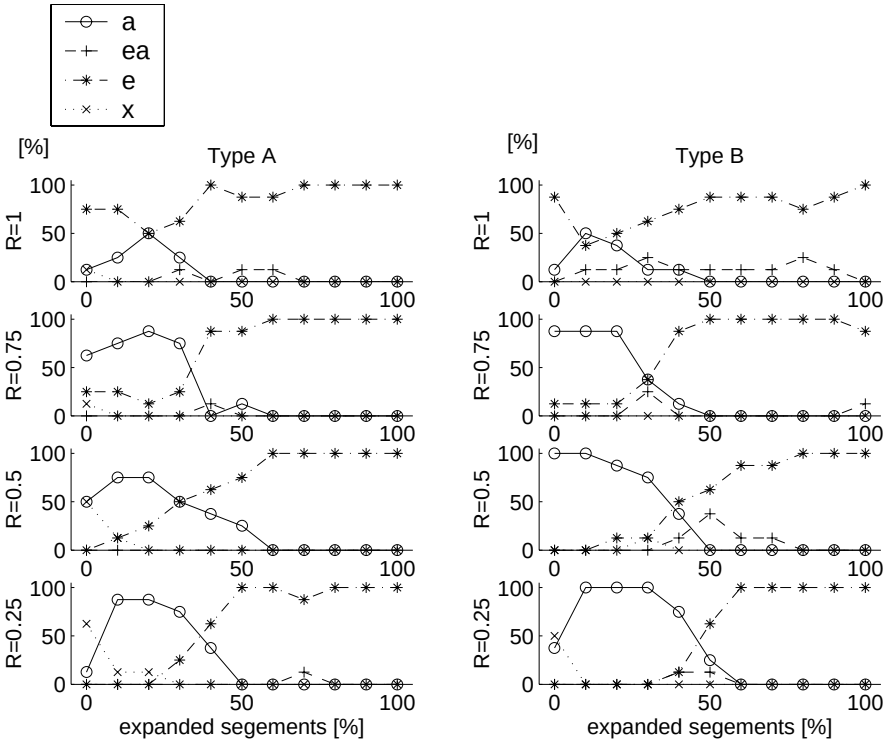


Figure 6: Perceptual performance of /iai/ by the subject 2.

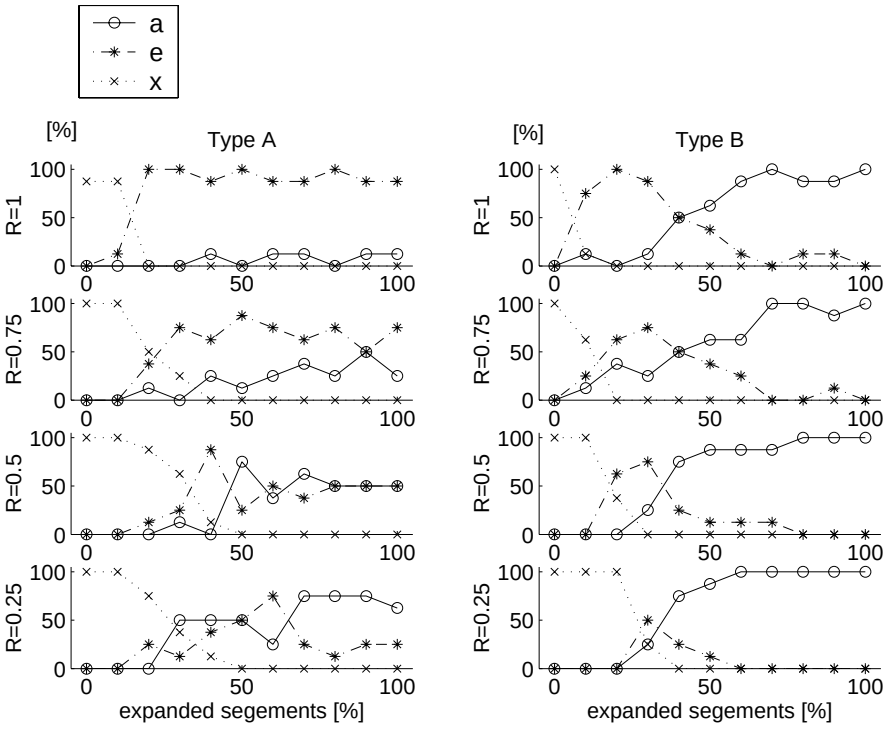


Figure 7: Perceptual performance of /iai/ by the subject 1.

Rates of contracting is presented with 'R'.

