A Study of a Human Interface Device Controlled by Formant Frequencies for the Disabled

Norihiro Uemi

Department of Architecture and Mechatronics, Faculty of Engineering, Oita University, Japan
uemi@oita-u.ac.jp

Abstract. We propose a human interface device using a formant frequency. It does not require complicated signal processing. The formant frequency means vocal tract resonance frequency which characterizes the phonemes. By using a microphone, we think it is possible to estimate the position of tongue and the opening of mouth. We investigated the relationship of formant frequencies and the manner of articulation. At first, we clarified the range of the formant frequencies by moving the mouth and the tongue freely. Next, we set the horizontal and vertical axes which correspond to the second- and first-formant frequencies respectively on a computer display screen. Subjects were asked to fit a mouse pointer controlled by their formant frequencies to a target appeared on a display. It is concluded that subject can be relatively easy to control the pointer within the range enclosed by five Japanese vowels formant points.

Keywords: Human Interface Device, Formant, Speech, the Disabled.

1 Introduction

A voice recognition system is one of the useful methods of human interface devices for the disabled. But it requires complicated signal processing and it is difficult to recognize the voice of the person who has articulation disorder.1,2,3

We propose a human interface device using a formant frequency that is one of the voice characteristic values. It does not require complicated signal processing. The formant frequency means vocal tract resonance frequency which characterizes the phonemes. It is known that fronting or backing of the tongue causes chiefly a raising or lowering of second formant frequency (F2) and closing or opening of the mouth causes a lowering or raising of first formant frequency (F1).[4]

By using a microphone, we think it is possible to estimate the position of tongue and the opening of mouth from formant frequencies. And the disabled might be able to utilize the function that was not used effectively before for a human interface device.

In this paper, we investigated the relationship of formant frequencies and the manner of articulation.

In the first experiment, we clarified the range of the formant frequencies by moving the mouth and the tongue freely.

P.L.P. Rau (Ed.): CCD/HCII 2013, Part I, LNCS 8023, pp. 340–345, 2013.

In the next experiment, we set the horizontal and vertical axes which corresponded to the second- and first-formant frequencies respectively on a computer display screen. Subjects were asked to fit a mouse pointer controlled by their formant frequencies to a target appeared on a computer display screen.

2 The Range of the Formant Frequencies by Moving the Mouth and the Tongue Freely

In this paper, we propose a human interface device using the first and second formant frequency. It is necessary to check the range of the formant frequencies, because the range differs according to person.

In particular, if we use them as an input device for the disabled, it is necessary that the disabled can generate the formant frequencies easily. We investigated a range and an individual difference of the formant frequencies by moving a mouth and a tongue freely.

2.1 Experimental Method

In this experiment, an electrolarynx was used as a sound source. An electrolarynx is one of the artificial larynges. In the electrolarynx, an electromechanical vibrator is attached to the neck to put a sound source into oral cavity.5 It has a big burden on subject for a long time to continue vibrating the vocal cords to measure the formants. Therefore we think that using an outside sound source is practical. We input the sound that is uttered using the electrolarynx on a computer. The sound is performed linear predictive coding and the formants are extracted. The program to analyze the formants was written using LabVIEW programming software (National-Instruments). It analyzed the first peak frequency (F1: first formant frequency) of the spectrum of the sound and following peak (F2: second formant frequency). We can apply this analysis method to an artificial voice that is generally used in the formant analysis of the vowel.6

Analysis condition

- Vibrating frequency of an electrolarynx……….66Hz
- Sampling frequency ………………………….8000Hz
- Analyzing points ……………………………512 points
- Window weighting……………….Hamming window
- Analyzing order……………………………..……12^{th}

At first the subject practiced the electrolarynx. He was asked to move his mouth and tongue freely without overdoing it. The sound was recorded for two minutes. And this recording was repeated three times. The formant frequencies were analyzed afterwards. Five male healthy subjects participated in this experiment.

2.2 Experimental Results and Discussion

Fig. 1 shows the F2-F1 chart of the five subjects. In this chart, the vertical axis shows the first formant frequency and horizontal axis shows a second formant frequency. The point on the F2-F1 chart shows the second and the first formant of the sound from the mouth. In this paper, we call this point ‘formant point’. The F2-F1 chart is related to the position of the tongue and the shape of the mouth. For example, a formant point rises when the mouth is closed, and it falls when the mouth is opened. The formant point moves to the right when the tongue moves backward, and it moves to the left when the tongue moves forward. The white line on the chart is connecting the formant points of the vowels which a subject uttered using the electrolarynx. The existing area of the formant points is different depending on the subjects. The area where the formant points gather is near the formant point of each vowel. It is easy to generate the formant points between /a/ and /o/ and near the vowel position in Fig.1. The area with a few formant points is between/a/ and /e/, between /i/ and /u/, and a center of the area enclosed by the white lines. We do not use this area by everyday utterance. We think that it is difficult to generate a formant point in this area freely.

