웹 기반의 화자확인시스템 설계에 관한 연구

A Study on the Design of Web-based Speaker Verification System

이 재 희*, 강 철 호*
(Jae Hee Lee*, Chul Ho Kang*)

요 약

본 연구에서는 인터넷 웹 기반의 화자확인시스템을 설계하였다. 웹 기반의 화자확인 시스템에 적용할 화자인식기법을 선정하기 위해 문자종속 화자인식기법들(DTW, DHMM, SCHMM)의 성능 및 특징들을 컴퓨터 시뮬레이션을 통하여 비교 평가하였다. 컴퓨터 시뮬레이션결과를 이용하여 웹 기반의 화자확인시스템에 적합한 인식성능 및 초기 학습발음수를 갖는 DHMM을 화자인식기법으로 선정하고 이를 분산처리환경에서 동작하도록 ActiveX, DCOM기술을 이용하여 3계층방식으로 설계하였다.

핵심용어: 웹 기반의 화자확인, 분산처리, 디컴, 액티브엑스, 문장종속, 컴포넌트, 이산 HMM
투고분야: 음성처리분야(2.6)

ABSTRACT

In this paper, the web-based speaker verification system is designed. To decide the speaker recognition algorithm applied to the web-based speaker verification system, the recognition performance and special features of the text-dependent speaker recognition algorithms(DTW, DHMM, SCHMM) are compared through the computer simulation. Using the results of computer simulation, select DHMM as speaker recognition algorithm at web-based speaker verification system because DHMM has the proper recognition performance and initial training utterance number. And by the three-tier method using the ActiveX, DCOM techniques web-based speaker verification system is designed to be operated in the distributed processing environment.

Key words: Wed-based speaker verification, Distributed processing, DCOM, ActiveX, Text dependent, Component, DHMM.

I. 서 론

컴퓨터의 발달과 함께 사회는 점점 복잡해지고 급속한 정보의 교류가 필요로 하게 되었다. 특히 인터넷의 등장으로 인해 모든 정보통신 기술이 인터넷과 공존해야만이 그가치를 인정받는 사회가 되어가고 있다. 최근에는 인터넷을 이용하여 정보를 제공하는 사이트(Site)에서는 특정인에게 양질의 정보를 제공하기 위해 사이트를 회원제로 운영하여 회원에게만 정보를 제공하는 사이트가 늘고 있는 추세이다. 회원제로 운영되는 사이트의 경우 회원번호(Member ID)와 패스워드(Password)을 입력한후 사이트를 이용할 수 있다. 그러나 패스워드는 쉽게 잊혀지거나 해킹(Hacking)등에 의해 타인에게 알려질 수 있어 타인에 의한 도용에 취약한 실정이다. 또한 최근에는 한 개인이 3~5개정도의 회원번호와 패스워드를 보유하고 있다고 한다. 그 결과 패스워드의 망각 또는 혼동으로 인해 사이트에 로그인할 수 없어 회원을 재등록하거나 관리자에게 패스워드를 문의하는 경우가 종종 있고 이것으로 인해 사이트이용의 활성화를 저해하는 요인으로 동작하고 있다. 이러한 문제를 해결하기 위해서는 패스워드에 의한 사용자 확인방식보다는 각 개인이 가지고 있는 고유한 특성(지문, 홍채, 음성)을 이용하는 것이 신뢰 및 편리성 면에서 유리하다[1,2]. 인간의 고유한 특성중에서 기본적인 컴퓨터환경이외에 부가적인 장비가 설치될 필요없이 각 개인을 확인할 수 있는 것은 각 개인의 음성특성을 이용한 기술이 가장 적합하다고 할 수 있다. 웹 기반에서 동작하는 화자확인시스템을 개발하기 위해서는 우수한 화자인식율뿐만 아니라 고속수행이 가능한 화자인식기법을 선택해야만 한다. 또한 학습이 온라인(On-line)상에서 이루어지는 현실성을 고려하여 최단시간의 학습이 요구되는 화자인식 기법을 선정해야만 한다. 본 연구에서는 대표적인 문자종속 화자인식기법들 중에서 웹 기반의 화자확인시스템 구현시 최적의 화자인식기법을 선정하기 위하여 대표적인 화자인식기법(DTW, DHMM,SCHMM)들의 성능들을 컴퓨터 시뮬레이션(Computer simulation)을 통해

* 광운대학교 전자통신공학과
접수일자 : 2000년 2월 15일

비교검토하였고 웹 기반의 화자확인시스템을 분산처리환경에서 동작하도록 ActiveX컨트롤기술[3,4]과 DCOM (Distribute Component Object Module)기술[5]을 이용하여 설계하였다.

