

Integrated Speech Training System for Hearing Impaired

Sang H. Park, *Member, IEEE*, Dong J. Kim, Jae H. Lee, and Tae S. Yoon, *Member, IEEE*

Abstract—Hearing impaired people have serious difficulties learning to produce intelligible speech. To facilitate speech training for hearing impaired people, many researchers have developed various speech training aids, such as F_0 (fundamental frequency) indicator, N (nasality) indicator, S (fricative /s/) indicator, and spectra displayer. However, most of these training aids treat only a single speech parameter.

In this paper, an integrated speech training system was implemented. The system makes it possible to display vocal tract shape, as well as other speech parameters, in real time. A self-training program for hearing impaired persons has been developed to help them correct errors.

To estimate speech parameters, it is assumed that the speech production process is an autoregressive (AR) model. Using a linear predictive analysis, vocal tract lateral shapes and log spectra are estimated. Log energy of speech signal is used as intensity. Fundamental frequency and nasality are also detected using a vibration sensor.

To check the reliability of the developed system, parameters of normal people were tested, and then, training tests for hearing impaired children were performed. The results showed that our proposed system is very useful for speech training for the hearing impaired.

I. INTRODUCTION

THE phonation of hearing impaired people is usually trained by observing the mouth of a tutor or by detecting motions of the vocal organs by the hands. To help speech training for hearing impaired people, many researchers have developed various speech training aids [1]. Most of them treat only a single speech parameter, such as fundamental frequency, nasality, fricative /s/, and frequency spectra. Shigenaga and Kubo [2] developed speech training systems using vocal tract lateral shapes.

These speech training aids have two severe drawbacks even though they are easy to operate. First, a hearing impaired person can train only a single speech parameter. Second, due to the limits of trainable speech parameters, even if a hearing impaired person has mastered a speech parameter completely, there is a great possibility that he/she may forget previous training during the training of other parameters.

The speech training system that displays the frequency spectra is useful in training of speech but, by using spectra displayer, a trainee can not understand how to correct spectral difference.

Manuscript received March 9, 1994; revised July 6, 1994.

S. H. Park, D. J. Kim, and J. H. Lee are with the Department of Electrical Engineering, Yonsei University, Seoul, Korea.

T. S. Yoon is with the Department of Electrical Engineering, Changwon University, Changwon.

IEEE Log Number 9406449.

Some speech training systems which can display multiple parameters have been developed [3], [4]. However, most of these systems lack well-designed self-training programs. Therefore, a trainee does not know how to correct errors. Without adequate correction, these systems share the same problem as the spectra displayer.

This paper describes an integrated speech training system which can display the vocal tract shapes and other speech parameters such as intensity, fundamental frequency, nasality, and log spectra. This system can display all parameters in real time. A self-training program was also developed to allow easy training for hearing impaired individuals.

To estimate speech parameters, it is assumed that the speech production process is an AR model. Using a linear predictive analysis, vocal tract area functions and log spectra are estimated. Log energy of speech signal is used as intensity. The estimated vocal tract area functions are mapped to corresponding section heights and these are displayed in vocal tract lateral graphics. The first two section heights can be determined by three formant frequencies [5]. The vocal tract graphics display trainee's phonation in the form of a cross-section of the vocal organs. This allows the trainee to know what to correct.

Fundamental frequency and nasality are detected using vibration sensors. To allow the trainee to use two sensors conveniently, sensor attachment tools are made. Wearing the sensor attachment tools, the trainee only needs to handle a microphone.

By using all the estimated parameters and trainee's errors, a self-training program is developed in which the error correction messages can be displayed.

II. SPEECH PARAMETER ESTIMATION

A. Design of Vocal Tract Graphics

The vocal tract graphics describe the cross-section of the vocal organs from the glottis to the lips. First, the axes in which vocal organs move are defined. Using these axes, the vocal tract lateral shapes are displayed. The motion axes are developed, based on the following principles, to represent the motion of vocal organs.

- 1) For the vocal tract model, an acoustic tube model with 15 equal tube sections is used.
- 2) In the vocal organs, the shortest path from the upper to the lower part of each section is selected.

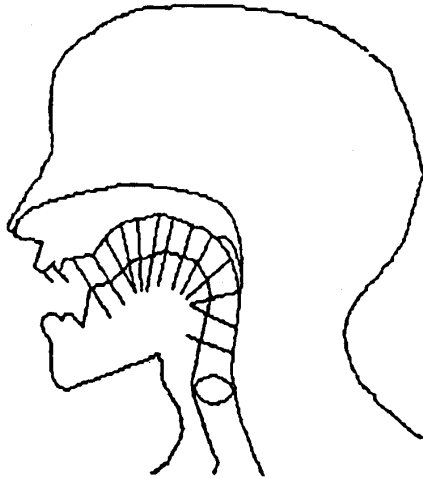


Fig. 1. Draft of vocal tract graphics.

- 3) The developed vocal tract graphics are designed to coincide with X-ray data of the lateral shape of the vocal tract. The X-ray data are based on the results which are used in PARAFAC analysis by Harshman *et al.* [6].
- 4) In the designed graphics, the upper jaw is assumed to be fixed and the lower one to be movable.