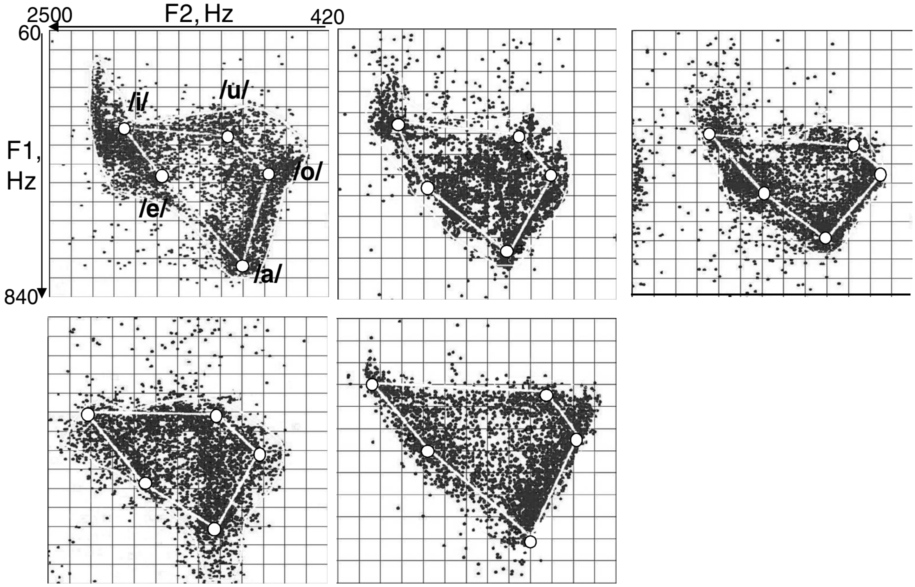

Fig. 1. The range of formant frequencies by moving the mouth and the tongue freely

3 The Range of Formant Frequencies When Subject Was Asked to Fit a Mouse Pointer Controlled by the Formant Frequencies to a Target Appeared on a Computer Display Screen

In the next experiment, we set the horizontal and vertical axes which correspond to the second- and first-formant frequencies respectively on a computer display screen. The plane on the screen is called “F2-F1 plane” and the point represented by first- and second-formant frequency on the F2-F1 plane is called “formant point” in this paper. As mentioned before, Subject can be relatively easy to control the pointer within the area enclosed by five Japanese vowels (/a/,/i/,/u/,/e/,/o/) formant points. But we thought that it was hard to generate the formant point at the center in this area because a few formant points existed at the center. Therefore we presented a target on the F2-F1 plane and asked subject to adjust a formant point to it. And we investigated the position that was easy to adjust the formant point and the position that was hard to adjust it in detail.

3.1 Experimental Method

Subjects were asked to fit a mouse pointer controlled by their formant frequencies to a target appeared on a computer display screen. Fig.2 shows the target position on the computer display screen. In this experiment, we investigated the formant point on the F2- F1 plane which could be controlled consciously.

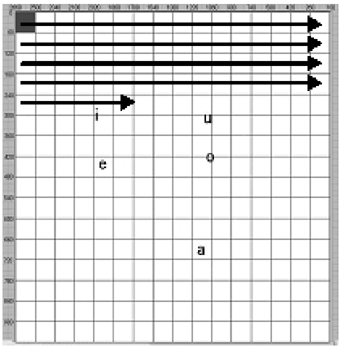

Fig. 2. Target position on the computer display screen

We set the horizontal and vertical axes which correspond to the second- and first-formant frequencies respectively on a computer display screen. We call the plane on the screen 'F2-F1 plane'. We prepared the target which moved on the F2-F1 plane in the display screen and asked the subject to fit a mouse pointer to the target. The target started from left-top position and moved to the right automatically every two seconds. When a target came to the right-side end, it moved to left-second position. The target was moved until right-bottom position. If a mouse pointer was fitted to the target and a switch was pushed, this target position was recorded. We performed this ten times and added up the number of times every position. Five male healthy subjects participated in this experiment.

3.2 Experimental Results and Discussion

Fig.3 is the experimental result. Target position where subject can fit a mouse pointer to a target many times is painted dark color. In Fig.3, we display the line connected the outline of the range of the formant points in Fig.1 and the line connected the position of the vowel. The area where the subject could generate the formant points expanded compared to the area in Fig.1. The target position chosen more than seven times were near the position of the vowel formant. In addition, the target positions chosen many times tended to be between /a/ and /o/. These positions were relatively easy to generate formants. The subject who could not choose the center area in Fig.1 could choose this area in Fig.3.