II. 화자확인 시스템의 개요 및 고려사항

2.1. 화자확인시스템 개요

화자인식은 인식대상에 따라 화자확인(Speaker verification)과 화자식별(Speaker identification)로 크게 나눌 수 있다. 화자확인은 입력된 음성이 본인의 것인지의 여부를 판정하는데 비해, 화자 식별의 경우는 입력된 미지의 음성이 이미 등록된 여러명의 화자 중 어떤 화자에 의해 발성된 음성인지를 판정하는 것을 말한다. 화자가 발성할 문장이 고정되어 있는 화자인식시스템을 문장 종속형(Text-dependent)이라 하고, 문장이 정해져 있지 않고 자유롭게 발음하는 경우를 문장 독립형(Text-independent)이라 한다. 문장종속형은 사전에 녹취한 음성정보를 이용하여 화자인식시스템에 손쉽게 침범할 수 있다는 단점이 있다. 이러한 단점을 보완하기 위해 본 연구의 웹 기반의 화자확인시스템은 문장지시형(Text-prompted) 화자인식방법으로 설계하였다.

2.2. 화자확인시스템의 구성

화자확인시스템의 전반적인 구성은 그림 1과 같다

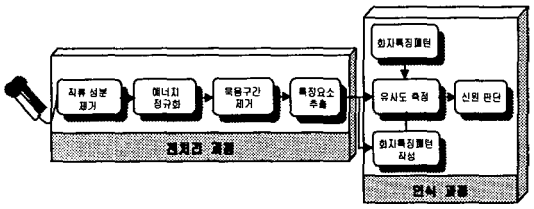


그림 1. 화자확인시스템의 구성도.
Fig. 1. Speaker verification system configuration.

전처리과정에서는 어떤 음성 특징요소(Feature parameter)를 선택하느냐 하는 것이 문제이다. 음성신호에는 여러가지 정보와 잡음이 섞여 있다. 이 가운데는 발음한 화자의 신원에 관한 정보도 포함되어 있고, 그 외의 정보도 있다. 화자 인식을 위한 관점에서 바라볼 때, 화자의 신원에 관한 정보 이외에 다른 모든 정보는 일종의 잡음이다. 이러한 잡음은 인식율 저하의 원인이 된다. 그러므로 화자신원에 관한 정보만 포함하고 그 이외의 모든 정보는 억압한 음성 특징요소를 전처리과정에서 추출해야 한다. 그러나 화자의신원에 관한 정보만 포함하는 음성 특징요소를 추출한다는 것은 현실적으로 불가능하다. 현재까지 화자인식에 사용하고 있는 음성 특징요소들에는 음성에 관한 정보도 많이 포함하고 있다. 그러므로, 이러한 특징요소들의 성능비교에 관한 연구가 수행되었다. 지금까지의 연구결과 기본주파수와 선형예측계수가 화자인식에 유용하며[6] 선형예측계수를 변환한 켑스트럼(Cepstrum)계수가 특히 우수한 것으로 알려져 있다[7]. 화자인식시스템의 인식과정에 수행되는 인식방법으로는 크게 벡터 양자화(VQ), 동적정합법(Dynamic Time Warping : DTW)과 같은 패턴정합을 이용하는 방법과 HMM(Hidden Markov Model)을 이용한 확률적인 방법 그리고 인간의 두뇌를 모델링한 신경회로망(Neural Network)을 이용한 방법등이 있다[8]. 이중에서 문장종속화자확인시스템에서 사용되는 대표적인 인식방법으로는 DTW, HMM이 사용된다[9].

2.3. 인터넷 웹 관련기술

2.3.1. 웹 프로그래밍 기술

웹상에서 클라이언트측 즉 웹브라우저상에서 프로그래밍이 가능한 현존하는 대표적인 기술로는 스크립트언어(JavaScript, VBscript)와 자바 애플릿(Java Applet) 그리고 ActiveX 컨트롤(Control)등이 있다. 스크립트언어의 경우 프로그램모듈의 소스파일(Source file)이 전송되므로 전송시간은 다른 기술에 비해 빠르지만 클라이언트측에서 컴파일하여 수행하므로 실행시간이 오래 걸리고 수행할 수 있는 기능에 한계가 있고 클라이언트가 사용하는 브라우저의 종류에 따라 다르게 동작한다. 다음 방문시 같은 다운로드 및 컴파일과정을 반복해야 한다[10]. 자바 애플릿은 프리컴파일(Precompile)한 바이트코드(Byte code)형태로 서버로부터 전송되므로 스크립트언어보다는 전송시간이 오래 걸리나 전송된 바이트코드을 웹 브라우저에 내장된 JVM(Java Virtual Machine)에 의하여 재컴파일되어 바이너리코드(Binary code)를 생성한다[11]. 자바 애플릿은 스크립트언어와 달리 복잡한 기능을 수행할 수 있고 운영시스템(Operating System)이 다른 클라이언트환경에서도 동작할 수 있다는 장점이 있으나 다음 방문시 스크립트언어와 같이 다운로드 및 컴파일과정을 반복해야 한다. ActiveX기술은 바이너리코드(Binary code)형태의 실행파일(.exe) 이나 DLL(Dynamic Linked Library)화일을 서버로부터 다운로드하므로 다른 기술에 비하여 전송시간이 오래 걸린다.

표 1. 웹 프로그래밍기술들의 비교
Table 1. The comparision of web programming techniques.