For simplicity, the vocal tract is modeled with an assumption that its length is 17 cm long. Generally, the male vocal tract is slightly longer than 17 cm. Children and women have a shorter length vocal tract [7]–[9], so the assumption may cause an error of area function. However, the error can be compensated for by the first two section heights, which are determined by formants. The average shape of each speaker is taken as the basic frame of the vocal tract. On the basis of the above-mentioned principles, the vocal tract length is divided into 15 sections: 2 sections from lips to teeth, 11 sections from teeth to tongue root, and 2 sections from tongue root to glottis. Each motion axis is made to be perpendicular to the vocal tract axis from the lips to the glottis. Fig. 1 shows the draft of vocal tract graphics.

B. Calculation of Vocal Tract Section Height

The vocal tract area function and three formants can determine the section heights of the vocal tract graphics. The vocal tract area, A_n in section n , is as follows:

$$A_n = A_{n-1} \frac{1 + \gamma_n}{1 - \gamma_n} \quad (1)$$

where A_n is in cm^2 and γ_n is the reflection coefficient.

The area functions from the lips to the glottis are estimated using Wakita's boundary condition [10].

The section height in the lips can be determined from the estimated three formants by (2) [5]

$$X_1 = C_1 F_2 + C_2 F_2 F_3 + C_3 F_1 / F_2 + C_4 \quad (2)$$

where $C_1 = 0.300 \times 10^{-3}$, $C_2 = -0.343 \times 10^{-6}$, $C_3 = 4.143$, $C_4 = -0.174$, X_1 is in cm. The constants $C_1 \sim C_4$ are

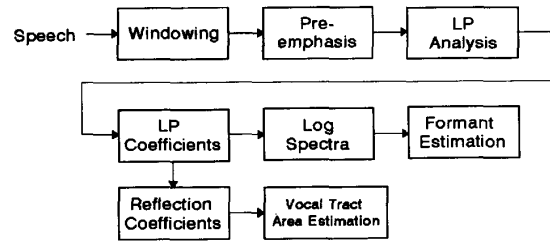


Fig. 2. Block diagram of formant and vocal tract area estimation.

the experimentally adjusted values to match Korean vowels, considering the difference between the formant frequencies of the English vowels and those of the Korean. The section height of the upper and the lower teeth is the average of X_1 and X_3

$$X_2 = \frac{X_1 + X_3}{2} \quad (3)$$

where X_3 is extracted from vocal tract area function of section 3. Fig. 2 is block diagram of the estimation of formants and vocal tract area.

In addition, the estimated formants determine the rounding degree of lips and the height of the jawbone, etc. In mapping the estimated section heights to each axis, the upper part of the vocal organs is assumed to be fixed. That is, there is no motion in the upper teeth, the teethridge, the hard palate, and the soft palate. Therefore, the lower points corresponding to each section height, started from the fixed points of the upper part on each axis, form the shape of the articulator.

C. Other Parameters

In addition to the vocal tract shape, the fundamental frequency, the sound intensity, the nasality, and the frequency spectra are extracted. The fundamental frequency is the most fundamental parameter among them. In the case of hearing impaired people, it is a very frequent phenomena that their tones are higher than normal phonation and considerably irregular. Another serious problem in the phonation of the hearing impaired is the control of the intensity. The intensity is closely related to the fundamental frequency. When a hearing impaired person tries to control the fundamental frequency, the intensity may be extraordinarily loud and/or irregular.

In this paper, the intensity is determined by log energy of speech, considering the characteristics of the ear that recognize the intensity as the log energy. Fundamental frequency and nasality are detected by obtaining the vibrations of the throat and the nose with well-designed devices. Frequency spectra is obtained through the general method of the AR spectrum estimation. Fig. 3 is the block diagram of monitor display procedure of all feature parameters obtained thus far.

III. HARDWARE IMPLEMENTATION

When a person produces a sound, each part of the vocal organs performs its own function.

In this paper, three data acquisition devices were designed. Fig. 4 is the block diagram of speech signal acquisition. In this figure, the sixth-order voltage-controlled voltage source (VCVS) low-pass Butterworth filter and the first-order RC high-pass passive filter are used.

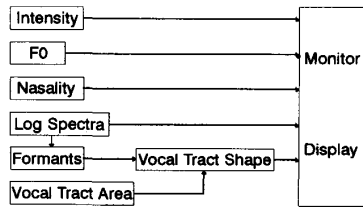


Fig. 3. Block diagram of monitor display.

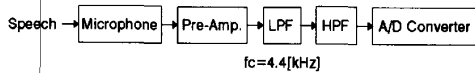


Fig. 4. Block diagram of speech signal acquisition.

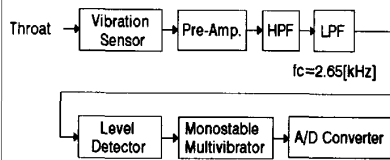


Fig. 5. Block diagram of fundamental frequency detector.

An input speech signal through the microphone is amplified by the total gain of about 350 and passes through the low pass filter (LPF) with cutoff frequency of 4.4 kHz. The low-pass filtered signal passes through the high pass filter (HPF) of $f_c = 1.6$ Hz to cut off the DC components and then, is digitized in an A/D converter of DSP (Digital Signal Processing) board.

A wide variety of software/hardware version pitch detectors have been proposed [11], [12]. Most of the pitch detectors require either a great deal of logic or an excessive amount of computation. The fundamental frequency of the hearing impaired is very irregular, so the detection by software is especially hard. Therefore, a simple hardware version fundamental frequency detector was implemented, as shown in Fig. 5.

