ISCA Archive
http://www.isca-speech.org/archive

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

ANALYSIS OF TEMPORAL COORDINATION BETWEEN ARTICULATORY MOVEMENTS AND PITCH CONTROL IN THE REALIZATION OF JAPANESE WORD ACCENT BY A PATIENT WITH APRAXIA OF SPEECH

Shinobu Masaki, Itaru F. Tatsumi and Sumiko Sasanuma

Tokyo Metropolitan Institute of Gerontology
35-2 Sakaecho, Itabashi-ku, Tokyo, 173 Japan

ABSTRACT

Characteristics of articulatory movements and pitch control were investigated in the realization of Japanese word accent for a patient with apraxia of speech. Japanese meaningful words beginning with [ai] which have an upward pitch transition at the end of the initial mora were uttered at normal (5.0 mora/s) and slow (3.3 mora/s) speaking rates by a patient with apraxia of speech and five normal controls. Formant trajectories for the initial vowel sequence and pitch contour were extracted and analyzed based on the models proposed by Fujisaki [1]. The results for the patient indicated marked delay of pitch control relative to articulatory movements for the slow speaking rate, and shortening of duration of pitch transition in proportion to the delay of initiation of pitch control. A positive correlation of amplitude of accent command with duration of pitch transition was observed for both normal subjects and the patient with apraxia of speech. The correlation indicates compensatory control of amplitude of accent command in order to produce sufficient upward pitch shift against the gradual fall of the voicing component of pitch contour.

INTRODUCTION

In 1977 Darley et al. defined apraxia of speech as "an articulatory disorder resulting from impairment, due to brain damage, of the capacity to program the positioning of speech musculature for the volitional production of phonemes and the sequencing of muscle movements for the production of words" [2] (p.255). In pure cases impairment in muscle function, which is shown in dysarthric patients, is not ordinarily noted. In addition, performance in all modalities of language is normal except in oral production, which differentiates apraxia of speech from aphasia.

The nature of the abnormalities in oral production has been investigated for both articulation and prosody. Many researchers have investigated the characteristics of articulation errors based on auditory impression and acoustical analysis [3]-[8]. Recently developed articulatory movement observation systems, such as the fiber scope and the X-ray microbeam system, have demonstrated the incoordination of various articulatory movements, causing articulatory errors [9][10]. However, only a few characteristics of the abnormalities in prosody have been reported, such as 1) slow speech characterized by lengthening and separation of syllables, and 2) less variation in peak intensity across syllables [4][11]. In particular, pitch contour, which is one of the most important features of prosody, has not been investigated intensively because an adequate analysis method has not been adopted.

In the analysis of normal speech, Fujisaki investigated the nature of pitch control in terms of its temporal coordination with articulatory movements [1]. He estimated underlying command for vowel transition (articulatory command) and for pitch transition for accent (accent command) in the production of a vowel sequence word [ai] with upward or downward pitch transition between [a] and [i]. The analysis of the temporal relationship between these commands revealed the tendency for accent command to be delayed relative to articulatory command.

In the present study, the dynamic characteristics of articulatory movements and pitch control for a patient with apraxia of speech were investigated using the analysis method proposed by Fujisaki.

DATA COLLECTION

Subjects

Subjects comprised a 27-year-old female patient S.U. with relatively pure apraxia of speech and five 23- to 33-year-old normal females.

S.U. suffered a Cerebro-Vascular Accident caused by an embolism resulting from heart disease six years before the data acquisition. A CT scan showed the circumscribed low density areas in the cortex and subcortical white matter of the second and third frontal gyri, and precentral gyrus of the left hemisphere. The results of the Standard Language Test of Aphasia (SLTA) [12] demonstrated performances within the normal range in all modalities except oral production. Paralyses in the right side of the upper and lower extremities and face were observed.

Speech Materials

Speech materials were 16 three- to five-mora words beginning with [ai], all having an upward pitch transition at the end of the initial mora. Table 1 shows examples of speech materials with their pronunciation, accent pattern, orthographic representation in Japanese and meaning.