특성 \ 기술	Script 언어	Java Applet	ActiveX
전 송 시 간	빠름	보통	늦음
수 행 속 도	낮음	보통	빠름
N회 방문시	N회 다운로드	N회 다운로드	1회 다운로드
운영 시스템	독립적	독립적	의존적
브라우저종류	의존적	독립적	독립적

그러나 한번 다운로드가 된후에는 컴파일과정없이 수행되므로 수행속도는 빠르다. 또한 클라이언트가 MS사의 운영시스템(Window98, WindowNT)을 사용하는경우 클라이언트측에 내장되어 있는 다양한 함수 및 컴포넌트를 활용할 수 있어 복잡한 기능을 손쉽게 구현할 수 있고, 다음

방문시 반복해서 다운로드 할 필요없이 예전에 다운로드 되어 시스템 레지스트리(System registry)에 등록된 ActiveX 컨트롤을 검색하여 실행시키므로 다음 방문부터는 모든 면에서 성능이 우수하다. ActiveX기술의 단점은 실행파일을 전송하므로 클라이언트측 운영시스템에 의존적이다. 즉 클라이언트 운영시스템체계와 ActiveX컨트롤을 개발한 운영시스템체계가 같아야만 ActiveX 컨트롤이 동작한다.

2.3.2. 분산처리기술(Distributed processing technique)

현재 가장 보편적으로 사용되는 분산처리기술은 OMG-(Object Management Group)의 CORBA(Common Object Request Broker Architecture)[12]와 MS사의 DCOM기술이 있다. 본 연구에서는 두 기술의 특성을 비교한 자료[13]을 검토하여 개발환경구축의 용이성 및 클라이언트측의 ActiveX 컨드롤기술과의 양립성을 고려하여 DCOM기술을 선택하였다. 그림 2는 DCOM의 기본 구조도이다.

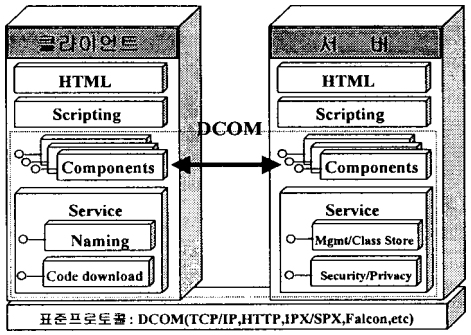


그림 2. DCOM의 기본 구조도
Fig. 2. The base configuration of DCOM.

2.4. 웹 기반 화자확인시스템 설계시 고려사항

웹 사이트에서 사이트이용회원들의 신분확인을 위한 화자확인시스템 설계시 고려해야할 사항은 다음과 같다.

첫째 웹 기반의 화자확인시스템에 적용할 적절한 화자 인식가법을 선정해야 한다.

둘째 회원등록시 회원의 음성을 학습시키기 위한 단어당 발음횟수가 적으면서도 화자인식률이 높아야 한다. 음성정보등록을 위한 학습과정이 온라인(On-line)상에서 이루어지므로 발음횟수가 많으면 회원등록시 학습과정의 불편함으로 회원의 음성정보등록을 기피할 우려가 있다.

셋째 기존의 패스워드의 단점인 패스워드망각으로 인한 불편함을 제거하기 위해 등록한 단어를 사용자가 기억할 필요가 없어야 한다.

넷째 화자인식율이 시간경과에 따라 감소되지 않아야 한다. 회원등록시 학습하여 생성한 기준패턴과 시간경과로 인한 회원의 음성정보변화로 인한 차이로 화자인식율이 저하되지 않아야 한다.

다섯째 인터넷상에서는 동시 사용자수를 예측할 수가 없다. 그러므로 동시에 많은 사용자가 화자확인서스템에 접속하였을 경우에도 시스템의 부하가 한 시스템에 집중되는 것을 방지하기 위해 분산 설계되어야 한다.