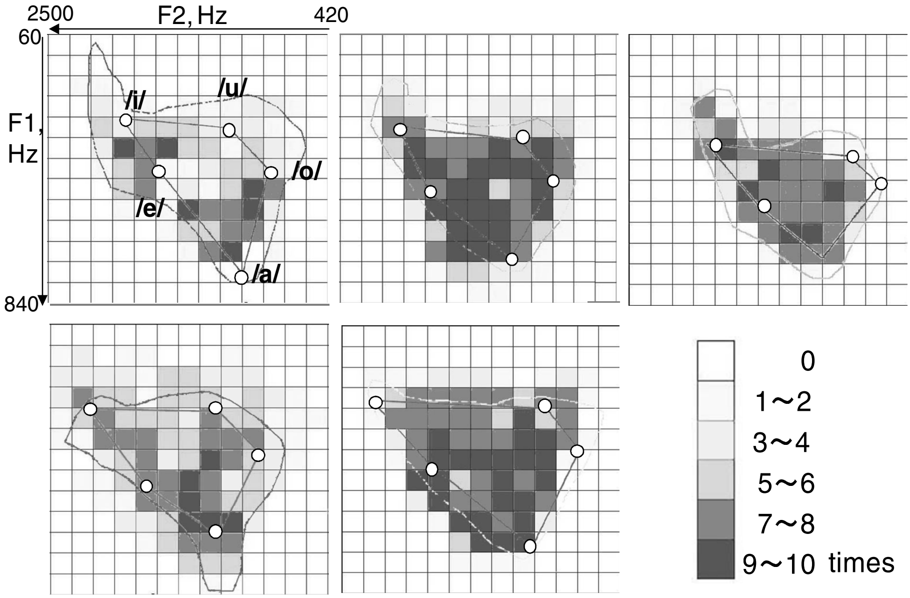

Fig. 3. Target position where subjects can fit a mouse pointer controlled by their formant frequencies to a target appeared on a computer display screen

Because subjects were able to choose the range near the center in the area enclosed by vowels, formants of these positions may be available for an input device. All subjects were able to choose the formant of the range near the vowel /a/. From this, it is easy to move a mouth and a tongue in the area that is close to the formant point of /a/. The area that was easy to produce a formant point differed with different persons. When we use formant for an input device, coordinating a function is necessary.

4 Conclusion

We propose a human interface device using a formant frequency that is one of the voice characteristic values. In this paper, we investigated the relationship of formant frequencies and the manner of articulation.

In the first experiment, we clarified the range of the formant frequencies by moving the mouth and the tongue freely. The Existing area of the formant points is different depending on the subjects. The area where the formant points gather is near the formant point of each vowel. It is easy to generate the formant points between /a/ and /o/ and near the vowel position. The area with a few formant points is between /a/ and /e/, between /i/ and /u/, and a center of the area enclosed by the formants of vowels.

In the next experiment, we set the horizontal and vertical axes which corresponded to the second- and first-formant frequencies respectively on a computer display screen. Subjects were asked to put a mouse pointer controlled by their formant frequencies over a target appeared on a computer display screen. In this experiment, we investigated the formant point on the F2- F1 plane which could be controlled consciously. The target positions chosen many times tended to be between /a/ and /o/. These positions were easy to generate formants relatively. Because subjects were able to choose the range near the center in the area enclosed by vowels, formants of these positions may be available for an input device.

The area that was easy to produce a formant point differed with different persons. When we use formant for an input device, coordinating a function is necessary.

References

1. Noyes, J., Frankish, C.: Speech recognition technology for individuals with disabilities. Augmentative and Alternative Communication 8(4), 297–303 (1992)
2. Kojima, H., et al.: Inarticulate speech recognition for the control of wheelchairs by the severely disabled. In: 29th Annual RESNA conference (2006)
3. Simpson, T., Gauthier, M., Prochazka, A.: Evaluation of Tooth-Click Triggering and Speech Recognition in Assistive Technology for Computer Access. Neurorehabilitation and Neural Repair 24(2), 188–194 (2010)
4. Fant, G.: Ericsson Technics 15(1), 21–23 (1959)
5. Uemi, N., Ifukube, T., Takahashi, M., Matsushima, J.: Design of a New Electrolarynx having a Pitch Control Function. In: IEEE International Workshop on Robot and Human Communication, pp. 198–203 (1994)
6. Markel, J.D., Gray, A.H.: Linear Prediction of Speech. Springer, New York (1976)