대한음성언어의학회지
제 7 권 제 1 호 1996

청각장애자를 위한 원격조음훈련시스템의 개발

연세대학교 전기공학과, 의용공학과*
이재혁 · 유선국* · 박상희

= Abstract =

Remote Articulation Training System for the Deafs

Jae-Hyuk Lee, M.D., Sun-Kook Yoo Ph.D.,* Sang-Hui Park, Ph.D.

Department of Electrical Engineering, Department of Biomeical Engineering, Yonsei University, Seoul, Korea*

In this study, remote articulation training system which connects the hearing disabled trainee and the speech therapist via B-ISDN is introduced. The hearing disabled does not have the hearing feedback of his own pronuciation, and the chance of watching his speech organs' movement trajectory will offer him the self-training of articulation. So the system has two purposes of self articulation training and trainer's on-line checking in remote place.

We estimate the vocal tract articulatory movements from the speech signal using inverse modelling and display the movement trajectoy on the sideview of human face graphically. The trajectories of trainees' articulation is displayed along with the reference trajectories, so the trainee can control his articulating to make the two trajectories overlapped. For on-line communication and ckecking training record, the system has the function of video conferencing and tranferring articulatory data.

KEY WORDS : Remote training · Articulation trajectory estimation.

서 론

최근의 신호처리기술과 컴퓨터기술의 발전에 힘입어 음성발성시의 조음궤적을 추정하고자 하는 연구가 활발히 전개되고 있다[1)2)]. 청각장애자의 경우 auditory feedback을 갖지 못하여 발음방법을 습득하기가 어려운데 거울이나 교사의 입모양을 관찰하는 것은 입술과 혀의 앞부분의 모습 등을 볼 수 있을 뿐이어서 조음기관의 움직임을 제한적으로밖에는 관찰하지 못하고 나머지 조음기관의 움직임은 설명이나 그림을 통해서 밖에 알 수 없다. 더우기 여러 조음기관이 동시에 각기 움직이는 협응과정은 더더욱 알기 어려운 사정이다. 조음운동을 디지탈 컴퓨터를 이용하여 추정하려는 연구는 바로 이러한 제약을 극복하고자 하는 시도이다.

청각장애자를 위한 음성발음훈련이나 진단은 종합병원이나 언어치료시설, 특수학교 등에서 이루어진다. 이것은 검사와 훈련진도의 확인 등을 모두 언어치료사 등의 전문가에게 의존해야 하기 때문이다. 만일 컴퓨터의 도움으로 평소에 가정에서 자가훈련을 계속하다가 약속한 시간에 언어치료사와 네트웍으로 접속되어 훈련진도를 확인받고 다음 훈련절차 등을 지시받을 수 있다면 보다 향상된 훈련효과를 거둘 수 있을 것이고 바로 이점이 본 연구의 목표이다. 즉 컴퓨터를 이용하여 훈련자에게

자가훈련의 기회를 부여하고 훈련자와 언어치료사간의 원격훈련을 가능케 하는 것이다.

본 시스템은 3개의 부분으로 나눌 수 있다. 첫째 음성신호로부터 성도궤적을 추정하고, 이를 얼굴횡단면 위에 그래픽으로 표현하며, 둘째 원격훈련을 가능케 하기 위해 서로의 영상과 음성 및 조음궤적데이타를 처리하며, 셋째 네트웍접속을 위한 인터페이스를 갖춘다. 성도궤적의 추정은 선형예측필터와 포먼트정보를 조합하여 사용한다. 영상회의를 위해서는 H.261 및 G.711 표준을 사용하였다. 또한 시스템은 초고속통신망의 선도시험망에 접속되는 인터페이스 기능을 갖는다.

조음운동의 추정

조음운동의 추정은 음향신호 즉, 마이크로폰으로 입력받은 음성신호로부터 그 음성신호가 생겨나게 된 조음운동을 역으로 추측하는 것을 의미한다. 이때 성문에서부터 입술까지의 성도(vocal tract)의 변화를 각 지역에서의 단면적의 변화로 나타내기 때문에 결국 조음운동의 추정은 성도면적의 형태를 띄게 된다. 본 연구에서는 선형예측필터-형성음 방식의 성도궤적추정(Linear Prediction Filter-Formant based Vocal Tract Profile Estimation)을 사용한다[3].