The subjects repeated the model utterance of each word presented on an audio tape recorder at a normal speaking rate of 5.0 mora/s, and a slow speaking rate of 3.3 mora/s which was the natural speaking rate of the patient S.U. One utterance for each word for each speaking rate was recorded for analysis.

ANALYSIS

Estimation of Articulatory Command

The first three formant frequencies were extracted pitch-synchronously using the "Analysis-by-Synthesis" method [13]. Articulatory command for the second vowel [i] was estimated based on the functional model proposed by Fujisaki [1]. In the present study, observed formant frequencies were considered to be the consequence of the smoothing process for an articulatory command which made the transition from the target formant frequencies of [a] to that of [i]. According to the model, the first three formant trajectories $F_n(t)$ (n=1 to 3) for the vowel sequence [ai] were represented by the following equations.

$$F_1(t)=F_{11}+(F_{12}-F_{11})G_f(t-T_f), \quad (1)$$

$$F_2(t)=F_{21}+(F_{32}-F_{21})G_f(t-T_f), \quad (2)$$

$$F_3(t)=F_{31}+(F_{22}-F_{31})G_f(t-T_f), \quad (3)$$

where,

$$G_f(t)\begin{cases}=\mathrm{Min}[1-(1+\alpha t)\exp(-\alpha t),\ 0.9] & \text{for } t\geq 0,\\ =0 & \text{for } t<0.\end{cases} \quad (4)$$

The symbols in Eqs. (1) to (4) represent

- F_{n1}: the n-th target formant frequency for the initial vowel [a],
- F_{n2}: the n-th target formant frequency for the second vowel [i],
- T_f : instant of onset of articulatory command for the second vowel [i],
- α : natural angular frequency of the second order system for the articulatory command.

$G_f(t)$ is a step response of a critically damped second-order linear system with a ceiling limit of 0.9. The ceiling limit was set so that $G_f(t)$ reached the target value within a finite time interval.

In the sequence from a back vowel [a] to a front vowel [i], the second and the third formant frequencies of [a] transfer to the third and the second formant frequencies of [i], respectively. However, observed formant trajectories do not show the intersection between the second and the third formant frequencies. This can be explained by an analogy with a double-tuned circuit, in which the coupling of two resonant circuits causes the separation between their resonant frequencies [14].

A set of parameters specifying the articulatory command (F_{n1}, F_{n2}, T_f and α) were estimated using "Analysis-by-Synthesis" of observed formant trajectories; i.e., the optimum set of parameters was obtained so as to synthesize the closest approximation to the observed trajectories [1].

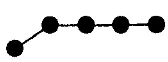
Estimation of Accent Command

Pitch contour was extracted from the speech wave by measuring intervals between peaks corresponding to the opening of the glottis. Accent command was estimated based on the functional model of the generation process of pitch contour proposed by Fujisaki [1]. According to Fujisaki's model, pitch contour $F_0(t)$ was assumed to be a combination of voicing and accent components. In the present study, logarithmic pitch contour $\ln F_0(t)$ was represented by the following equation.

$$\ln F_0(t)=\ln F_{min}+A_vG_v(t-T_v)+A_aG_a(t-T_a), \quad (5)$$

where,

Table 1. Examples of speech materials

Accent pattern and Pronunciation	Orthographic representation in Japanese (Meaning)
 a- i-ma	合間 (interval)
 a- i- i-ɾo	藍色 (indigo blue)
 a- i-da-ga-ɾa	間柄 (relation)

