FORMANT FREQUENCIES FOR INCREMENTALLY VARYING VOCAL TRACT

Hisashi J. Wakita

Panasonic Technologies Inc.
Santa Barbara, CA 93105, USA

ABSTRACT

This paper concerns the behavior of the formant frequencies for a dynamically varying vocal tract to investigate the sensitivity of the formant frequencies to different parts of the vocal tract. The sensitivity functions for diphthongs were obtained by simulating the varying vocal tract using measured vocal tract area functions for five steady-state vowels.

1. INTRODUCTION

A better understanding of the relationships between vocal tract configurations and their acoustic characteristics has always been important not only in the area of speech production, but also more recently in the inverse problem of estimating the vocal tract shape from acoustic data. Insight into the dynamics of the vocal tract is expected to especially contribute to better speech synthesis based on articulatory models and also is expected to clarify some of the physiological constraints to be imposed on a method for estimating either vocal tract shapes or the midsagittal view of the vocal tract based only on the resonance frequencies of the vocal tract. In some of the recent studies on static vowels, attention has been directed to the clarification of the detailed sound production processes. For instance, some vocal tract spatial parameters derived from the distribution of kinetic and potential energies inside the vocal tract at a particular resonance mode have been studied by Fant and others[2] to clarify the affiliation of formant frequencies with the particular part of the vocal tract and also to clarify the sensitivity of the formant frequencies to local perturbations in the cross-sectional areas of the vocal tract. Along this line of thinking, this paper concerns the behavior of the formant frequencies for a dynamically varying vocal tract. Particularly, the contribution of different parts of the vocal tract to increments in a given formant frequency were investigated together with the sensitivities of the formant frequencies to given parts of the vocal tract.

2. EXPERIMENT

As a preliminary step, the vocal tract area functions for five Russian vowels measured by Fant[1] were used to simulate the five diphthongs which typically appear in American English. Each area function was represented by the concatenation of 34 acoustic tube sections. In order to compute the formant frequencies for a dynamically varying vocal tract, the difference in the areas between two vowels was linearly interpolated into ten intervals. Between each two time successive intervals, the formant frequencies were computed by varying two sections at a time, from glottis to the lips, thus resulting in a total of 170 steps to represent the movement of the vocal tract from one vowel to the other. In computing the formant frequencies, a transmission line analog of the vocal tract was simulated on a digital computer, taking all the energy losses within the vocal tract and the lip radiation into account[3]. The glottis was assumed to be closed in this study. The transfer function of the vocal tract was computed by sweeping the frequency in two Hertz steps, so that the first three formant frequencies could be obtained by peak-picking, with an accuracy of one Hertz.

As a result of this analysis, a set of formant frequency increments, F_{ij} , can be obtained. In this case, ΔF_{ij} represents the increment of the i-th formant frequency affiliated with the j-th section of the area function. Thus, the total formant frequency increment is represented by the sum of all the ΔF_{ij}'s as given by

$$\Delta F_i = \sum_{j=1}^{N} \Delta F_{ij} \qquad (1)$$

The sensitivity of the formant frequency increment is then defined in this study as a ratio of the formant frequency increment to the area increment of the j-th section given by

$$S_{ij} = \frac{1}{K_i} \left| \frac{\Delta F_{ij}}{\Delta A_j} \right| \qquad (2)$$

where S_{ij} is normalized by the total sum of the ratios of the formant frequency increment to the area increment as given by

$$K_i = \sum_{j=1}^{N} \left| \frac{\Delta F_{ij}}{\Delta A_j} \right| \qquad (3)$$

This normalization is made so that the sensitivity of a section of the vocal tract will be bounded by zero and one.

3. RESULTS AND DISCUSSION

Fig. 1 shows the results for the vowel movements from /a/ to /i/ and from /a/ to /u/. The top figures show the distribution of the area increment along the vocal tract from the lips to the glottis. The successive three figures for each vowel movement are the sensitivity functions for the first three formant frequencies. In these two cases, the sensitivity for the vowel movement from /a/ to /i/ is rather dispersed along the vocal tract, whereas a relatively high localized sensitivity is observed in the lip and the pharynx regions for the vowel movement from /a/ to /u/. The most noticeable feature of the sensitivity functions among five vowel movements is that there is a place in the vocal tract where the area increment is relatively small and also the sensitivity experiences a local minimum. This is indicated by the small arrows in the figure. This location corresponds roughly to either the 8-th or 9-th section from the lips.