1. 조음궤적의 추정

성도필터를 all-pole model로, 성문여기신호는 pulse열이나 랜덤노이즈로 단순화하였다. 또한 성도필터와 성문필터를 분리가능한 선형결합으로 가정한다[1)3)]. 선형예측계수로부터 반사계수를 얻어 성도 15개 각 부분의 상대면적을 구한다. 또한 개방경계의 턱과 입술의 opening을 포먼트주파수로부터 구하여 인간의 성도구조위에 매핑시켜 그래픽으로 표현한다. 이 그래픽은 발음이 진행됨에 따라 계속 새로 그려져 결국 animated된 trajectory를 형성하게 된다.

2. 조음도 그래픽의 설계

조음도(vocal tract profile)는 발성할 때 각 조음 기관의 모습을 성대에서 입술까지, 즉 성도(vocal tract)를 따라 단면도의 형태로 나타낸 것이다. 음성 신호로부터 각 조음 기관의 조음 형태를 그래픽으로 복원하기 위해서는 우선 각 조음 기관의 운동 범위를 정의해야 한다. 본 연구에서는 조음기관의 생리학적 운동범위와 X-선 촬영 데이터에 기반하여 위턱에 설정된 조음점은 고정되어 있고 아래턱에 연결된 조음체는 변화되는 운동축을 설정하였다[3)4)].

성도는 길이 17cm의 성인남성을 기준으로 입술과 치아에 2개 section, 치아에서 설근까지(연구개(velum)까지) 11개, 연구개에서 성문까지 2개로 나누어 총 15개의 section으로 나누었다. 각 축의 기울기는 조음점에 수직이 되도록 하였고 앞뒤축의 기울기와 가능한 한 평행을 유지하도록 하였다.

또한 치경으로부터 연구개까지의 거리와 연구개에서 성문까지의 거리가 해부학적으로 거의 같지만 연구개에서 성문까지는 그 조음 거리가 혀의 움직임에 직접적인 영향을 받고, 화자 스스로 그 부분의 조음 거리를 직접 제어하기 어려울뿐더러, 또 그래픽 표현상 aspect ratio가 2.1 : 1인 점을 감안하여 두 부분의 가중치를 2 : 1로 주어 구성하였다. Fig. 1은 완성된 조음도 그래픽을 나타낸 것이다.

추정된 조음 거리를 각각의 축위에 사상(mapping)함에 있어 조음점을 고정점으로 보았다. 즉 위턱뼈는 고정되어 윗치아, 치경, 경구개, 연구개는 움직임이 없다고 보고 고정 좌표를 할당하였다. 또한 설근에서 성문까지의 성도의 뒷벽, 즉 연구개에서 성문에 이르는 목의 뒷부분 역시 고정되었다고 가정하였다. 따라서 각 축 위의 고정된 조음점에서부터 시작하여 해당 조음 거리에 해당하는 점이 조음체의 좌표가 된다.

3. 영상 및 음성의 처리

네트웍의 전송을 통하여 훈련자의 발음 및 조음도 그래픽이 의사에게 전달되지만 조음도 그래픽이 얼굴의 횡단면을 기준으로 조음체의 운동 궤적을 볼 수 있게 설계되

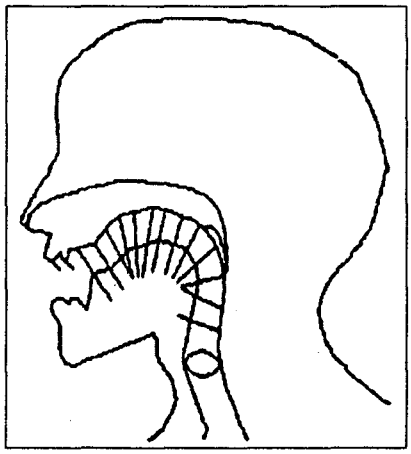


Fig. 1. Motion axses of vocal tract graphics.