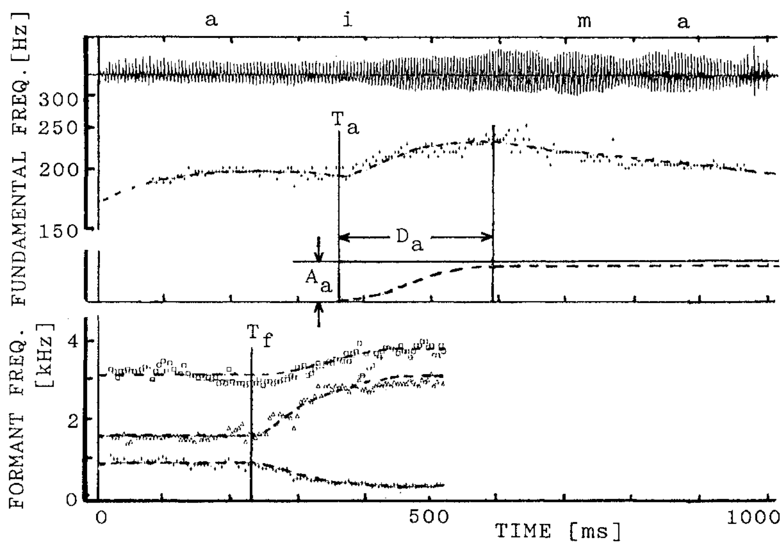


Fig.1 An example of extracted speech wave, pitch contour and formant trajectories with estimated results of accent command and articulatory command for the patient with apraxia of speech.

$$G_v(t) = \begin{cases} \beta^2 t \exp(-\beta t) & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases} \quad (6)$$

and

$$G_a(t) = \begin{cases} \mathrm{Min}[1-(1+\gamma t)\exp(-\gamma t),\ 0.9] & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases} \quad (7)$$

The symbols in Eqs. (5) to (7) represent

- F_{min}: bias level upon which the voicing and accent components are superposed,
- A_v : amplitude of voicing command,
- A_a : amplitude of accent command,
- T_v : instant of onset of voicing command,
- T_a : instant of onset of accent command,
- β : natural angular frequency of the second order system for the voicing command,
- γ : natural angular frequency of the second order system for the accent command.

In Eq.(5), the second and the third terms represent the voicing and accent components, respectively. $G_v(t)$ in the voicing component is an impulse response of a critically damped second-order linear system. $G_a(t)$ in the accent component is a step response of another critically damped second-order linear system with a ceiling limit.

In the analysis of pitch contour, a set of parameters was estimated using "Analysis-by-Synthesis" of observed pitch contour [1].

Measurement

Figure 1 shows an example of the results of the estimated articulatory and accent commands for a Japanese word "aima" (meaning "interval") uttered at a slow speaking rate by the patient S.U. On the top, the observed pitch contour is shown by an array of crosses. A broken line, which overlaps with the array of crosses, represents a synthesized pitch contour which gives the closest approximation to the extracted pitch contour. In the middle of the figure, the extracted accent component is represented by a broken line. At the bottom, the first three formant trajectories are represented by arrays of crosses, triangles and squares. Broken lines represent synthesized trajectories which give the closest approximation to the extracted formant trajectories.

In the present study, the following parameters were measured (see Fig.1).

- T_a: instant of onset of accent command,
- A_a: amplitude of accent command which is represented by difference in natural logarithmic frequency,
- D_a: duration of pitch transition, whose termination is defined as the instant at which the value of $G_a(t)$ saturates at the ceiling value (0.9),
- T_f: instant of onset of articulatory command for the second vowel [i].

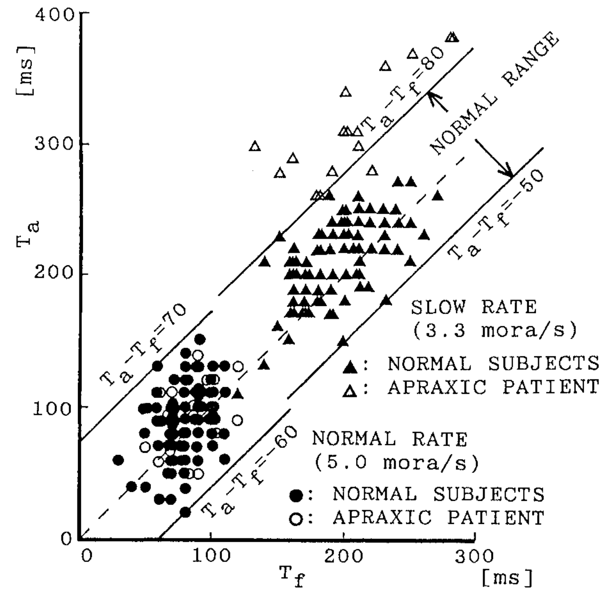


Fig.2 A scatterplot of onset of accent command (T_a) and onset of articulatory command (T_f).