A vibration sensor is attached to the neck where the vocal cord vibration is easy to detect. The PKS-4A10 (Murata Co.) is used as the vibration sensor. A vocal cord vibration signal is amplified by a total gain of about 11.9, and passes through a first-order passive HPF and a fourth-order VCVS LPF of $f_c \cong 2.65$ kHz. This signal has several high peaks in a period if input is a voiced sound. The designed level detector generates a pulse if the input signal is beyond a certain level. For voiced sound, the level detector periodically generates pulse trains. The generated pulse trains are stabilized by a monostable multivibrator and each pulse train forms a stable pulse. The pulse signal is finally digitized. Fundamental frequency is determined as the reciprocal of these pulse periods.

Nasal sounds are periodic and their amplitude is bigger than that of other consonants but smaller than that of vowels. In order to detect nasality by the similar method of fundamental frequency detection, a vibration sensor is attached at the side part of the nose. The designed hardware device, as shown in Fig. 6, detects the envelope of the nose vibration signal.

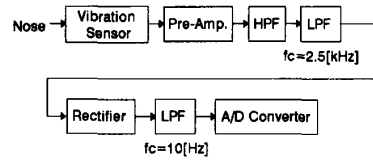


Fig. 6. Block diagram of nasality detector.

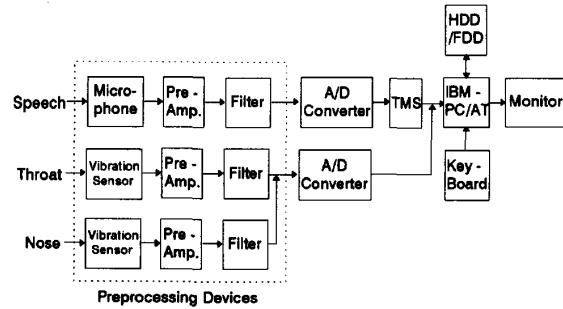


Fig. 7. System configuration.

When a nasal sound is produced, the nose vibration can be detected using the vibration sensor. The detected signal is amplified by a total gain of about 357 and passes through an HPF and an LPF of $f_c \cong 2.5$ kHz. And, the output signal passes through a rectifier and an LPF of $f_c \cong 10$ Hz, then the envelope of the nasality is detected and digitized. The filter types in Fig. 6 are the same as those of Fig. 5.

Among parameters of speech, the fundamental frequency and the nasality can be detected directly from developed devices, but the estimation of the log spectra and the vocal tract shapes require many computations. To obtain near real-time performance, a TMS 320C25 board was used to perform high-speed computations. All the devices were combined, as shown in Fig. 7.

In this figure, the data acquisition devices were made on a board. Speech signals enter the A/D converter of the TMS board, and fundamental frequency and nasality signals are digitized with the DT2821 A/D board by Data Translation Company. These devices are operated in a personal computer system.

IV. DESIGN OF SELF-TRAINING PROGRAM

Because most of the existing speech training systems for the hearing impaired lack proper training programs, it is impossible for the hearing impaired to train speech systematically for themselves. Therefore, it is necessary to guide a trainee carefully.

In this paper, a self-training program was developed by using all the estimated parameters and errors of the trainees. This information was used to display correction messages to the trainee.

A. Self-Training Program for Vowels

The parameters for normal pronunciations are detected experimentally and used as references. Then, the trainee's parameters are compared with the references.

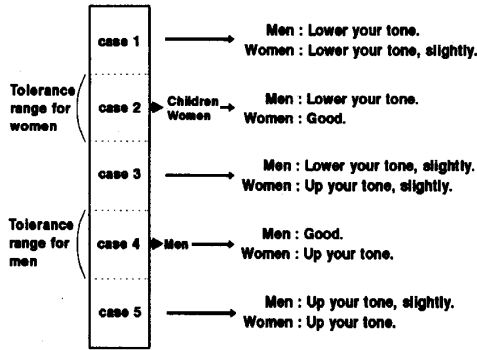


Fig. 8. Judgement criteria for fundamental frequency.

To determine the reference fundamental frequency, first of all, the fundamental frequency of men was detected. Considering the difference between the fundamental frequencies of men and women, each reference fundamental frequency was determined. For the simple display of the reference fundamental frequencies, it is assumed that the fundamental frequency range of children is similar to that of women [11].

1) *Single Vowels*: A self-training program was developed for nine Korean single vowels. The correction messages of the self-training program were produced on the basis of the following principles:

a) *Intensity*: If a trainee's intensity is in the tolerance range, it is normal. Otherwise, the corresponding correction messages are displayed. The tolerance range is decided experimentally. For the typical intensity, the range is 55~58 dB, thus 56 dB was selected as the reference intensity.

Letting I_{ref} be the reference intensity, I_{st} the trainee's intensity, and θ_I the tolerance range, the judgement is performed as

$$\begin{cases} |I_{st} - I_{ref}| \leq \theta_I & \Rightarrow \text{"Good,"} \\ I_{st} - I_{ref} > \theta_I & \Rightarrow \text{"Pronounce weakly,"} \\ I_{ref} - I_{st} > \theta_I & \Rightarrow \text{"Pronounce loudly."} \end{cases}$$

b) *Fundamental Frequency*: For fundamental frequency, first, it is judged whether the trainee's fundamental frequency is in the tolerance ranges of men's or women's. Then, the correction messages for these cases are displayed.