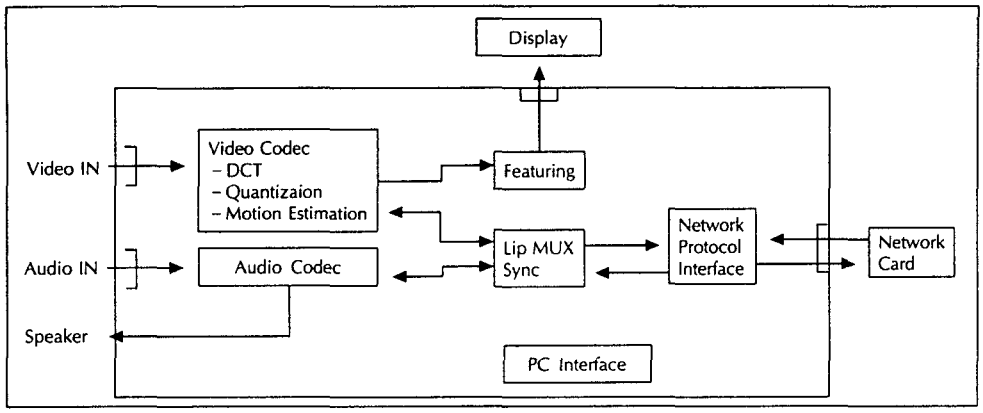


Fig. 2. Diagram of video conferencing coder.

어 입술의 내밀음(protrusion)이나 양입술의 벌어짐(opening)은 표현되나 정면에서 보는 입술의 오므림(rounding)은 표현할 수 없다. 따라서 훈련자의 입술 모습을 입체적으로 파악하기 위해서는 훈련자 얼굴의 동영상을 볼 필요가 있다. 또한 자택진료를 위해서는 의사와 훈련자(청각장애자)간의 대화가 어떠한 수단으로든 이루어져야 하는데 음에 대한 듣기 및 말하기가 불가능한 청각장애자의 경우 일반 음성 대화의 통신은 어려운 경우가 많아 수화(signed language)에 의한 의사 교환이 이루어져야 한다. 따라서 정면에서 본 훈련자 입술 영상과 수화를 위한 영상을 전송하기 위하여 동영상 압축기법을 적용한 영상전화기를 시스템에 통합하여 구성하여야 한다.

종합정보 통신망에서의 영상전화 영상회의 등과 같은 서비스의 수요에 부응하여 CCITT에서는 64 kbps에서 2 Mbps의 통신망에서의 영상전화에 대한 표준안 H.261을 제안하였다. 따라서 청각장애자 자택진료 시스템에서도 동영상의 압축은 컴퓨터에 의한 멀티미디어 시스템의 호환성과 확장성을 고려하여 Fig. 2와 같은 Audio/Video Codec을 사용하였다.

Fig. 2의 코덱보드는 영상을 H.261규약에 맞추어 64 Kbps에서 2Mbps정도의 비트발생률로 영상을 실시간 압축 복원하며, 음성은 G.722, G.711 등의 규약을 기반으로 실시간 부호화하여, Ethernet-ATM switch를 경유하여 초고속 통신망을 통해 상대방과 데이터를 전송하고, 수신한 데이터를 실시간으로 부호화한다.

시스템 구성

본 연구의 목적인 발음 훈련의 자택진료를 위해서는 음

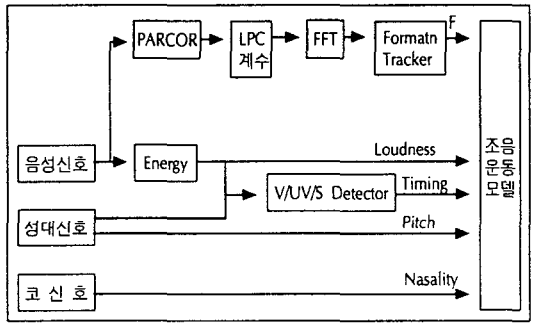


Fig. 3. Block diagram of articulation estimation.

성 분석, 조음도 훈련, 음성 및 동영상 압축, network interface 등의 각 부분이 하나로 결합되어야 한다.

