

SPEECH TRAINING SYSTEMS USING LATERAL SHAPES OF VOCAL
TRACT AND F1-F2 DIAGRAM FOR HEARING-IMPAIRED CHILDREN

Minoru Shigenaga
Department of Computer Science
Faculty of Engineering
Yamanashi University
Takeda-4, Kofu 400
Japan

Yoshihiro Sekiguchi
Department of Computer Science
Faculty of Engineering
Yamanashi University
Takeda-4, Kofu 400
Japan

Three speech training systems for hearing-impaired children were designed and constructed using a minicomputer and a microprocessor. The first system displays the lateral shape of the vocal tract for each vowel estimated from the speech sound. In this system, three storages are prepared. One of them can be used to store a reference shape which may be estimated from a teacher's voice or a prepared standard shape. A child can articulate seeing the reference shape and re-articulate comparing his own first articulation with the reference shape. The system can also compare the estimated shape with the reference one and produce instructions with animated cartoons which show where any articulatory defects exist. The second system displays successively the lateral shapes for articulation of any phoneme sequences containing consonants. The third system displays the first two formant frequencies extracted from speech as a spot on the F1-F2 plane, where the regions of vowels are shown in color. Preliminary use of these devices has shown that they complement each other to form a more effective system of speech training.

1. INTRODUCTION

If hearing-impaired children can see their articulation and compare it with a reference articulation in order to bring their articulation closer to the reference, or if they can see the formant frequency locus for their utterance on a F1-F2 plane, where the regions of the vowels are shown, they might be able to understand how they articulated or how they should articulate. In addition, teachers might be more able to teach them how to articulate or know where their articulatory defects exist. Therefore, we have constructed three speech training systems using a minicomputer and a microprocessor. For vowel articulation the first system estimates a lateral shape of the vocal tract from the speech sound and displays it on a storage CR-tube. And it can compare the estimated shape with the reference one and produce instructions with animated cartoons which show where any articulatory defects exist. For consonant articulation we cannot yet estimate a vocal tract shape from the speech sound, but our second system can display successively the lateral shapes of the vocal tract for the articulation of any phoneme sequence containing consonants. Our third system displays the first two formant frequencies extracted from a speech sound as a spot on the F1-F2 plane, where the regions of the vowels are also shown in color. It can also display pitch frequencies.

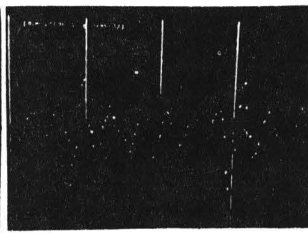
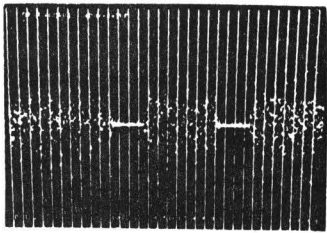
2. SYSTEM 1

This system estimates shapes of the vocal tract for vowels by evaluating PARCOR coefficients and instructs where any articulatory defects exist and also indicates how a child should correct his articulation. Fig. 1 shows the main flow of our procedures. The operation of this system is simple, and one has only to type a number or YES/NO in reply to the system inquiry " = " or " ? " shown in the state ☐ of this diagram. Then the system proceeds to the next state executing the action described along the line, types the characters shown in the new state and waits for the next reply. Referring to the marks in the diagram we will explain what this system can do chiefly.

(a) In order to avoid disturbance from the room noise, you can specify a threshold value in reply to the inquiry "Cut level=" at which the system begins to store the input speech wave for 3.2 sec.

(b) The system displays the reduced input speech wave on a storage CR-tube as shown in Fig. 2 by typing "N" at the state (ii) ☐ Again? . Therefore you can examine the speech wave envelope and select the places where you want to obtain the vocal tract shape.

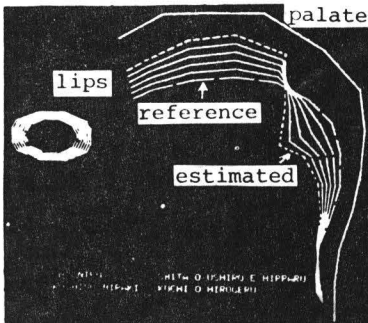
(c) At the state (iii) you can enlarge the wave form at any specified place and find its pitch period as shown in Fig. 3.



(d) Two storages labelled "1,2" for storing the estimated shapes are prepared and you can respond with one of them to the inquiry "Pattern=" at (v). While, in the third storage labelled "0" you should store the shape which you want to use as a reference.