III. 화자확인시스템을 위한 화자인식기법의 성능평가

웹 기반의 화자확인시스템 설계시 중요한 사항은 웹 환경에 적합한 화자인식기법을 선정하는 것이다. 기존의 자료들은 대부분 화자인식기법의 인식율에만 초점을 맞추었다. 그러나 인터넷상에서 동작하는 화자확인시스템에서는 인식률 이외에도 고려해야할 여러 가지 사항들이 있다. 위에서 언급한 인터넷상에서 동작하는 화자확인시스템 설계시 고려사항을 반영한 최적의 화자인식기법 및 학습 발음수를 찾아내기 위하여 기존의 대표적인 문장종속의 화자확인기법들을 학습발음수별로 컴퓨터 시뮬레이션을 통해 화자인식률을 비교 평가하였다. 또한 화자인식기법의 실행속도를 평가하기위해 각 기법의 인식 및 학습과정의 계산량을 근사적으로 구하였다. 실험에서 사용된 단어는 5개월간에 걸쳐 7명의 남성화자에 의해 30번, 또 다른5명의 남성화자에 의해 30번씩 발음된 단어 "안녕하세요"로써, 7명의 화자는 시스템에서의 회원으로써 쓰여졌고, 또 다른5명의 화자는 사칭자(impostor)로써 사용되었다. 컴퓨터 시뮬레이션은 7명의 학습발음수(2~30번)에 따라 3가지 화자인식기법(DTW, DHMM, SCHMM)별로 각각의 화자기준모델(Speaker reference model)을 만들었고, 7명의 회원발음 30번에 대해 회원들간의 거부률(False rejection)실험과 허용률(False acceptance)실험을 하였다. 그리고 5명의 사칭자 발음의 허용률(False acceptance)실험을 행한후 각 실험의 인식률의 평균을 구하여 각 화자인식기법의 인식률로 나타내었다. 컴퓨터시뮬레이션에 사용된 음성 데이터는 11.025Khz로 샘플링한 후 끝점검출(End-point detection)을 통해 순수음성 구간만을 분리하고 256포인트(Point)를 한프레임(Frame)으로 하여 128포인트가 중첩되도록 해밍창(Hamming Windows)을 취한후 음성특징요소로서 14차 LPC-캡스트럼을 구하였다[15]. DTW에서의 화자 모델은 각 발음수에 따른 단어를 사용하여 대표평균 벡터를 만들었다. DHMM과 SCHMM에서는 14차 LPC-캡스트럼계수를 이용하여 한국어에 존재하는 모든 자음과 모음을 k-means 기법을 이용해 64개의 코드워드로 구성된 코드북 벡터를 생성하였다. 그리고 이 기법을 통해 각각 코드워드의 공분산(Covariance)을 구해 SCHMM의 14차 멀티가우시안분산(Multigaussian variance)값으로 사용하였다. 또한 실험에 사용된 HMM모델은 상태수(State number)를 6개로 하고 Left to right 모델을 적용하였다. DTW의 경우 최적의 정합함수를 통해 총 누적거리는 임계치(Threshold value)와 비교되어서 수락할 것인지, 거부할 것인지 결정된다. 본 실험에서는 적당한 임계치를 찾기 위해서 학습발음들과 채택된 참조발음간의 거리를 측정한후 가장 큰 것을 임계치로 채택하였다[15]. HMM은 월드모델(World model)방식[16]을 이용하여 임계치를 설정

하였다. 각 화자인식기법의 학습발음수에 따른 화자인식율은 표 2와 그림 3이다.

표 2. 화자인식율
Table 2. The speaker recognition rate.

인식기법 / 발음수	DTW	DHMM	SCHMM
2번	92.74	95.13	95.53
3번	94.31	95.60	96.25
4번	94.78	97.29	99.18
6번	97.29	97.85	99.54
8번	96.51	98.28	99.48
10번	95.36	98.28	99.07
14번	96.96	98.52	99.27
20번	96.89	98.32	99.48
30번	95.88	98.32	99.75

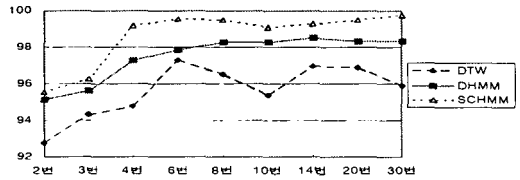


그림 3. 화자인식율
Fig 3. The speaker recognition rate.

본 연구에서 실험한 각 화자인식기법의 시스템의 처리속도를 예측하기 위해 화자인식과 학습과정에서 필요로 하는 계산량을 근사적으로 산출하여 비교한 결과가 표 3과 같다.

표 3. 화자인식기법별 근사적 계산량
Table 3. The approximated calculation number for the each speaker recognition algorithm.

인식기법		학습시	인식시
DTW	곱셈	T	LT+2
	덧셈	(T+2)P-T-1	2P(LT+1)+1
DHMM	곱셈	I{N2(6T-1)+N[(2L+1)T+L+1]+1 +T(P+1)L }	N+N2(T-1)
	덧셈	I{N2(6T-1)+N[(2L+1)T+L+1]+1 +T(P+1)L }	
SCHMM	곱셈	I{N2(6T-1)+N[(T+3)L+T+1]+2PLT+5}	N+N2(T-1)+LT
	덧셈	I{4N2(T-1)-2N(T-1)+N2+LT(3N+P+1)}	LT
I : 반복(Iternation)수, N : 상태(State)수, L : 코드북 크기(Codebook size), P : LPC-Cepstrum차수, T : 프레임(Frame)수			

DTW와 HMM간의 인식률은 SCHMM, DHMM, DTW 순으로 나타남을 알수 있다. 계산량면에서는 DTW, DHMM, SCHMM임을 알 수 있다. DTW는 인식율에 있어서 HMM기법에 비해 떨어지지만 수행속도 면에서는 유리한 장점이 있다. 학습 발음수가 증가함에 따라 DTW보다는 DHMM, SCHMM의 인식율이 점진적으로 향상됨을 알수 있다. SCHMM, DHMM의 경우 학습발음수가 3회에서 4회로 증가할 때 인식율의 증가율이 가장 높음을 그림 3에서 알 수 있다. 본 연구에서는 화자인식기법들의 성능평가를 통하여 인식율과 실행속도 양쪽측면에서 적절한 성능을 보인 DHMM기법을 화자확인시스템의 인식기법으로 선정하였고 초기학습의 학습발음수는 4회로 결정하였다.

IV. 제안한 웹기반의 화자확인시스템

4.1. 웹기반의 화자확인시스템의 분산설계

본 연구에서 구현하고자하는 화자확인 시스템에서는 시스템 전체의 부하분배(Load balancing)와 클라이언트, 서버 간의 통신트래픽(Traffic)감소를 위해 전처리과정은 클라이언트측에서 인식과정은 서버측에 수행하도록 분산설계하였다.