To judge the fundamental frequency and make correction messages, the judgement range is divided into five cases, as depicted in Fig. 8.

c) *Log Spectra*: The square error between the reference log spectra to be trained and the trainee's log spectra is calculated. Using this error, the judgement and correction ordering is performed. Letting S_{ref} be the reference log spectra, S_{st} be the trainee's log spectra, and θ_s be the tolerance range, the judgement is performed as

$$\begin{cases} \sum_{i=1}^{128} (S_{ref} - S_{st})^2 \leq \theta_s & \Rightarrow \text{"Good,"} \\ \text{else} & \Rightarrow \text{"Wrong."} \end{cases}$$

d) *Vocal Tract Graphics*: If the reference height of a jaw is H_{ref} , the trainee's jaw height is H_{st} and the tolerance range

TABLE I
JUDGEMENT CRITERIA FOR CONSONANTS

| Type | Consonants | Judgement Criteria | |
|-------------|------------|--|---------|
| 1. Labial | m | inten +, | nasal + |
| | b | inten +, | nasal - |
| | p | inten -, | nasal - |
| 2. Alveolar | n | inten +, | nasal + |
| | d | inten +, | nasal - |
| | t | inten -, | nasal - |
| | l | "Good" if spectral peak in 1953~2540[Hz] | |
| | s | "Good" if spectral peak in 4199~4785[Hz] | |
| 3. Palatal | c | inten +, | nasal - |
| | c' | inten -, | nasal - |
| 4. Velar | g | inten +, | nasal - |
| | k | inten -, | nasal - |

is θ_H , then

$$\begin{cases} |H_{st} - H_{ref}| \leq \theta_H & \Rightarrow \text{"Good,"} \\ H_{st} - H_{ref} > \theta_H & \Rightarrow \text{"Up your lower jaw and pronounce,"} \\ H_{ref} - H_{st} > \theta_H & \Rightarrow \text{"Down your lower jaw and pronounce."} \end{cases}$$

2) *Complex Vowels*: For 12 Korean complex vowels, the intensity and fundamental frequency are the same as those of single vowels. But for the log spectra and vocal tract shape, two components of a complex vowel are displayed together with the same color, and judgements are also performed in each component.

B. Self-Training Program for Consonants

It is impossible to estimate the exact vocal tract area function for consonants. This is because consonants are pronounced in the tightly constricted state in the vocal tract and there are no resonances or reflections. So, using all the estimated parameters and considering the acoustic features of consonants, a self-training program was developed for consonants.

First of all, the reference vocal tract shapes for corresponding consonants are displayed, and it is assumed that a trainee's vocal organs fit well to the reference shape. Then, by using the judgement criteria in Table I, it is decided whether or not the phonation is correct.

In Table I, "inten-" and "inten+" mean that the starting intensity of phonation should be relatively small or big, and "nasal+" requires that a trainee articulate nasalized phonation. If any conditions are not satisfied, then the corresponding correction message will be displayed.

V. EXPERIMENTS

The vocal tract shape is represented in the right space of the monitor and other parameters are in the left. Before using the training test for the hearing impaired, we tried to check the reliability of the developed system.

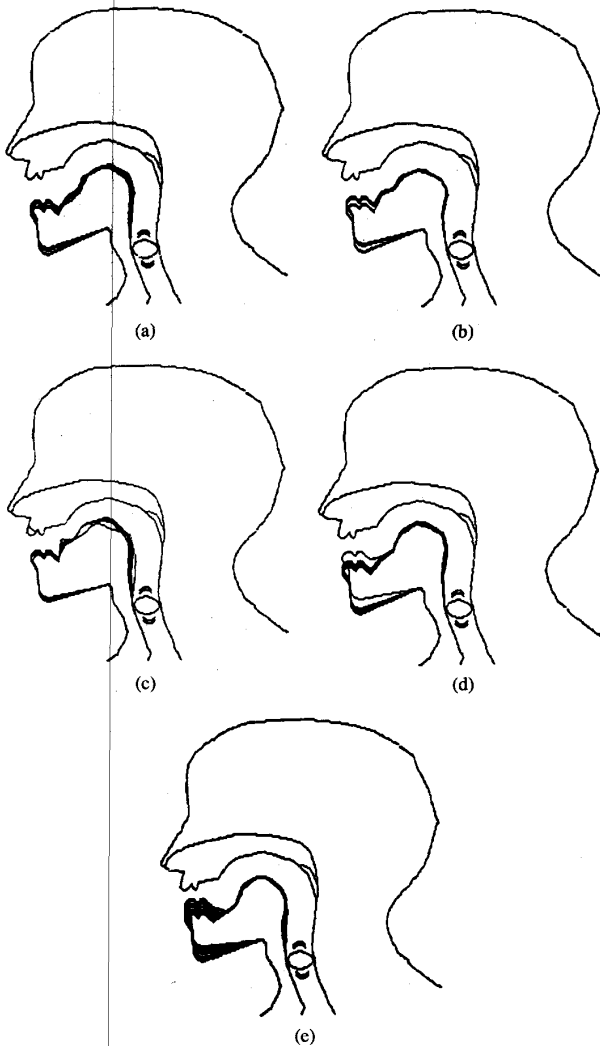


Fig. 9. Estimated vocal tract shapes for five Korean vowels (full frame). (a) /a/. (b) /e/. (c) /i/. (d) /o/. (e) /u/.