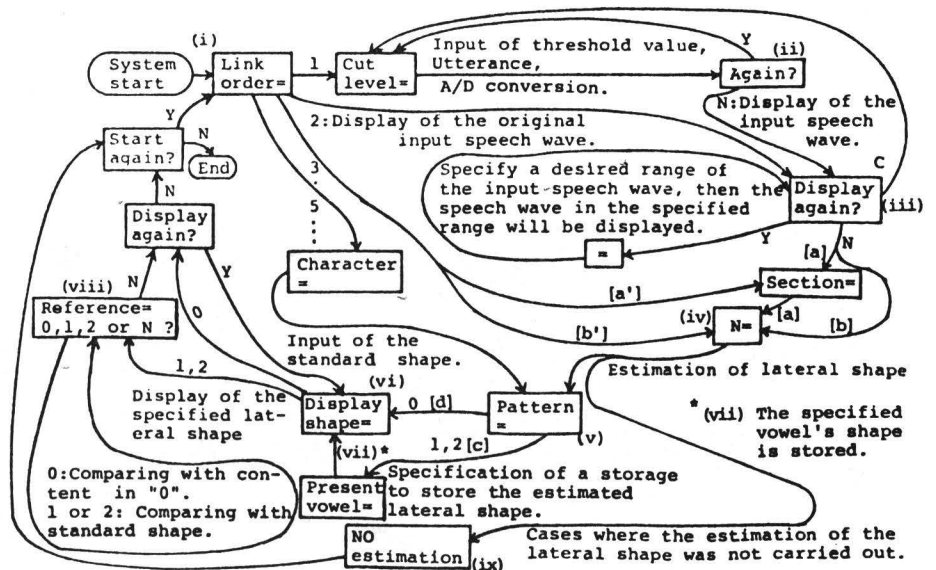
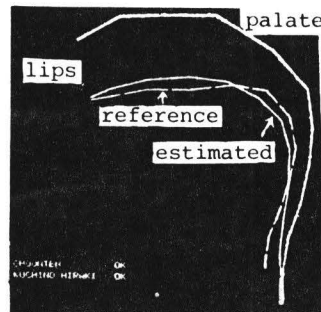
(e) For the inquiry "Display shape=" at (vi) you can specify any storage containing the shape. Then the specified shape is displayed.

(f) At the state (viii) Reference = 0,..?, if you input "0" then the estimated shape is compared with the content in the storage "0", and if you input "1" or "2" then the estimated shape is compared with the standard reference shape which was specified at the state (vii) Present vowel =. Comparison between the estimated shape and the reference one is done for the two features: the place of articulation and the lip-opening. If the difference between the places of articulation of both shapes is over a certain



threshold value, an instruction such as "Draw back your tongue" is shown and the estimated lateral shape moves gradually toward the reference shape as shown in Fig. 4. And if the difference between lip-openings of both shapes is large, an instruction such as "Open your lips" is written and the lips represented by an ellipse move gradually toward the reference opening (Fig.4). If the difference is a little, the instruction "OK" is written as shown in Fig. 5. The reference shape stored in the storage "0" can be the prepared standard shape or the estimated shape which has been estimated from the teacher's voice through the above-described routine.

(g) Going along the route 3 by inputting "3" at the state (i) Link order=, we can estimate and display successively the vocal tract shapes of the vowels at the pre-specified places.



For practice, at first you should display a standard shape prepared previously or a shape estimated from your own utterance. Then a child can articulate seeing the reference shape, and re-articulate comparing his own articulation with the reference shape (Fig. 5). The practice shapes of the child should be stored in either storage "1" or "2", then he can successively compare his articulation with the previous articulation or the reference shape, if he wants, looking at instructions.

3. SYSTEM 2

We have prepared a system which displays synthetic articulatory movements for monosyllable utterances or words containing consonants, because it is very difficult to estimate a vocal tract shape from a speech wave for any of the consonants. The lateral shape of the vocal tract for Japanese phoneme sequences are produced by our speech synthesizer using an articulatory model [1]. The synthesized lateral shapes are stored on a drum storage and displayed with arbitrary speed. A child can easily look at the articulatory movements containing stop or fricative consonants. Fig. 6 shows the synthesized shapes in the sequence /ika/.

4. SYSTEM 3

This system displays the first two formant frequencies extracted from speech as a spot on the F1-F2 plane, where the regions of the vowels are shown in color as shown in Fig. 7. It can also display pitch frequencies. This system uses a microprocessor and a color TV set.

This system is useful by itself, and also helpful in order to know the quantitative vowel quality.



(a) from uniform tube to /i/ (b) from /i/ to /k/ (c) from /k/ to /a/

Fig. 6 Synthesized lateral shapes of the vocal tract for /ika/. In our synthesizer the shape of the vocal tract begins to move from a uniform tube, therefore in (a) it starts from the shape parallel to the upper wall and moves toward the reference shape of /i/.

5. DISCUSSION AND CONCLUSION

Our estimated shapes of the vocal tract are not accurate, but we can approximate the place of articulation for vowels and point out the defect of articulation and, of course, the systems function properly for hearing-impaired children. But an infant is unable to understand instructions such as "Draw back your tongue body", so these systems should be used to offer information to a teacher.

We also hope that these systems are used by older persons for self-teaching.

Though we have not yet evaluated the effectiveness of these systems, preliminary use of these devices has shown that they complement each other to form a more effective system of speech training. Especially it is instructive to use the systems 1 and 3 simultaneously, because we can get articulatory information in both articulatory and frequency domains.

Authors would like to express their gratitude to O. Tanaka, T. Nishide, Y. Furuya and other students in our laboratory for their help.

REFERENCE

- [1] Shigenaga, M. and Ariizumi, H. "Articulatory Movements by Rule" in *Modeles Articulateurs et Phonetique* edited by Carre, R. et al. (G.A. L.F. Groupe de la Communication Parlée), 1977, pp. 193-202.

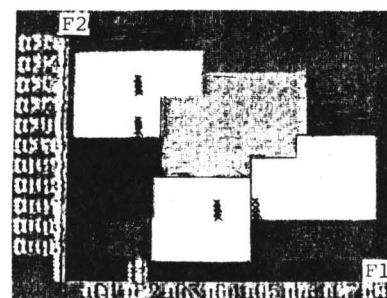


Fig. 7
An output
in the
system 3
for /aoi/.