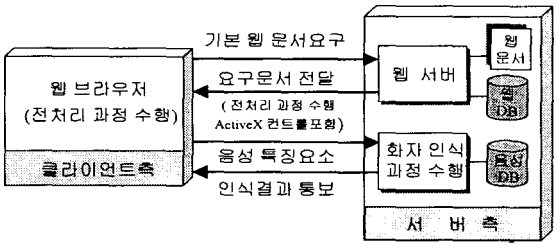


그림 4. 화자확인시스템의 분산설계
Fig. 4. The distributed design of speaker verification system.

본 연구에서는 사용의 편리성 및 실행속도, 개발의 용이성, 그리고 우리나라의 클라이언트 운영시스템이 대부분 MS사의 윈도우시스템(Window System)이라는 점을 고려하여 ActiveX기술을 이용하여 클라이언트측의 전처리과정을 설계하였다. 또한 ActiveX기술이 클라이언트측 운영시스템에 종속적이라 단점을 보완하기 위해 서버측에서 클라이언트측 운영시스템을 파악한 후 이미 개발되어 서버측에 등록된 ActiveX컨트롤중에서 클라이언트의 운영시스템과 일치하는 ActiveX컨트롤을 다운로드 받게하는 방식으로 설계하였다. 클라이언트가 웹 사이트에 접속하였을 경우 웹 서버에 등록된 응용프로그램에서 클라이언트측의 운영시스템을 파악하여 클라이언트측 운영시스템에 맞는 ActiveX 컨트롤을 선택하여 웹 서버를 통해 ActiveX 컨트롤을 HTML문서를 내장하여 클라이언트측에 전송한다.

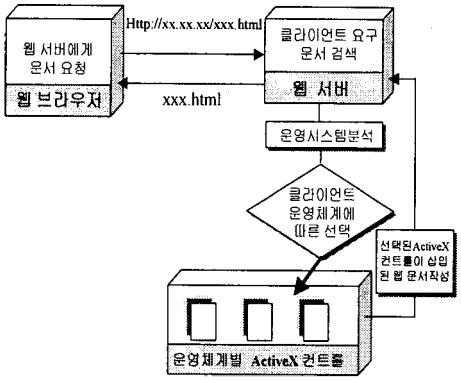


그림 5. 운영시스템에 따른 ActiveX 컨트롤선정 방식
Fig. 5. The method of selecting the ActiveX control according to operating system.

다음그림 6은 ActiveX 컨트롤기술을 이용한 클라이언트측 사용자 인터페이스이다.

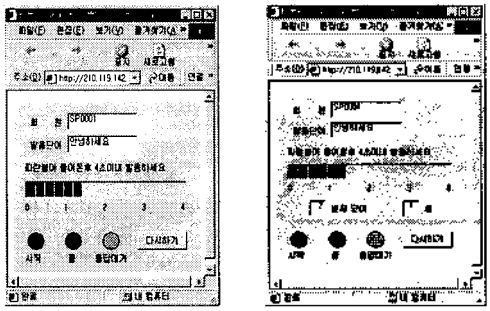


그림 6. 클라이언트측 사용자 인터페이스
Fig. 6. The client side user interface.

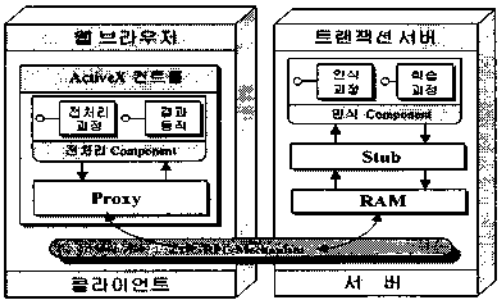


그림 7. DCOM을 사용한 화자확인시스템의 구성
Fig. 7. The diagram of speaker verification system using the DCOM.

서버측은 클라이언트측으로부터 전송된 음성 특징요소들을 이용하여 인식과정을 수행한다. 클라이언트측으로부터 음성특징요소들을 받아 인식과정을 수행하고 그결과를 다시 클라이언트에 통보하는 전과정을 오브젝트기반(Object based)의 분산처리기법인 DCOM을 이용하여 설계하였다. 다음그림 7는 DCOM기술을 이용하여 화자확인시스템을 분산설계한 것이다.

본 연구의 화자확인시스템은 크게 클라이언트 컴포넌트와 서버 컴포넌트로 구성되어 있다. 클라이언트 컴포넌트(Component)는 2개의 인터페이스(Interface)로 구성되어 있다. 디지털화된 음성비트스트림으로부터 음성 특징요소(LPC-Cepstrum)를 추출하는 인터페이스와 인식결과를 서버측으로부터 통보받는 인터페이스로 구성된다. 전처리과정 인터페이스는 추출한 음성 특징요소를 프록시객체(Proxy Object)에 전달한다. 프록시객체에서는 추출된 음성특징요소를 패킷(Packet)화 하는 마샬링(Marshalling)과정을 수행한 후 채널을 통해 서버로 전송하고 스텁객체(Stub Object)를 통해 서버로 전송된 패킷으로부터 음성요소 데이타를 복원하는 언마샬링(Unmarshalling)과정을 수행하여 서버의 인식 컴포넌트내 인식과정 인터페이스에 음성 특징요소를 전달한다. 인식과정을 수행한 후 그 결과를 앞와 전송된 절차의 역순으로 하여 클라이언트측의 결과통지인터페이스에 통보한다. 각 컴포넌트내부는 그림 8과 같다.