A. Simulation

The vocal tract shape and frequency spectra are estimated by using 15 reflection and linear prediction coefficients. First, we checked the estimated vocal tract shapes of four male speakers. Fig. 9 shows the overlapped graphics of the estimated vocal tract shapes in full phonation for five Korean vowels. In Fig. 9, a considerable similarity can be found with actual vocal tract shapes in phonation. The estimated vocal tract shapes and X-ray data are compared in Fig. 10. In Fig. 10, the bold line is X-ray data and the fine line is the estimated data. Fig. 11 shows the estimated intensity and fundamental frequency for vowel /a/. For nasal consonant /m/ + vowel /a/, the estimated nasality and log spectra are shown in Fig. 12.

B. Training Test for the Hearing Impaired

To train speech using our system, the trainee must handle a microphone and two sensors simultaneously. Therefore,

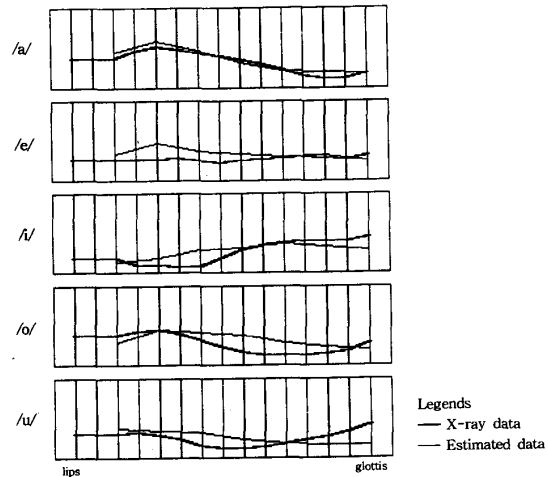


Fig. 10. Comparison of estimated vocal tract shapes and X-ray data.

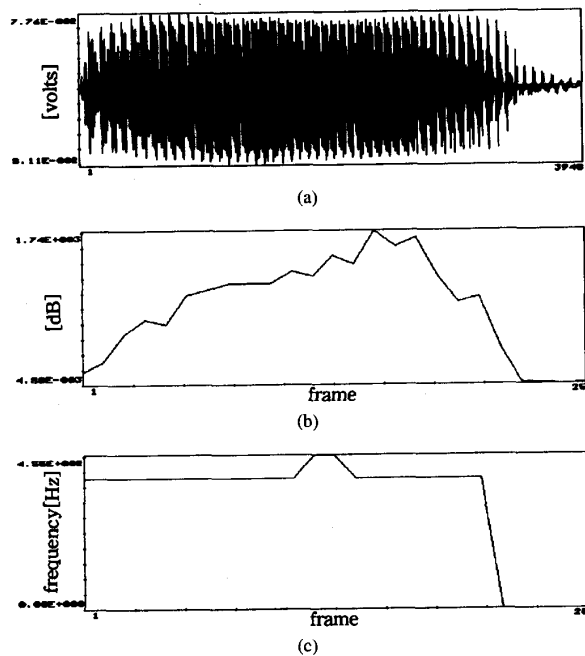


Fig. 11. Intensity and fundamental frequency for Korean vowel /a/. (a) Original waveform of /a/. (b) Intensity for /a/. (c) Fundamental frequency for /a/.

sensor attachment tools were made. An attachment tool for the nasality sensor was designed after the model of the Pierrot's nose. For the fundamental frequency sensor, a tool like a neck band was designed. Wearing these tools, the trainee can process three parameters easily.

The reference data for each parameter was stored. When a trainee tries to learn a phonation, the reference parameters and trainee's data are displayed simultaneously in different colors. The well-designed self-training program can teach the trainee how to correct this error.

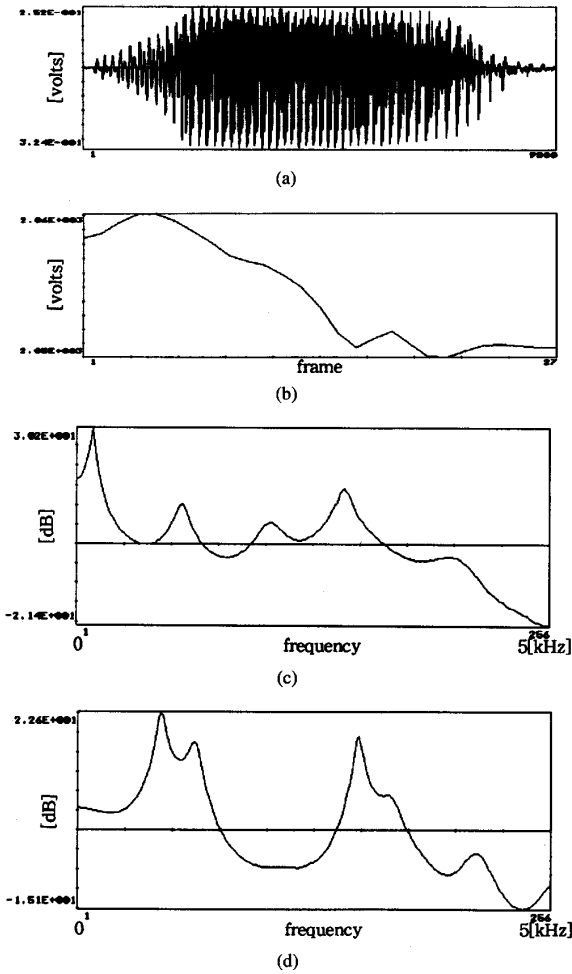


Fig. 12. Nasality and log spectra for /ma/. (a) Original waveform of /ma/. (b) Nasality trajectory. (c) Spectra for /m/. (d) Spectra for /a/.

Two hearing impaired children were selected as subjects for the training test. They are 12~13 years old and have very poor phonation ability. Training tests were performed for complex vowel /ja/ and consonant /p/ for 20 min in a day.

