

다중사용자 처리기술을 이용한 전자상거래용 화자확인 사용자 인증 시스템

정석영*, 유완선**, 강선미*

*서경대학교 컴퓨터학과

** (주)아이커머스코리아

e-mail : syjeong@ihci.seokyeong.ac.kr, smkang@seokyeong.ac.kr

Automatic User-identification verification system using speech signatures based on multi-user processing technology for secured electronic commerce systems

Seokyeong Jeong*, Wansun Yu**, Sunmee Kang*

*Dept. of Computer Science, Seokyeong Univ.

**iCommerce Korea

요약

전자상거래 시장이 활발해 지면서 인터넷 쇼핑몰 업체들은 보다 강력한 보안체제를 갖추기 위해 노력하고 있다. 특히 최근 생체정보의 인식사례가 전보이면서 이런 인식서비스를 전자상거래에 접목시키려는 노력이 활발히 이루어지고 있는 중이다. 그러나 다중사용자를 실시간으로 처리해야하는 전자상거래 서비스 특성상 부하가 많이 소요되는 인식엔진을 전자상거래용 서버에 포함시키는 것은 적지 않은 부담이 든다. 본 논문에서는 이러한 문제의 해결방안으로 고객의 목소리를 통한 사용자 인증 시스템을 별도의 다중 처리 시스템으로 구성하는 것을 제안하며 이의 구현 사례를 보이고자 한다. 부하가 많이 따르는 인식엔진 등의 서비스를 별도로 관리함으로써 다중사용자 접속을 요구하는 많은 서비스에 유용한 해결 방안을 보이고자 한다. 본 서비스는 인터넷 쇼핑몰 프로그램 개발 업체인 (주)아이커머스 코리아의 전자상거래 솔루션과 연동하여 프로그램 개발이 완료된 상태이다.

1. 서론

최근 청소 용역업체의 직원이 우연히 주문 다른 사람의 카드매출전표를 통해 인터넷 쇼핑몰에서 다량의 물품을 구입하다가 경찰에 적발된 사례가 있었다. 물품 구입이 가능했던 것은 쇼핑몰 업체에서 카드번호와 만료 기간만을 기재하면 자동으로 결제가 되도록 서비스를 제공하기 때문이다. 본인임을 증명하는 비밀번호를 기재하기 위해서는 카드 회사에 건당 일정액의 수수료를 부가해야 하기 때문에 인터넷 쇼핑몰 업체에서는 이를 부담스럽게 여기고 본인확인 서비스를 지원하지 않은 것이 문제였다.

이와 같이 전자상거래는 그 특성상 모든 결제를 본인의 신용카드로 이루어지며 신용카드의 각종 정보가 쉽게 유출되어 도용될 수 있는 위험부담을 지니고 있다. 이에 인터넷 쇼핑몰 솔루션 회사는 이러한 문제를 종식시켜줄 해결책과 함께 보다 저렴하고 부담이 적은 방법을 찾고자 노력하고 있다.

이에 따라 비밀번호 등의 외부유출이 용이한 해결책 대신 개개인의 고유한 정보인 생체 정보를 이용한 보안연구

는 해를 거듭할수록 실용화되고 있으며, 이미 사인이나 지문, 홍채의 경우가 그렇고, 얼굴 인식 등의 기술도 부분적으로 사용되고 있다. 하지만 이런 생체정보들은 인식을 위해 특수한 장비가 요구되어 지며, 또한 인식을 위해서 서버측 컴퓨터에 많은 부담을 주게되므로 실용도 측면에서는 적합하지 못하게 된다.

따라서 본 논문에서는 이러한 추가적인 작업을 위한 엔진이 보다 효율적으로 처리되게 하기 위한 다중사용자 처리 프로세스 시스템의 구성을 제안하고 이의 구현 사례를 보이고자 한다. 전체 시스템은 고객의 목소리를 통한 화자확인 시스템으로 고객의 PC에 장착된 일반 마이크를 이용하여 본인이 정한 음성비밀단어를 입력함으로써 사용자임을 증명하는 음성확인 전자상거래 시스템이다. 특히, 다중사용자 처리가 필요한 인터넷 쇼핑몰 상에서 3-tier 환경에서의 다중사용자 시스템 구성에 중점을 두었다.