RESULTS

Figure 2 shows the temporal relationship between onset of accent command and that of articulatory command (T_a vs. T_f). Circles represent the data points for a normal speaking rate of 5.0 mora/s, and triangles for a slow speaking rate of 3.3 mora/s. At each speaking rate, filled marks represent data points for the five normal subjects, while open marks represent data points for S.U.

For normal subjects, accent commands occurred either before or after articulatory commands. The delay of accent command relative to articulatory command specified by T_a-T_f ranged from -60ms to 70ms at the normal speaking rate, where a negative value represents cases in which accent command precedes articulatory command. At a slow speaking rate, delay of accent command tends to be greater than at a normal speaking rate. The range of accent command delay was between -50ms and 80ms. In addition, two subjects out of five demonstrated that all accent commands occurred simultaneously with or after articulatory commands.

For the patient S.U., the accent command delay for a normal speaking rate ranged from -40ms to 50ms, i.e. within the normal range. At the slow speaking rate, however, all accent commands occurred after the articulatory commands, and the range of accent command delay was between 60ms and 170ms. In 13 speech samples out of 16, the delay of accent command exceeded the range for normal subjects.

Figure 3 shows a scatterplot of the instant of onset of accent command and the duration of pitch transition (T_a vs. D_a). At a normal speaking rate, ranges of both the instant of onset of accent command (T_a) and the duration of pitch transition (D_a) for the patient S.U. were almost the same as those for normal subjects. However, at a slow speaking rate, the patient demonstrated shorter duration of pitch transition and delay of instant of onset of accent command when compared to normal subjects. As seen in Fig.3, data points for a slow speaking rate demonstrated a negative correlation between D_a and T_a ($r=0.41$, $p<.001$). The negative correlation suggests that delayed initiation of accent command in the patient S.U. is compensated for by a shortening of the duration of pitch transition.

Figure 4 shows a scatterplot of the amplitude of accent command and the duration of pitch transition (A_a vs. D_a). Filled circles and open circles represent data points for normal subjects and the patient S.U., respectively. For normal subjects, positive correlation between the amplitude of accent command and the duration of pitch transition was observed (r=0.79, p<.001). It indicates the compensatory control of amplitude of accent command so as to realize the sufficient upward pitch shift against the gradual fall of the voicing component of pitch contour (see Fig.1); i.e., higher amplitude of accent command is necessary for the longer pitch transition since the voicing component, on which accent component is superposed (see Eq.(5)), gradually decays during the pitch transition. For the patient S.U., most data points were within the normal range. This indicates that the compensation in controlling the amplitude of accent command is preserved even in the patient.

Finally, as shown in Figs.3 and 4, minimum duration of the pitch transition for the patient S.U. (20ms) was shorter than for normal subjects (30ms), suggesting that the ability for the patient S.U. to control pitch was as fast as normal subjects. Therefore, the influence of the

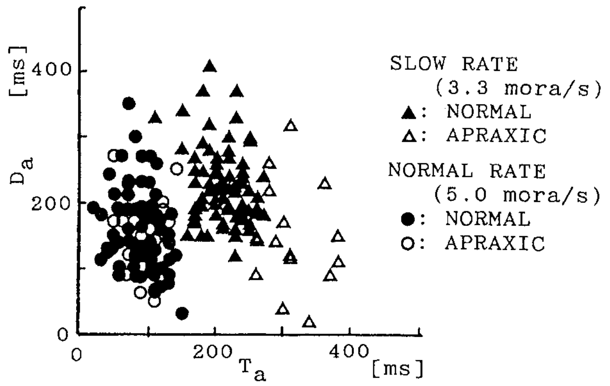


Fig.3 A scatterplot of the duration of pitch transition (D_a) and instant of onset of accent command (T_a).