# 4 General discussion

## 4.1 Dynamical structure of spectral transition and the factors contained in the spectral transition.

In experiment 1 and 2, when subjects can perceive the central vowel in /VVV/ stimuli as original phoneme, the stimuli have spectral transitions which change from transition segment to static segment toward central vowel segment. The dynamical structures contain the typical relationship between transition rates and lengths of static segments.

The result that transition rates effect on perception of central vowels supports the study of Lindblom and Studdert-Kennedy[4] and of Suzuki[5].

Moreover, if studies of Strange[2]Str will be taken into the result, transition rates and static segment lengths may be used when perceptually relevant dynamical information is defined over initial and final transition. Dynamical information in replacement noise segments may be also defined by the factors of transition rates and static segment lengths. From these suggestions, it is hypothsized that overall dynamical structures of syllables are defined by the factors of transition rates and static segment lengths and used for perception of central vowels in /VVV/.

## 4.2 Fitting of dynamical structure

To study the above hypotheses, the dynamical structures of stimuli used for experiment 2 are decided by fitting the factors with a polynomial function. The perceived phoneme presented by dynamical structure are compared with vowel boundaries measured in preceding experiment. In this fitting, difference in influences of static segments by each subject is not considered, because causes of the phenomenon are not clear.

Before the fitting, F1-3 transitions of the stimuli of /VVV/ are analyzed using the Speech Tools and reconstruct in ERBrate. Then, transition segments and static segments of the formant transition are fit to a polynomial function.

Perceptual performances of the subject 2 for a stimulus (/iai/, type B, contracting rate R=0.25) are shown in figure 8. Results of fitting and vowel boundaries of subject 2 are shown in figure 9. The vowel decided by peak of F1 transition in figure 9 almost correspond to perceived vowel shown in figure 8.

Perceptual performances of the subject 1 for a stimulus (/ioi/, type B, contracting rate R=0.5) are shown in figure 10. Results of fitting and vowel boundaries of the subject 1 are shown in figure 11. The vowel decided by peak of F2 transition in figure 10 also correspond to perceived vowel shown in figure 11.

However, these results are a part of all results, there is possible that dynamical structures of syllables are decided by systematic approximation and used for perception of central vowels in /VVV/ concatenated vowels. The possibility also suggests that there is a prediction mechanism of the central vowel.

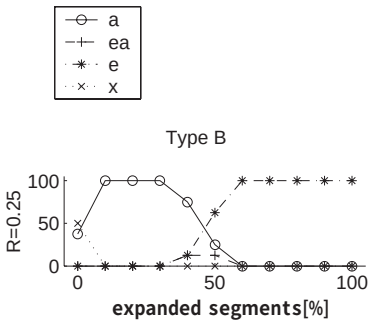


Figure 8: Perceptual performance of the sbujects 2 in experiment 2. Stimuli are /iai/ modified with R=0.25.

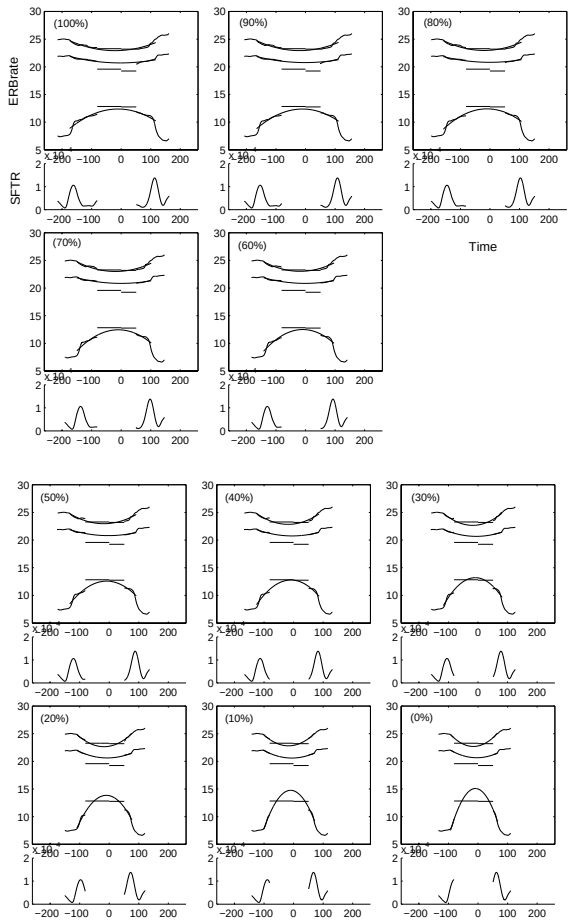


Figure 9: The results of fitting the stimuli used in figure 8. (The subject 2)
The graph in top of left correspond to the right of horizontal axis in figure 8. The graph in bottom of right correspond to the left of horizontal axis in figure 8. 0-ms point in horizontal axis of each graph is the center point of central vowel in /iai/. Boundaries of vowels are plotted in a straight lines. /e/ is in below the lines and /a/ is in above the lines.