2. 화자 확인용 전자상거래 서버의 다중 사용자 처리

전자상거래 서비스의 특성상 다수의 사용자가 동시에

서버에 접속할 가능성을 배제할 수 없다. 초기에 구현된 시스템들은 web 기반 C/S(Client Server) 환경의 2-tier 시스템으로 전자상거래 등의 서버를 구축하였지만, 인터넷의 대중화로 인해 수많은 사용자의 접속을 감당하기에는 많은 부담이 야기되었다. 따라서 많은 업체들이 이러한 문제를 Client-Server 환경에서 벗어나 미들웨어를 추가한 3-tier 형식의 시스템으로 구현하고 있는 상황이다[1]. 하지만 미들웨어의 구현을 위해 CORBA 나 DCOM 기술을 사용한 분산객체 기술을 도입하려 해도 그 구현이 복잡하고, 인식엔진 등의 별도 프로세서를 처리하기 위해서는 인식엔진 자체의 구현도 분산객체의 인터페이스에 맞게 개발되어야 하기 때문에 제품 구현에 많은 부담감을 안고 있다.

따라서 본 시스템에서는 전자상거래 업체의 서버 내부에 보다 간편한 방식으로 client의 요구와 인식엔진의 연결이 가능하도록, 별도의 다중사용자 통신 관리 시스템을 구현하였다. 본 시스템을 독자적인 시스템으로 구현함으로써 기존에 구현된 웹서버와 보안을 위해 구성된 다양한 인증 시스템을 보다 안정적인 환경 하에서 서비스가 이루어지도록 했다.

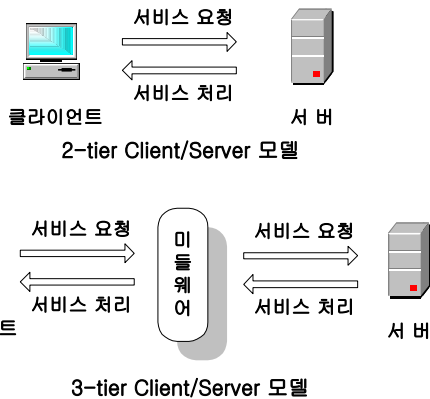


그림 1. 2-tier 및 3-tier Client/Server 모델

프로세서의 제어는 VCM(Voice password system Communication Manager) 이라는 총괄 시스템에서 관리하게 된다. 프로세서간의 통신은 각 모듈 상위에 Message Queue를 두어 FIFO 형태의 서비스를 받게 된다.

VCM은 크게 3 부분으로 구성되어 있다. Webserver 등의 외부 시스템으로부터 전송되어진 데이터를 처리하는 통신관리 프로세스(Communication Manager) 부분. 입력 데이터와 인증 처리 엔진을 연계시키는 인증처리 프로세스 관리 (Voice password system Manager)부분. 그리고 두 개의 프로세스를 message queue로 연동하여 guide 하는 연결처리 프로세스(Rotary) 부분이다.

각각의 프로세스를 살펴보면 다음과 같다.

1) Commgr(통신관리 프로세스)

Commgr 프로세스는 외부의 webserver 등의 시스템으로부터 입력되어진 데이터들을 소켓통신을 통하여 message queue로 전송시키는 부분을 담당한다. 우선 VCM이 실행되면 commgr 프로세서는 소켓에 접속되기를

대기하고, 소켓이 연결되면 subcom 프로세서를 실행시켜 전송작업을 수행한다. 전송작업이 완료되면 subcom 프로세서를 닫는다.

Commgr을 임의로 종료시키거나 에러로 종료될 때 subcom은 자동적으로 수행이 종료된다.

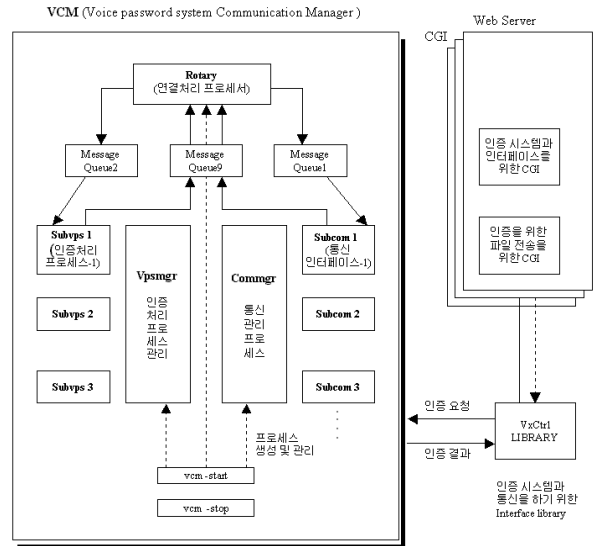


그림 2. VCM 전체 흐름도

2) Vpsmgr (인증처리 프로세스 관리)