Fig. 13 illustrates displayed parameters in training the phonation /ja/. In real monitors, the reference parameters are displayed in a sky-blue color and the student's parameters in violet. In the windows for intensity, F_0 , and nasality, a bar moves up or down indicating the currently estimated value. Typical values for each parameter are marked in the left side of the windows. The frequency spectra and the vocal tract shapes for the portions of /j/ and /a/ were displayed in the same color. The student must follow them one by one. To catch rapid changes in consonant parameters, display type was modified to display the parameter trajectories. The training results for /ja/ and /pa/ are given in Tables II and III.

The values in Table II represent the achievement rates of each parameter and those in Table III are the achievement rates of the total parameters. Fig. 14 shows the graphical achievements of each parameter for /ja/.

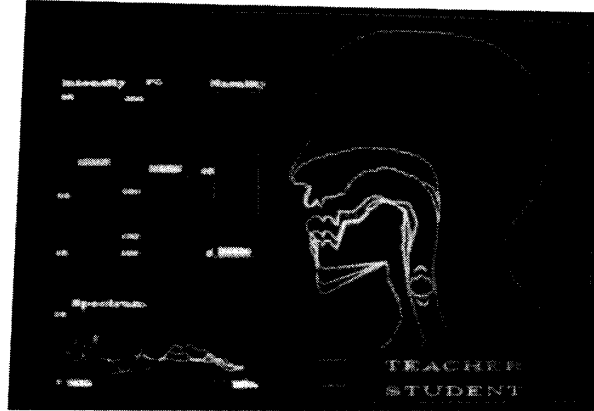


Fig. 13. Displayed parameters in training the phonation /ja/.

TABLE II
PROGRESS OF TRAINING FOR /ja/

| Subject | Date | Inten- sity | F0 | Vocal Tract Shapes | | Spectra | |
|---------|------|----------------|----|--------------------|-----|---------|-----|
| | | | | /j/ | /a/ | /j/ | /a/ |
| P P S | 1st | 25 | 50 | 0 | 25 | 0 | 0 |
| | 2nd | 50 | 0 | 90 | 30 | 90 | 30 |
| | 3rd | 22 | 0 | 78 | 56 | 56 | 22 |
| | 4th | 44 | 36 | 64 | 44 | 0 | 32 |
| | 5th | 71 | 64 | 100 | 43 | 64 | 50 |
| | 6th | 90 | 50 | 0 | 30 | 100 | 50 |
| | 7th | 87 | 53 | 0 | 40 | 100 | 80 |
| | 8th | 95 | 55 | 25 | 90 | 95 | 100 |
| | 9th | 75 | 44 | 100 | 44 | 100 | 95 |
| | 10th | 75 | 45 | 10 | 25 | 100 | 100 |
| | 11th | 90 | 50 | 10 | 35 | 95 | 100 |
| K S W | 1st | 0 | 20 | 80 | 20 | 0 | 0 |
| | 2nd | 70 | 30 | 100 | 30 | 100 | 40 |
| | 3rd | 44 | 0 | 0 | 78 | 78 | 11 |
| | 4th | 76 | 84 | 100 | 24 | 48 | 12 |
| | 5th | 86 | 57 | 100 | 0 | 100 | 79 |
| | 6th | 100 | 36 | 55 | 18 | 100 | 18 |
| | 7th | 80 | 53 | 0 | 100 | 93 | 20 |
| | 8th | 63 | 37 | 16 | 21 | 100 | 100 |
| | 9th | 88 | 50 | 100 | 13 | 100 | 100 |
| | 10th | 70 | 95 | 100 | 50 | 100 | 100 |
| | 11th | 80 | 60 | 50 | 45 | 95 | 100 |

TABLE III
PROGRESS OF TRAINING FOR /pa/

| Date | Achievement Rates | |
|------|-------------------|-------|
| | P P S | K S W |
| 1 st | 10 | 10 |
| 2 nd | 0 | 10 |
| 3 rd | 30 | 10 |
| 4 th | 50 | 10 |
| 5 th | 70 | 70 |

VI. CONCLUSION

This paper proposed an integrated speech training system with a well-designed self-training program. The proposed