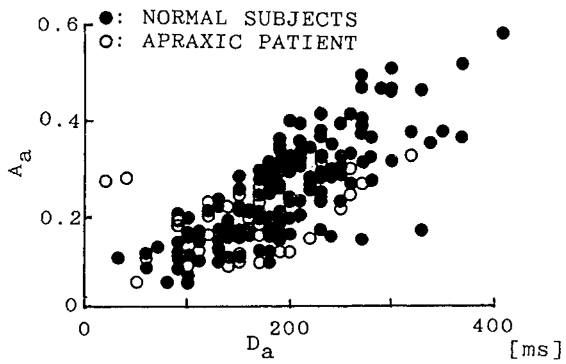


Fig.4 A scatterplot of the amplitude of accent command (A_a) and the duration of pitch transition (D_a).

facial paralysis of the right side on the control of pitch is minimum.

SUMMARY

(1) The patient of apraxia of speech S.U. demonstrated a disturbance of temporal coordination between pitch control and articulatory movements at a slow speaking rate.
(2) Duration of pitch transition for the patient S.U. tended to be shortened when initiation of pitch control was delayed.
(3) The amplitude of accent command (A_a) was positively correlated with the duration of pitch transition (D_a) for both normal subjects and the patient S.U. This correlation suggests the compensatory control of amplitude of accent command in order to produce sufficient pitch shift against the gradual fall of the voicing component of pitch contour no matter how long the duration of pitch transition is.
(4) The fastest pitch transition found in the patient S.U. suggests that the disturbance of pitch control was not attributable to the impairment of muscle function in controlling pitch.

ACKNOWLEDGMENTS

We are grateful to Mrs. Mariko Fujibayashi and Dr. Shoichi Tsuchida at Toranomon Hospital for their help in obtaining speech samples. Above all, we would like to thank our patient S.U. for her cooperation and patience.

REFERENCES

[1] Fujisaki, H., In M. Sawashima and F. S. Cooper (Eds.), "Dynamic Aspects of Speech Production," Tokyo, Univ. of Tokyo Press, 348-369, 1977.
[2] Darley, F. L., Aronson, A. E. and Brown, J. R.,"Motor Speech Disorders," Philadelphia, London, Toronto, Saunders Company. 1975.
[3] Johns, D. F. and Darley F. L., J.S.H.R., 13, 556-583, 1970.
[4] Sasanuma, S., Ann. Bull. of Res. Inst. of Logopedics and phoniatrics, Univ. of Tokyo, 5, 85-89, 1971.
[5] Deal, J. L. and Darley, F. L., J.S.H.R., 15, 639-653, 1972.
[6] Trost, J. E. and Canter, G. J., Brain and Language, 1, 63-79. 1974.
[7] Shankweiler, D., Harris, K. S. and Taylor, J. L., Archives of Physical Medicine and Rehabilitation, 49, 1-8, 1968.
[8] LaPointe, L. L. and Johns, D. F., Journal of Communication Disorders, 8, 259-269.
[9] Itoh, M., Sasanuma, S. and Ushijima, T., Brain and Language, 7, 227-239, 1979.
[10]Itoh, M., Sasanuma, S., Hirose, H., Yoshioka, H. and Ushijima, T., Brain and Language, 11, 66-75, 1980.
[11]Kent, R. D. and Rosenbek, J. C., J.S.H.R., 26, 231-249, 1983.
[12]Hasegawa, T. (Ed.),"Aphasia: Mechanisms and Management," Tokyo, Kongo Shuppan, 1980. (in Japanese)
[13]Fujisaki, H., Nakamura, N. and Yoshimune, K., J. of Acoust. Soc. Jpn., 26(3), 152-154, 1970.
[14]Fujisaki, H., Yoshida, M. and Sato, Y., National Convention Record, I.E.C.E., 1437, 1974. (in Japanese)