인증시스템(VPS)의 처리는 vpsmgr에서 수행한다. VPS(Voice Password System)는 실질적인 화자 인식 엔진 시스템으로 시스템의 환경을 고려하여 필요한 개수만큼의 프로세스를 subvps 로 정의하여 생성한다. VCM이 실행되면 vpsmgr 프로세서는 우선 subvps 프로세서 3개를 실행시킨다. Rotary 프로세서를 통하여 message queue에 대기하고 있는 화자 데이터 정보를 받게되면 호출 순서대로 인식을 수행하고 종료하게 된다. vpsmgr을 임의로 종료시키거나 에러로 종료될 때 subvps는 자동으로 종료된다. 또한, 수행되는 subvps의 개수가 특정개수 이하 일때는 vpsmgr은 부족한 수만큼의 subvps를 자동으로 수행시킨다.

2) Rotary (연결처리 프로세스)

Rotary 프로세서는 subvps와 subcom의 교량역할을 하는 프로세스이다. VCM이 실행되면 message queue를 open하여 메시지가 전달되기를 대기하고 메시지가 들어오면 목적지를 구분하여 메시지를 전송한다.

Rotary 프로세서는 차후 다른 인증절차나 확장된 프로세서를 접목할 때 편리함을 고려하여 설계되었다.

3. PC 기반 사용자를 위한 Browser용 인터페이스[2]

본 논문에서 소개한 화자확인기술은 인터넷 쇼핑물에서 고객이 쇼핑을 마친 후 제품 구매시에 적용된다. 따라서 고객은 인터넷상에서 웹 브라우저를 사용하다가 화자확인 서비스를 받게 되며, 그에 따른 모든 처리는 인터넷상에서 이루어져야 한다.

현재 구현된 프로그램은 Active-X의 기술을 사용하여 ocx형태의 component로 개발하였다. 사용자는 화자확인 사이트에 접속할 때 최초 1회의 다운로드를 받게되며, 이후에는 registry에 등록되어 별도의 지연시간 없이 서비스를 받게 된다. 화자확인 인터페이스 프로그램(VPSR)은 사용자와 사용자 인증 서버(Client-Server) 간의 데이터 통신 및 동작을 수행하도록 구현하였고, 웹브라우저 안에서 음성이 입력되는 상황을 과정으로 관찰하며 수행되도록 구현되었다.

VPSR component는 Visual C++ 6.0 에서 MFC 기반 하에 구현되었다[3]. 본 VPSR 프로그램의 내용은 다음 그림 3과 같은 순서로 실행된다.

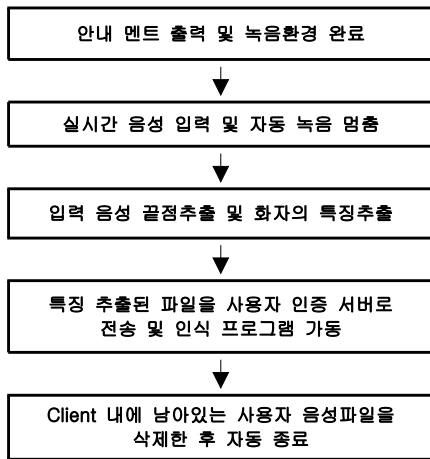


그림 3. VPSR component flowchart

먼저 프로그램이 실행되면 안내문구와 함께 사용자의 녹음을 기다린다. 이때 사용자는 녹음 버튼을 눌러 마이크로 자신의 비밀단어를 입력하고, 입력이 끝나치면 VPSR 프로그램이 자동으로 녹음을 중지시킨다. 그 후에 입력되어진 wave 형태의 파일을 분석하여 끝점을 검출하고, 검출된 파일의 특징을 추출하여 화자인증 서버 시스템에 추출된 특징벡터 파일을 전송한다. 특징 파일들을 전송할 때에는 서버 측에서 방화벽등의 보안장치를 해 놓았을 경우를 대비해 일반 http 사용 포트인 80번 포트를 통해 CGI 명령 중 post 명령을 사용하여 하나의 stream 파일 형식으로 전송하게 된다.



그림 4. 화자 인증 인터페이스 프로그램

파일이 전송되면 화자인증 서버의 인증 프로그램을 실행

시키고 client 컴퓨터 내에 남아있는 특징추출 파일을 삭제한 후 자동으로 종료된다. 서버측의 인증 프로그램은 CGI 형태로 만들어져 있어 URL을 입력하는 것으로 프로그램을 실행시킬 수 있다.