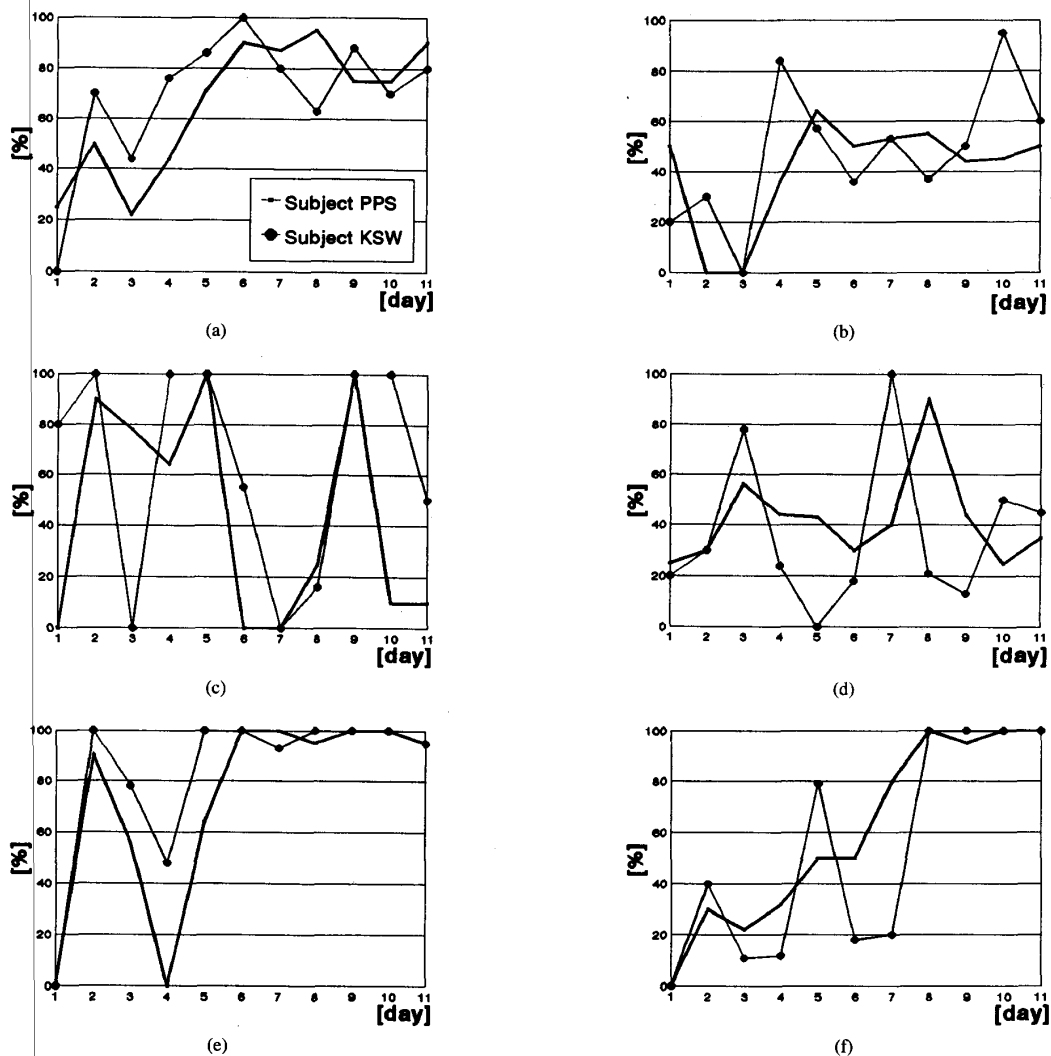


Fig. 14. Achievements of each parameter for /ja/. (a) Progress of intensity. (b) Progress of fundamental frequency. (c) Progress of vocal tract shape for /j/. (d) Progress of vocal tract shape for /a/. (e) Progress of spectra for /j/. (f) Progress of spectra for /a/.

system can display five speech parameters, as well as vocal tract shape, intensity, fundamental frequency, nasality, and frequency spectra. The reference and trainee's parameters are displayed simultaneously. This will be helpful for hearing impaired persons to recognize their errors. The self-training program in this system can teach hearing impaired persons how to correct these errors.

In developing this system, the obtained results are summarized as follows: First, by displaying vocal tract shapes, it is possible for hearing impaired people to easily recognize the difference between their shape and the reference. Second, displaying all the trainee's parameters and reference phonations in different colors can help hearing impaired people train efficiently by making efforts to match their color to that of the reference. Third, five speech parameters are displayed in a single system. This makes it possible for hearing impaired persons to observe the integrated features of their

own phonation. Fourth, for five Korean vowels, the estimated vocal tract shapes are considerably similar to the X-ray data. And finally, in training tests for hearing impaired children, it is known that all parameters improved simultaneously. Hearing impaired people frequently forget previous training. The proposed system can eliminate this problem. The hearing impaired children have mastered two syllables /ja/ and /pa/ in 5~6 days per each utterance. It is expected that the developed system can be a useful tool for hearing impaired people in speech training.

REFERENCES

- [1] C. D. Choi, "A review on development of visual speech display devices for hearing impaired children," *Commun. Disorders*, vol. 5, no. 2, pp. 38-44, 1982.
- [2] M. Shigenaga and H. Kubo, "Speech training systems for handicapped children using vocal tract lateral shapes," in *Proc. ICASSP*, 1986, pp. 637-640.