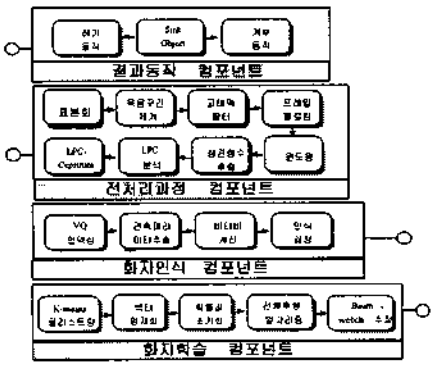


그림 8. 각 컴포넌트의 내부구성
Fig. 8. The internal diagram of each component.

4.2. 웹 기반의 화자확인시스템의 동작설계

본 연구에서 제안하는 웹 기반의 화자확인 시스템에서 사용되는 화자확인인식기법은 화자인식율을 높이고 녹취로 인한 사칭방지를 위해 화자의 성별, 나이에 따라 화자의 개인특성을 잘 나타내는 단어집합, 즉 단어군을 서버측에서 클라이언트측에 전송하여 학습시키고 화자확인시 서버측에서 클라이언트측에 학습한 단어들중 임의의 단어를 전송하여 발음하게 하는 문자지시형 화자확인기법으로 시스템을 설계하였다. 본 연구에서 제안한 웹 기반의 화자확인시스템은 클라이언트측으로부터 음성특징요소를 추출하는 ActiveX부분과 일반적인 웹 서비스를 수행하는 웹서버, 화자인식작업을 수행 및 관리조정하는 트랜잭션서버와 회원의 신상정보 및 일반적인 사이트 정보를 저장하는 웹 사이트 데이터베이스(DataBase), 회원의 음성정보를 저장하는 음성 데이터베이스로 구성 되어있다.

본 연구에서는 MS사(Microsoft Company)의 NT4.0서버 상에서 웹서버로 IIS4.0(Internet Information Server), 트랜잭서버로 MTS2.0(Microsoft Transaction Server), 웹 사이트 데이터베이스(DataBase), 음성 데이터베이스(DataBase)로 MS_SQL7.0을 이용하여 최근에 컴퓨터시스템환경구축에서 각광받고 있는 3계층 (Three tier)방식[17]으로 화자확인시스템을 설계하였다. 본연구에서 제안하는 웹 기반의 화자확인 시스템의 주요동작은 화자확인동작과 화자학습동작으로 나눌수 있다. 화자확인기능의 동작절차를 요약하면 그림 9와 같다.

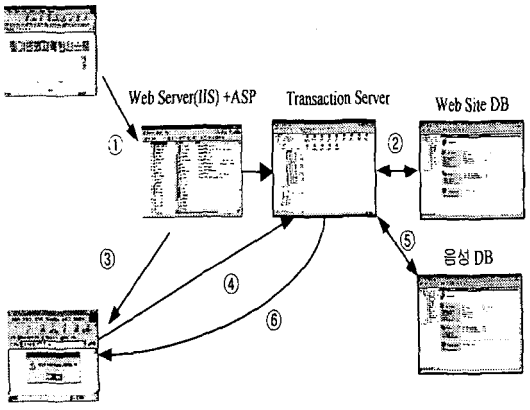


그림 9. 화자확인동작도
Fig. 9. The speaker verification operation diagram.

① 클라이언트가 웹 서버에 접속하여 기본 HTML문서를 요구하고 웹 서버는 클라이언트에세 회원식 별번호(ID)을 입력할 수 있는 폼(Form)형태의 HTML 문서를 클라이언트측에 보낸다. 클라이언트는 회원ID을 입력하여 웹 서버로 전송한다.

② 클라이언트로부터 전송된 회원ID를 이용하여 웹 사이트 데이터베이스(Web site database)의 회원 관리 테이블에서 회원의 단어군ID을 검색하여 단어군,단어테이블로부터 클라이언트가 발음할 단어를 임의로 선택하여 단어를 식별하기위한 단어ID와 함께 클라이언트측에 전송한다.

③ 웹서버로부터 클라이언트측에 전송된 단어를 클라이언트가 발음하여 취득된 음성 비트스트림(Bit stream)을 전처리(Preprocessing) 하여 음성 특징요소을 추출하여 트랜잭션서버에 전송하는 기능을 수행하는 컴포넌트를 HTML문서내에 ActiveX컨트롤형태로 삽입(Embedding)하여 클라이언트 측에 전송한다.

④ 클라이언트측 HTML문서의 ActiveX컨트롤이 클라이언트가 발음한 음성신호를 A/D(Analog to Digital)한 음성 비트스트림으로부터 추출한 음성특징요소, LPC-켑스트럼을 클라이언트의 회원ID, 발음한 단어ID과 함께 트랜잭션서버로 보낸다.