1. 음성신호처리 기술

음성분석을 위한 기본 방법은 all-pole 디지털 역필터인 LPC(Linear Predictive Coefficient) 방법을 기초로 하여 Fig. 4와 같이 음성훈련에 적합한 스펙트럼, 5개의 포먼트 주파수, 피치, timing, 음의 세기, 리듬, 비음, 피치 정보를 활용한다. LPC방법의 해는 정규화된 음성신호의 Yule-Walker 방정식으로부터 구한다.

2. 조음상태의 추적 및 그래픽 디스플레이

음성신호를 분석하여 얻은 매개변수를 inverse modeling을 통과시킴으로써 발음 당시의 조음 운동 즉 혀, 턱, 치아의 운동궤적을 추정할 수 있다. 본 연구에서는 Fig. 4와 같이 LPC 반사계수와 스펙트럼 상의 포먼트 주파수 정보를 이용하여 단음 및 단어의 조음상의 운동궤적을 실시간으로 추정한다. 이의 표현을 위해 얼굴의 횡단면을 그래픽화 하고, 이것의 위에 조음 운동체 즉 혀, 치아, 턱의 추정된 궤적을 에니메이션화 하여 디스플

레이 한다.이러한 조음도 그래픽으로써 기존의 발음훈련으로는 교정하기 어려운 입 속의 조음운동을 교정할 수 있게 해준다.

3. 전체 시스템의 결합

전체 시스템은 Fig. 5와 같다. 훈련자측의 조음 거리 파라미터와 압축된 동영상/음성이 결합된 bit stream, 문자 다이얼로그 등은 control software에 의해 network을 통해 의사 측으로 전달된다. 또 마찬가지로 의사 측의 동영상/음성/문자 다이얼로그도 훈련자측으로 전달된다. 개발된 전체 시스템의 특징은 다음과 같다.

- 운영체계 : Windows 95
- 개발도구 : C 4.5 compiler
  Windows API
- 운영 환경 :
  Pentium 120MHz, 16M memory
- 음성분석 :
  10KHz sampling 12 bit quantization
  LPC (Linear Predictive Coding) 15차

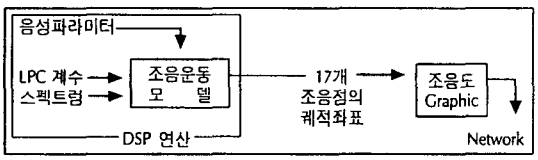


Fig. 4. Calcuration and display of vocal tract graphics.

- 조음도 graphic data :
  17개 조음점의 궤적 좌표
  데이터 전송율 10.8Kbps
- DSP : TMS 320C30
  32 bit floating point processor
  40 MFLOPS, 20MIPS
- 영상 압축 : H.261
  CIF규격, 64K-768Kbps
- 음성 압축 : G711
- Dialogue : 리턴 키에 의한 일괄 전송방식
- Input Device
  · CCD Camera : Thosiba IK-M28
  · Mikerophone : Condensor Type
  · Accelerometer : Mutura PKS1
- Network 접속 환경
  · ATM switch의 LAN Interface
  · PC/TCP protocol
  · Ethernet / 10baseT

결 과

무반향의 녹음실에서 4명의 남성화자가 발음한 5개의 한국어 모음 /아,에,이,오,우/에 대한 전체 발음의 조음도를 Fig. 6에 나타내었다. 이 결과는 추출된 성도면적