- [3] L. E. Bernstein, J. B. Ferguson, III and M. H. Goldstein, "Speech training devices for profoundly deaf children," in *Proc. ICASSP*, 1986, pp. 633-636.
- [4] S. Aguilera, A. Borrajo, J. M. Pardo and E. Munoz, "Speech-analysis-based devices for diagnosis and education of speech and hearing impaired people," in *Proc. ICASSP*, 1986, pp. 641-644.
- [5] P. Ladefoged, R. Harshman, L. Goldstein and L. Rice, "Generating vocal tract shapes from formant frequencies," *J. Acoust. Soc. Am.*, vol. 64, no. 4, pp. 1027-1035, Oct. 1978.
- [6] R. Harshman, P. Ladefoged and L. Goldstein, "Factor analysis of tongue shapes," *J. Acoust. Soc. Am.*, vol. 62, no. 3, pp. 693-707, Sept. 1977.
- [7] G. Fant, *Acoustic Theory of Speech Production*. The Hague, The Netherlands: Mouton, 1970.
- [8] H. Wakita, "Normalization of vowels by vocal-tract length and its application to vowel identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, no. 2, pp. 183-192, Apr. 1977.
- [9] R. L. Kirlin, "A posteriori estimation of vocal tract length," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 6, pp. 571-574, Dec. 1978.
- [10] H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-21, no. 5, pp. 417-427, Oct. 1973.
- [11] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 5, pp. 399-418, Oct. 1976.
- [12] J. J. Dubnowski, R. W. Schafer and L. R. Rabiner, "Real-time digital hardware pitch detector," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 1, pp. 2-8, Feb. 1976.
- [13] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
- [14] D. O'Shaughnessy, *Speech Communication—Human and Machine*. Reading, MA: Addison-Wesley, 1987.
- [15] T. Chiba and M. Kajiyama, *The Vowel, Its Nature and Structure*. Tokyo: Kaiseikan, 1941.
- [16] R. W. Schafer and L. R. Rabiner, "System for automatic formant analysis of voiced speech," *J. Acoust. Soc. Am.*, vol. 47, pp. 637-648, 1970.
- [17] J. P. Olive, "Automatic formant tracking by a Newton-Raphson technique," *J. Acoust. Soc. Am.*, vol. 50, pp. 661-670, 1971.
- [18] J. D. Markel, "Digital inverse filtering—A new tool for formant trajectory estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, no. 2, pp. 129-137, June 1972.
- [19] S. J. Orfanidis, *Optimum Signal Processing—An Introduction*. New York: Macmillan, 1985.
- [20] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [21] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. New York: Springer-Verlag, 1980.
- [22] R. L. Christensen, W. J. Strong and E. P. Palmer, "A comparison of three methods of extracting resonance information from predictor-coefficient coded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 1, pp. 8-14, Feb. 1976.
- [23] J. D. Markel, "Application of a digital inverse filter for automatic formant and F0 analysis," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, no. 3, June 1973.
- [24] R. G. Crichton and F. Fallside, "Linear prediction model of speech production with applications to deaf speech training," in *Proc. IEE Control & Science*, vol. 121, pp. 865-873, Aug. 1974.
- [25] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, 2nd ed. Berlin: Springer-Verlag, 1972.
- [26] S. Singh and K. S. Singh, *Phonetics—Principles and Practices*. Baltimore, MD: University Park Press, 1976.
- [27] C. Y. Suen and R. D. Mori, *Computer Analysis and Perception-Auditory Signals*. Boca Raton, FL: CRC Press, 1982.
- [28] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [29] M. E. Van Valkenburg, *Analog Filter Design*. New York: Holt-Saunders International Editions, 1982.
- [30] *PC System Board User Manual*, Loughborough Sound Images Ltd., TMS320C25, 1988.
- [31] *Digital Signal Processing Applications with the TMS 320 Family*, Texas Instruments, 1986.
- [32] I. R. Titze, "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.*, vol. 85, no. 4, pp. 1699-1707, Apr. 1989.
- [33] R. S. Nickerson and K. N. Stevens, "Teaching speech to the deaf: Can a computer help?," *IEEE Trans. Audio Electroacoust.*, vol. AU-2, no. 5, pp. 445-455, Oct. 1973.
- [34] D. J. Kim, S. P. Ahn, Y. S. Song, J. H. Lee, T. S. Yoon and S. H. Park, "A speech (vowels) articulatory motion estimating system," in *16th Int. Conf. on Medical and Biological Eng.*, Young Investigator's Award Symposium, Kyoto Japan, vol. 29, no. 1, p. 92, July 1991.
- [35] D. J. Kim, J. H. Lee, T. S. Yoon and S. H. Park, "Development of speech training aids for hearing impaired," in *Proc. Korean Automatic Control Conf.*, vol. 1, pp. 746-751, Oct. 1991.
- [36] S. H. Park, D. J. Kim, J. H. Lee and T. S. Yoon, "Development of speech training aids using vocal tract graphics," *Trans. Korean Inst. Elec. Eng.*, vol. 41, no. 2, pp. 209-216, Feb. 1992.



Sang H. Park (S'92-M'94) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1962, 1964, and 1971, respectively.

From 1964 to 1969, he was an Assistant Professor at Soodo Engineering College. He has been the Director of the Institute of Medical Instruments Technology at Yonsei University. Since 1970, he has been with Yonsei University, where he is currently a Professor in the Electrical Engineering Department. His research interests are currently biomedical signal and image processing, neural network, and rehabilitation engineering.

Dr. Park is a member of KOSOMBE, KIEE, KITE, KISS, and AKS.



Dong J. Kim received the B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1988, 1990, and 1994, respectively.

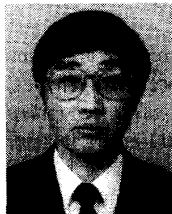
His research interests are currently speech signal and biomedical signal processing, speech training aids for hearing impaired, and rehabilitation engineering.

Dr. Kim is a member of KOSOMBE, KIEE, and ASK.



Jae H. Lee received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1986 and 1988, respectively. He is currently a Ph.D. candidate in electrical engineering at Yonsei University.

He has been with the Agency for Defense Development, Korea, since 1991. His research interests include auditory processing of speech, sound localization and aids for the hearing disabled.



Tae S. Yoon (M'90) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1978, 1980, and 1988, respectively.

From March 1980 until June 1983, he was an Instructor at the Second Naval Academy, Jinhae, Korea. Since March 1989, he has been an Assistant Professor at Changwon, Korea. He is currently a Visiting Assistant Professor at Vanderbilt University, Nashville, TN, USA. His research interests are biomedical and speech signal processing, signal modeling, and deconvolution.

Dr. Yoon is a member of KIEE and KITE.