⑤ 트랜잭션서버는 클라이언트로 전송된 회원ID와 단어ID을 이용하여 클라이언트가 회원등록시 학습등록한 단어들중에 단어ID와 일치하는 화자파라미터ID(Speaker parameter ID)를 구한 후 화자파라미터ID와 일치하는 화자파라미터데이터를 음성DB로부터 검색한다.

⑥ 입력된 음성패턴과 기준패턴을 화자확인기법을 이용하여 화자확인전차를 수행한 후 그결과를 클라이언트측에 통보한다.

화자확인동작시 클라이언트와 각 서버간에 주고받는 통신정보를 나타내면 그림 10과 같다.

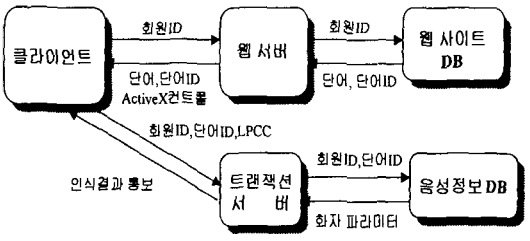


그림 10. 화자확인시 통신정보
Fig. 10. The communication information at the speaker verification time.

화자학습과정은 웹사이트의 회원등록과정중에 이루어진다. 화자학습기능의 동작절차를 요약하면 그림 11과 같다.

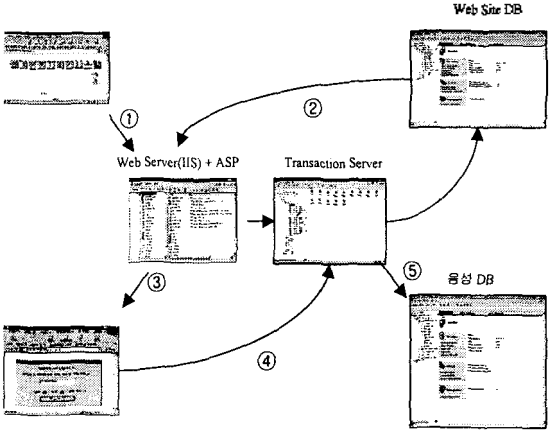


그림 11. 화자학습동작도
Fig. 11. The speaker trainning operation diagram.

① 클라이언트가 웹사이트에 회원등록을 하면 클라이언트의 신상정보가 웹사이트DB에 저장된다.

② 클라이언트의 신상정보를 웹 사이트DB에 입력한 후에 이를 이용하여 사전에 만들어진 단어테이블에서 학습할 단어군을 결정하고 단어군ID을 회원관리테이블의 단어군ID필드(Field)에 입력한다. 단어군은 각군마다 4~5의 단어들로 이루어져 있다.

③ 클라이언트의 신상정보에 의해 결정된 학습할 단어군에 속한 단어들과, 각 단어ID을 서버에서 클라이언트측에 전송한다.

④ 클라이언트는 각 단어별로 4회 발음하고 음성신호로부터 LPC-켑스트럼을 구해 회원ID,단어ID와 함께 트랜잭션 서버로 보낸다.

⑤ 트랜잭션 서버는 학습과정을 수행한 후 추출된 화자 파라미터데이터와 회원ID,기준패턴데이터와 관계된 기준패턴ID를 음성정보DB에 저장하고 음성정보DB에서는 전달된 단어ID와 화자 파라미터ID를 연결하는 테이블을 구성한다.

⑥ 트랜잭션서버는 다음 단어의학습이나 학습종료를 클라이언트에 통보한다.

여기서 ⑤과정은 클라이언트로부터 학습할 모든단어의 음성 특징요소들을 받은 후 임시파일에 저장하고 시스템 작업스케줄링(Job secheduling)에 의해 시스템의 부하가 없는 한가한 시간대에(Idle time)수행하면 학습과정에 걸리는 시간을 클라이언트가 기다릴 필요가 없이 짧은 시간내에 등록을 끝낼 수 있다. 화자학습동작시 클라이언트와 각 서버간에 주고받는 통신정보를 나타내면 그림 12과 같다.

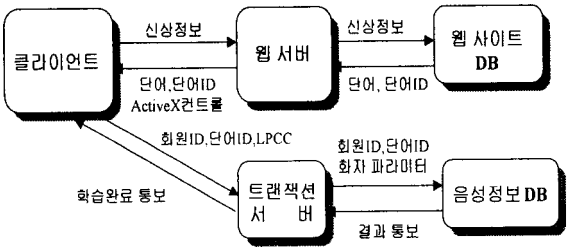


그림 12. 화자학습시 통신정보
Fig. 12. The communication information at the speaker training time.

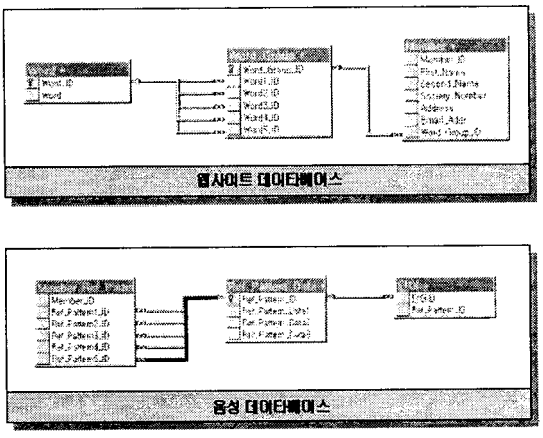


그림 13. 각 데이터베이스의 테이블간의 관계
Fig. 13. The table relation of the each database.