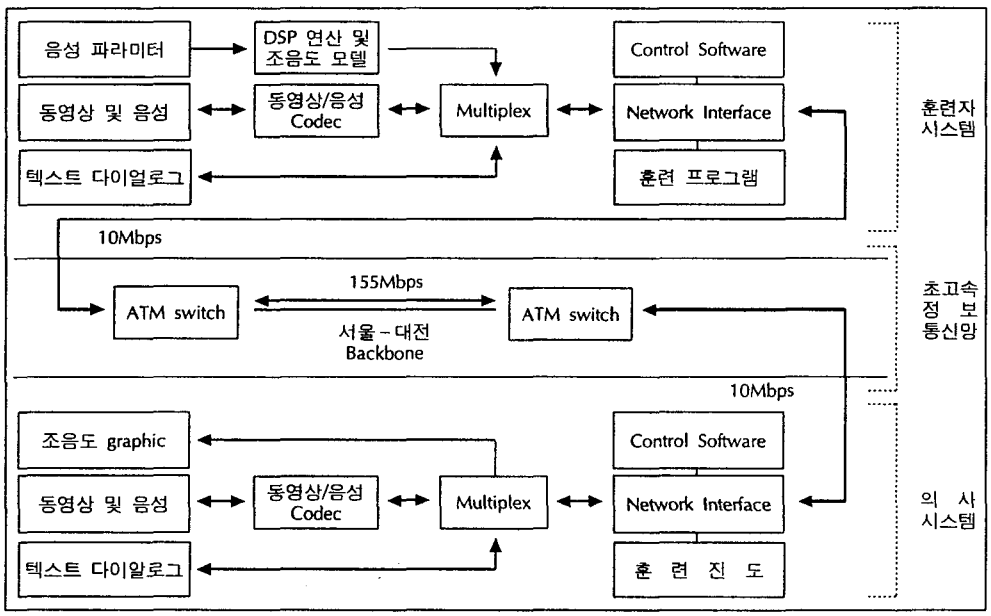


Fig. 5. Overall system configuration.

과 포먼트주파수로부터 조음거리를 계산하여 추출된 것인데 실제 조음기관의 움직임과 유사성을 보인다[4].

Fig. 7은 조음훈련을 위한 화면구성을 보여준다. 화면구성중 좌측의 네 개의 윈도우는 각기 intensity, F0 (pitch), Nasality, spectrum을 나타낸다. 이중 spectrum은 사실상 피훈련자의 조음훈련을 위한 것이라기보다는 개발자들의 성능확인을 위한 것이며, 나머지 intensity, pitch, nasality는 각기 피훈련자가 발성의 크기와 음조, 비음을 확인할 수 있는 형태를 띄고 있다. 우측은 조음도로서 발음시의 조음운동궤적을 보여준다. 훈련프로그램에 의해 훈련을 받을 때에는 컴퓨터에서 지정하는 발음을 하게 되는데 이때 화면에는 정상발음의 궤적이 다른 색으로 겹쳐 그려지게 되어 발음의 오류를 교정할 수 있게 되어 있다. Fig. 7의 (b)는 /아/발음이 지정되었을 때 정상적으로 /아/ 발음을 한 경우로 컴퓨터의 궤적과 훈련자의 궤적이 거의 일치하지만 (a)는 /아/가 지정되었을 때 훈련자가 /오/에 가까운 발음을 한 경우로 궤적의 차이가 크게 드러난다. 훈련자는 궤적이 서로 차이가 날 때 자신의 조음운동을 제어하여 컴퓨터의 궤적을 따라감으로써 발음을 교정하게 된다.