In the midsagittal view of the vocal tract, this location corresponds roughly to the boundary between the oral and pharyngeal cavities. This seems to be a rather strong physiological constraint in moving the vocal tract configuration from one vowel to another. Particularly, when the vocal tract configuration moves from a front vowel to a back vowel or vice versa, the area in-

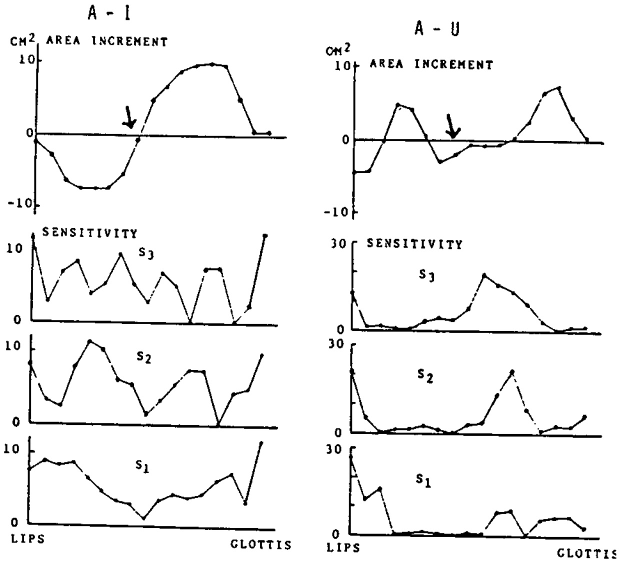

Fig. 1 Area increments and sensitivity functions for varying vocal tracts from /a/ to /i/ and from /a/ to /u/.

crement distribution becomes antisymmetric around this region, to maintain an approximate constant volume of the vocal tract.

Fig. 2 shows the vowel movement from /o/ to /i/ and from /o/ to /u/. The local minima of the sensitivity functions are again indicated by the arrows in the Figure. In the movement from /o/ to /i/, the lip area is highly sensitive, whereas in the movement from /o/ to /u/ the pharynx area as well as the lip area is sensitive.

Fig. 3 shows the vowel movement from /e/ to /i/. A similar tendency is also observed in this case, although the local minima of the sensitivity functions are somewhat obscured, due to the absence of an area increment over a rather broad region around the boundary between the oral and pharyngeal cavities. The sensitivity is high in the front cavity for this case.

4. CONCLUSION

The data obtained in this study will help clarify the relationships between dynamic vocal tract configurations and their acoustic correlates. The data will also be useful first in designing some speech synthesis rules on the basis of an articulatory model, second in gaining some insight into compensatory articulation, and third in finding some physiological constraints in the estimation of vocal tract dynamics from acoustic data.

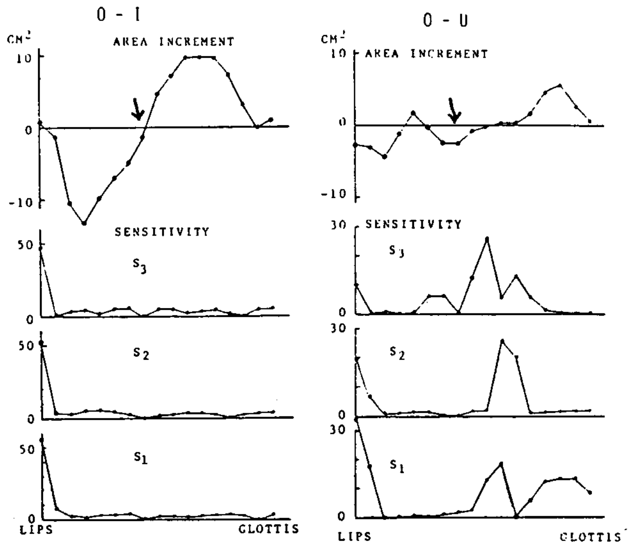

Fig. 2 Area increments and sensitivity functions for varying vocal tracts from /o/ to /i/ and from /o/ to /u/.

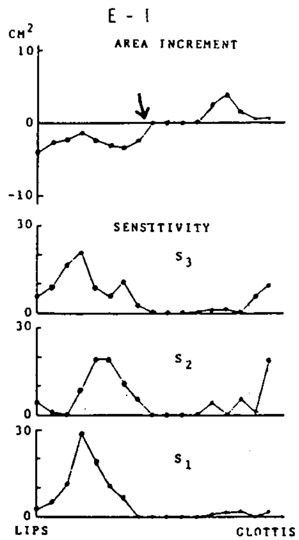

Fig. 3 Area increments and sensitivity functions for a varying vocal tract from /e/ to /i/.

6. REFERENCES

[1]. FANT, G. (1960), "Acoustic theory of speech production", The Hague : Mouton & Co.

[2]. FANT, G. and PAULI, S. (1976), "Spatial characteristics of vocal tract resonance modes", Almqvist & Wiksell Int. : Speech Comm. Seminar, Fant ed.., Stockholm, pp. 121–132.

[3]. Wakita, H. and Fant, G. (1978), "Toward a Better Vocal Tract Model", Quarterly Progress Status Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, Number 1, pp. 9-29.