본 연구에서 제안하는 웹기반의 화자확인시스템은 2개의 DB를 구축한다. 하나의 데이터 베이스는 일반적인 웹서비스를 위해 사용되는 웹사이트DB이고 다른하나는 음성정보를 저장관리하기위해 별도의 음성 DB를 구축한다. 각 DB에 구축되는 테이블간의 관계는 그림 13와 같다.

V. 결론 및 고찰

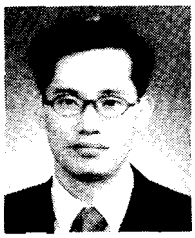
본 연구에서는 웹 사이트의 사용자확인을 위해 패스워드를 입력하는 대신 사용자의 음성정보를 이용하여 사용자 확인절차를 수행하는 웹기반의 화자확인 시스템을 설계하였다. 본 연구에서 제안한 화자확인 시스템은 음성신호 전처리과정과 인식과정을 분리하여 전처리과정은 클리이언트측 즉 웹브라우저상에서 수행하고 인식과정은 서버측에서 수행하도록 분산설계하였다. 웹 브라우저상에서 전처리과정을 수행하기위해 ActiveX기술을 이용하였고 추출된 음성 특징요소를 서버측에 전달하고 그 결과를 전달받기 위한 방법으로 DCOM기술을 선택하였다. 인터넷상에서 동작하는 화자확인 시스템에 적합한 화자인식기법을 도출하기 위하여 대표적인 화자확인기법들의 성능과 특성을 컴퓨터 시뮬레이션을 통해 비교 평가한후 적합한 화자인식기법과 초기 학습발음수를 선정하였다. MS사의 MTS를 이용한 3 계층방식으로 설계하였다. 본 연구를 통해 화자인식기술을 인터넷분야와 접목 시킬수 있는 방안을 제안하였다. 추후과제로 화자확인시스템의 인식율을 높이기위한 음성 특징요소에 대한 연구, 시스템 동작의 안정화를 위한 설계보완과 더불어 동시 사용자수가 증가할 때에도 서버의 처리성능을 유지하기위한 멀티스레딩(Multithreading)프로그래밍기법을 이용한 서버측 프로그래모듈의 설계 및 구현에 대한 연구가 필요하다.

참 고 문 헌

1. Daugman J., "High confidence visual recognition of person by a test of statistical independence," IEEE Trand.on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, 1993.
2. B.G.Sherlock, "Fingerprint Enhancement by Dire ctional Fourier Filtering," IEEE Proc.on Umage Dignal Processing, vol. 141, on. 2, 1994.
3. David Chappell, "Understanding ActiveX and OLE," Microsoft Press, 1996.
4. Brain Farrar, "Special Edition Using ActiveX," Que Corp., 1997.
5. Richard Grimes, "Prefessional DCOM Programming," Wrox Press, 1997
6. S.E.Rosenberg and M.R.Sambur, "New techmiques for automatic speaker verification," IEEE Trand. on ASSP. vol. 23, no. 2, 1975.
7. S.Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Trans. on ASSP., vol. 29, 1981.
8. L.R.Rabiner and B.H. Jung, "Fundamentals of Speech Recognition," JPrentice-Hall, 1993.

9. S.Furui, "Recent advances in speaker recognition," Patterm Recognition Letters 18, 1997.
10. Scott Lsaacs, "Inside Dynamic HTML," Microsoft Press, 1998.
11. Gray Comell and Cay S. Horstmann, "Core Java," SunSoft Press, 1997.
12. Object Management Group, "CORBA:Common Object Request Request Broker Architecture and Specifi cation Revision 2.0," 1996.
13. P.Emerald and J.C.Shih, "DCOM and CORBA Side by Side, Step by Step, and Layer by Layer," Http://www.bell-lab.com/~emerald/dcom-corba/paper.html, 1999.
14. J.D.Markel and A. H.Gray, "Linear Prediction of Speech," Springer-Verlag, 1976.
15. M.Pandit and J. Kittler, "Feature selection for DTW_based verification," Proc. ICASSP, vol. 2, 1998.
16. Yong Gu and Trvor Thomas, "A hybrid score measurement for HMM-based speaker verification," Proc. ICASSP, vol 1, 1999.
17. Island Hopper, "Designing Component-Based Applications," Microsoft Press, 1999.

▲ 이 재 희(Jae Hee Lee)



1986년 2월 : 광운대학교 전자통신과 졸업(학사)
1988년 2월 : 광운대학교 전자통신과 졸업(석사)
1988년~1993년 : 국방과학연구소 근무 (선임연구원)
1993년~1999년 : 대덕대학 정보통신과 (조교수)
1999년~현재 : 동서울대학 전자통신과 근무

▲ 강 철 호(Chul Ho Kang)
한국음향학회지 제17권 8호 참조