Fig. 8에서부터 Fig. 13까지는 완성된 시스템의 자가훈련 및 원격훈련의 모습이다. Fig. 8 및 Fig. 9는 단독

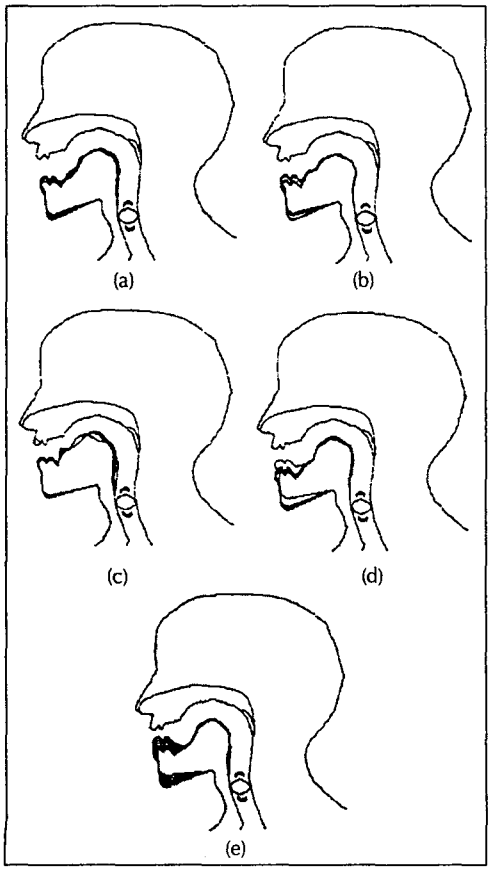
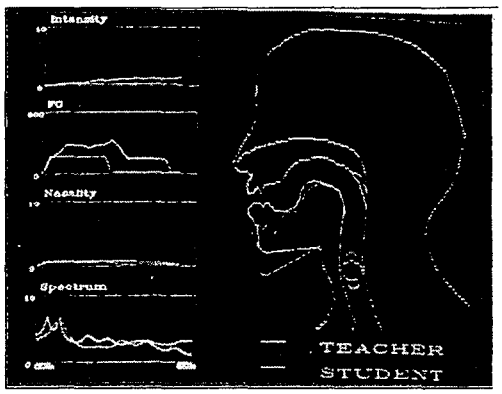
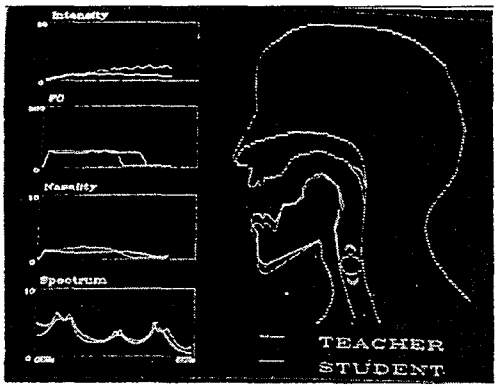


Fig. 6. Estimated vocal tract graphics for 5 vowels.
(a) /아/ (b) /에/ (c) /이/ (d) /오/ (e) /우/



(a)



(b)

Fig. 7. Display screen of training(a,b).

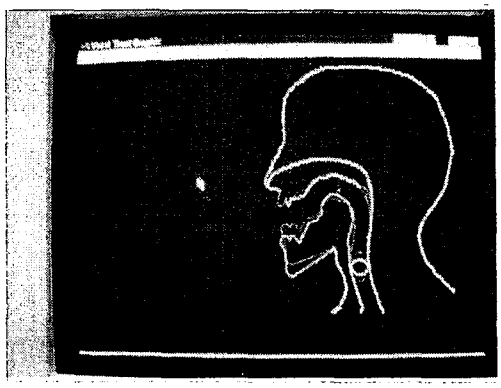


Fig. 8. Vocal tract graphics.

훈련에서의 화면모습이다. Fig. 10은 단독훈련에서의 모음선택절차이며 Fig. 11은 수화를 보조하는 다이알로그 윈도우이다. Fig. 12는 환자와 언어치료사 간의 영상회의 모습으로 좌측상단의 작은 윈도우가 본인의 모습, 큰 윈도우가 상대방의 모습이다. Fig. 13은 최종적으로 모든 기능이 활성화된 화면모습으로 영상회의, Vocal tract graphics, 다이알로그의 모습이 보인다.

Fig. 9. Self training.

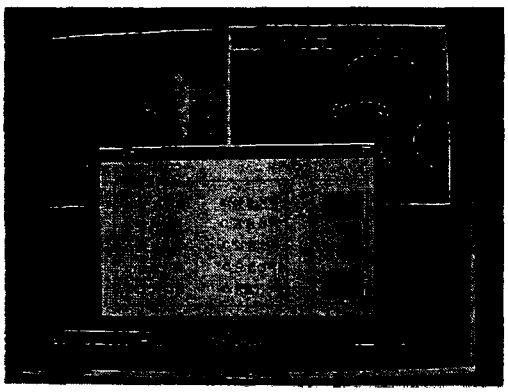


Fig. 10. Vowel selection.

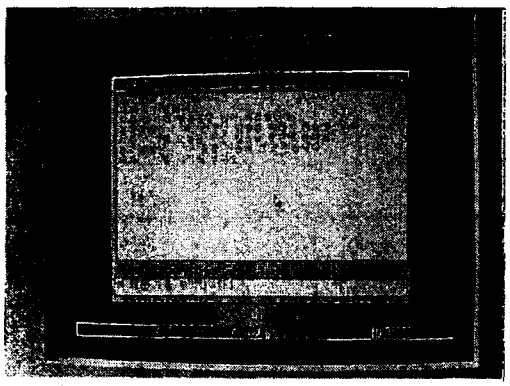


Fig. 11. Dialogue window.