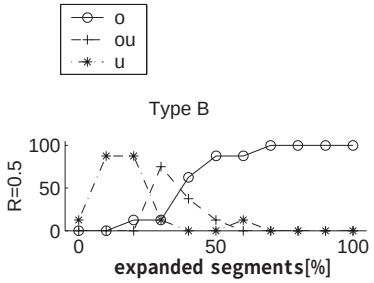


Figure 10: Perceptual performance of the sbujects 1 in experiment 2. Stimuli are /ioi/ modified with R=0.5.

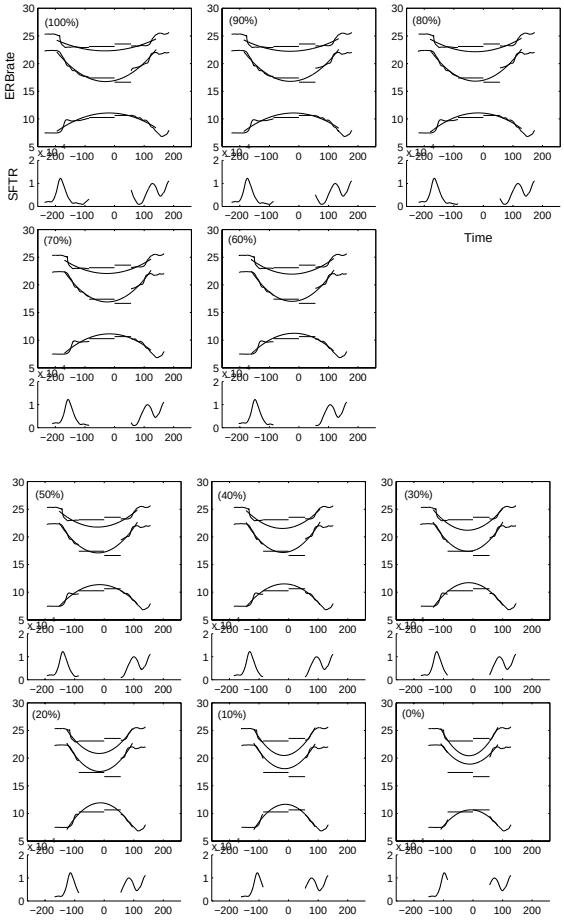


Figure 11: The results of fitting the stimuli used in figure 8. (The subject 1)
The graph in top of left correspond to the right of horizontal axis in figure 8. The graph in bottom of right correspond to the left of horizontal axis in figure 8. 0-ms point in horizontal axis of each graph is the center point of central vowel in /iai/. Boundaries of vowels are plotted in a straight lines. /u/ is in below the lines and /o/ is in above the lines.

## 5 Conclusion

This study investigates the effects of spectral transition modified in time domain on speech perception and the factors contained in the spectral transition. The results suggest that:

- when subjects can perceive the central vowel in /VVV/ stimuli as original phoneme, the stimuli have spectral transitions which change from transition segment to static segment toward central vowel segment.
- When perceived central vowel varies, there is the typical relationship between rate of transition segments and length of static segments.
- there is a prediction mechanism of the central vowel using transition rates and static segment lengths.

## References


[1] Kuwabara and Sakai. (1973). "Normalization of coarticlation effect for a sequence of vowels in connected speech,", J. Acoust. Soc. Jpn. 29, 2, 91-99.

[2] Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels,", J. Acoust. Soc. Am. 74, 695-705.

[3] Strange, W. (1989). "Dynamic specification of coarticulated vowels spoken in sentence context,", J. Acoust. Soc. Am. 85, 2135-2153.

[4] Lindblom, B. E. F., and Studdert-Kennedy, M. (1967). "On the role of formant transitions in vowel recognition,", J. Acoust. Soc. Am. 42, 830-843.

[5] Suzuki. (1974). "Mutually complementary effect between amount and rate of formant transition in perception of vowels, semivowels and voiced stops and a possible mechanism for their identification,", J. Acoust. Soc. Jpn. 30, 3, 169-180.

[6] Kawahara. (2001). "GUI-STRAIGHT:Getting started (for ver.23),".