4. 화자 인증 시스템 [4]

본 시스템에서는 화자확인기술을 사용하여 인터넷 쇼핑 물 가입자의 사용자 확인 서비스를 제공해 준다. 화자 확인(speaker verification)은 발생된 음성이 원하는 화자(의뢰인, client speaker)인지 아닌지(사칭자, impostor)를 구분해 내는 것으로 의뢰인에 대한 초기 등록을 요구하게 된다. 이 방법은 발화된 음성과 확인을 바라는 의뢰인의 음성비밀단어의 입력으로 시작되어 확인과정을 거친 후 의뢰인으로서 수락할 것인지(accept) 거부할 것인지(reject)를 결정하게 된다. 이 서비스는 여러 보안 시스템에서 사용되는 사용자 인증 과정과 유사하므로 화자 인증 시스템이라고도 한다[5].

화자 인증 시스템을 구현하기 위해서는 개별 화자의 특징을 올바르게 추출하는 방법이 중요하다. 본 연구에서는 인식 성능이 좋고 주변 잡음과 채널 왜곡에 강한 MFCC(Mel Frequency Cepstral Coefficient)를 사용하여 MFCC 계수 12개와 각 차원의 켈스트럼 계수의 미분 값, 그리고 파워 및 파워의 미분값을 특징값으로 사용했다.

화자확인 기본 동작은 그림 5와 같다. 테스트할 화자의 음성은 먼저 끝점 검출이 되고 특징 파라미터의 추출을 통해 분석된다. 추출된 특징 파라미터는 기준에 저장되어 있는 기준모델 및 요구되는 화자 개인 모델과 비교된다. 이렇게 해서 나온 likelihood를 이용해 화자확인이 이루어진다.

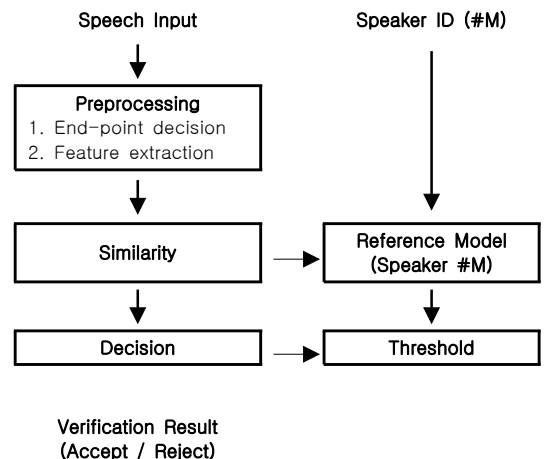


그림 5. 화자 인증 과정

인식 모델로는 8개의 Gaussian mixture를 사용한 5 state left-to-right 연속 HMM을 사용하였다. 입력된 화자의 정보를 HMM 파라미터로 생성하며 훈련을 위해서는 clustering 작업이 필요하다. 본 시스템에서는 Segmental k-means 알고리즘을 사용하여 훈련과정을 수행한다.

다음 단계는 최적 state sequence가 나올 확률을 구하는 과정으로 Viterbi algorithm을 사용한다. 최적 state

sequence는 재귀적인 방법으로 $\delta_i(j)$ 가 최대값이 될 때까지 반복 계산한 후에 최적의 sequence를 구하게 된다.

입력 음성에 대해 요구한 사람이 맞는지 아닌지를 결정하는 화자확인 과정은 hypothesis testing이 사용된다. 요구한 화자가 맞는지 아닌지를 결정하기 위해 입력 음성에 대해 likelihood ratio를 적용하는 것이다. 입력음성이 $X = \{x_1, \dots, x_T\}$ 이고 모델 λ_c 에 해당하는 화자라고 요구했을 경우 likelihood ratio는

$$\frac{\Pr(X \text{ is from the claimed speaker})}{\Pr(X \text{ is not from the claimed speaker})} = \frac{\Pr(\lambda_c | X)}{\Pr(\lambda_r | X)} \quad (3)$$

이 likelihood를 Bayes's rule과 uniform한 priors 확률 $P(S_j)$ 를 적용하여 log domain에서 보면 식(4)와 같이 표현된다.

$$\Lambda(X) = \log p(X | \lambda_c) - \log p(X | \lambda_r) \quad (4)$$

이 likelihood ratio $\Lambda(X)$ 가 threshold Θ 와 비교하여 만약 $\Lambda(X) > \Theta$ 이면 요구한 화자를 accept하고 $\Lambda(X) < \Theta$ 이면 reject한다. 따라서 이 threshold를 설정하는 점은 시스템의 성능을 결정하는 중요한 문제이다. Threshold는 false rejection rate(FRR)와 false acceptance rate(FAR) 사이의 trade-off를 고려하여 정해진다.