결 론

본 연구에서는 소리의 궤환이 안되어 자신의 발음을 제어하기 어려운 청각장애자들의 발음훈련의 효과를 극대화시킬 수 있는 원격진료 시스템의 개발을 수행하였다.

개발된 시스템은 음성신호, 비음신호, 성대신호를 전처리하여 음성 분석을 수행하고 조음모델을 계산하는 전처리부, DSP 연산부와, 구강내의 조음운동 궤적을 애니메이션화 하여 디스플레이 하는 조음도 그래픽, 의사와의 원격진료를 위한 동영상 및 음성 압축/복원, 네트웍 인터페이스 등의 주요 기능을 가지며 수화 통신을 보조하는 다이얼로그 기능, 영상의 저장, 재생, 반복기능, 훈련 프로그램 등을 보조기능으로 갖는다.

개발된 시스템을 이용하여 청각장애자는 평상시 스스로 훈련할 수 있으며, 정기적으로 병원의 언어치료사와

Fig. 12. Video conferencing.

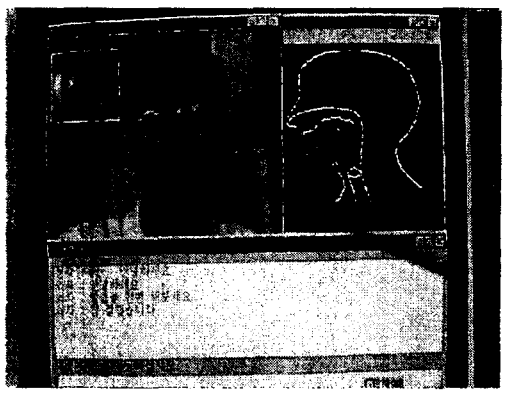


Fig. 13. Total display of all functions.

연결되어 발음 훈련 진도를 확인 받고 틀린 곳을 교정 받을 수 있다. 언어치료사는 원격지의 훈련자의 훈련 정도를 동영상과 음성, 그리고 조음도 그래픽을 통하여 현장에 있는 것처럼 진단하고 교정할 수 있어 결론적으로 자택 진료가 가능하다 하겠다.

본 연구를 통하여 얻을 수 있는 직접적인 효과로 첫째, 청각장애자의 발음조절능력을 증대시켜 언어생활의 폭을 넓힘으로써 정상인과의 공동생활의 기회를 넓히며 둘째, 병원의 언어치료실이나 재활원으로의 통원치료의 횟수를 대폭 감소시켜 장애자와 의사 양쪽의 치료부담을 경감시킬 수 있다. 또한 본 연구에서 자택진료를 위해 채택한 영상전화 및 다이알로그 기능, 문서전송기능은 청각장애자들 사이의 또는 청각장애자와 정상인 사이의 통신수단으로 바로 사용할 수 있다.

References

1) Rahim MG, Goodyear CC, Kleijin WB, Schroeter J and Sondhi M : *On the use of Neural Network in Articulatory Speech Synthesis. Journal of Acoustic Society of America 93(2) : 1109-1121, 1993*

2) Guenther F : *A Neural Network Model of Speech Acquisition and Motor Equivalent Speech Production. Biological Cybernetics 72 : 43-53, 1994*

3) 박상희 · 김동준 · 이재혁 · 윤태성 : 조음도를 이용한 발음훈련기기의 개발. 대한전기학회논문지 *41(2) : 209-216. 1992*

4) Ladefoged P, Harshman R, Goldstein L and Rice L : *Generating Vocal Tract Shapes from Formant Frequencies. Journal of Acoustic Society of America 64(4) : 1027-1035, 1978*

5) Park SH, Kim DJ, Lee JH, and Yoon TS : *Integrated Speech Training System for the Hearing Impaired. Institute of Electrical and Electronic Engineering transactions on Rehabilitation 2(4) : 189-196, 1994*