현재 구현된 알고리즘의 성능을 평가하기 위해 20대 남녀 50명에 대한 음성 데이터베이스를 구축하였고 이에 대한 실험 결과는 FAR이 0.6% 이하, FRR가 3% 이하의 성능을 나타내고 있다.

5. 제안된 전체 시스템 구성

현재 구현된 시스템에 적용하여 본 전체적인 구성은 그림 6과 같다. 각 기능을 요약하면 다음과 같다.

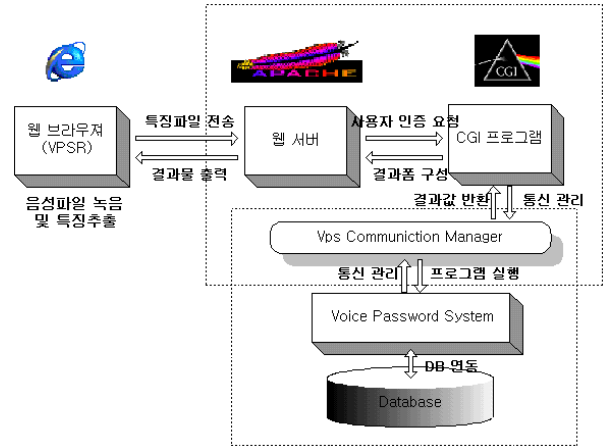


그림 6. 인증 처리 흐름도

- ① ID확인 및 음성녹음 요구
- ② 음성 녹음 및 특징추출
- ③ 특징추출 파일 전송
- ④ 인증 CGI 프로그램 실행
- ⑤ 다중 사용자 연결 처리
- ⑥ 인증 프로그램 DB와 연동실행
- ⑦ 파라미터 계산 및 인증 여부 판단
- ⑧ 결과값 반환
- ⑨ 인증결과 CGI 프로그램상 통보
- ⑩ 웹 브라우저 상 결과확인

6. 결 론

본 논문에서는 사용자 인증 엔진을 전자상거래 서버시스템에 독립적인 시스템으로 구축하여 인증작업을 처리해 줌으로서 다중사용자 접속환경에서의 전자상거래 서버의 부담을 줄여주기 위한 방안을 제시하였다. 현재 시스템의 다중사용자 처리 시스템은 화자인식 엔진과 더불어 구현된 시스템이지만, 별도의 엔진을 요구하는 여러 응용분야에 서버시스템의 부하를 줄이며 안정적인 처리를 요하는 곳에 응용될 수 있다.

따라서 인터넷 쇼핑몰 업체들의 요구에 맞는 인식 엔진들과 연동하여 다중 사용자 처리 시스템을 도입하면, 보다 효율적인 비용과 시스템으로 신뢰성 있는 사용자 인증 서비스를 지닌 전자상거래 사이트를 구축할 수 있을 것이다.

본 연구는 한국과학기술평가원(KISTEP)의 국책사업 중 '98 핵심S/W기술개발 사업 과제인 『음성확인 기술을 이용한 전자상거래용 사용자 인증 시스템 개발』 과제 지원으로 수행되었습니다.

[참고문헌]

- [1] 박재호, "3-tier 프로그래밍 올 가이드", 프로그램세계, 2000. 3
- [2] 정석영, 안성주, "음성확인 기술을 이용한 전자상거래용 사용자 인증 시스템", 추계종합학술발표회, 한국통

신학회, 1999. 11.

- [3] 이상엽, "Visual C++ 6 Programming Bible", 영진출판사, 1999
- [4] 한국과학기술평가원, "음성확인 기술을 이용한 전자상거래용 사용자 인증 시스템 개발", 1차년도 최종보고서, 1999. 9
- [5] 오영환, "음성언어정보처리", 홍릉과학 출판사, 1998. 1.
- [6] 이상엽, "Internet Programming Bible 2nd edition", 영진출판사, 1998
- [7] H. Bourlard, N. Morgan, " Speaker verification a quick overview," IDIAP Research report, 1998.
- [8] C.H.Lee, F.K.Soong, K.K.Paliwal, " Automatic speech and speaker recognition-advanced topics," Kluwer Academic publishers, second printing, pp.31-56, 1997.