

VoIP와 음성인식에 기반한 통합솔루션 서비스 동향

오재삼* · 윤용근**

The Trend of Integrated Solution Service Based on VoIP and Voice Recognition

Jae-Sam Oh* · Young-Keun Yoon**

■ Abstract ■

We are looking at the two different kinds of IT on this paper. One is VoIP and the other is VR (voice recognition). We are more interesting at the evolving techniques and services produced by combining the two techniques mentioned above. Recently, there are so many services and products appeared in the market using voice recognition technique. Now the technique has progressed on the level that can even replace the user interfaces using the GUI or general DTMF. Therefore, we are expecting so many various new services showed up in the market which is combination of the VoIP and VR. Up until now, three models are available in the field which are wired telephone, wireless telephone, and wireless internet. We know the effectiveness of the VoIP is maximized more when this technique is combined with others rather than used alone without other techniques.

Keyword : VoIP, Voice Recognition Technique, Call Center, PSTN, Mobile Networks

1. 서 론

음성은 가장 자연스러운 의사소통 수단이며 가장 널리 쓰이는 정보교환수단으로 이를 사용하기 위한 방법은 매우 다양하게 발전해왔다. 일반적으로

로 통신의 발달 과정 중 중요한 위치를 차지하는 음성통화가 그 중 하나이다. 인터넷을 통해 음성을 전달하는 기술인 VoIP(Voice over Internet Protocol)는 이제 국내에 소개된지 약 4년이 지났고, 인터넷을 아는 사람은 대부분 들어본 적이 있는 쉽게

* 국민대학교 BIT전문대학원

** 국민대학교 정보관리학부 교수

접할 수 있는 기술이 되었다. VoIP 기술에 있어서 우리나라는 다른 나라 못지않게 많은 개발업체가 생겨났고 서비스 업체도 상당히 많은 편이다. 하지만 VoIP 기술이 저렴한 국제전화로만 인식되던 기술에서 다양한 음성 서비스를 제공하는 기술로 사람들의 인식이 전환된 것은 최근의 일이다. 프로토콜에 매여있던 VoIP가 이제는 본격적인 서비스로 나아가고 있는 중인 것이다.

한편 음성인식에 관한 연구는 약 40여년 간의 역사를 가지고 있으며 그동안 많은 변화를 거듭해 왔다. 컴퓨터가 문자대신 사람의 음성을 입력 받아 그 의미를 알아내는 음성인식기술과 임의의 문장을 음성으로 만들어 내는 음성합성기술을 통칭하여 음성처리기술이라 한다. 주로 마이크를 통하여 컴퓨터에 전달된 사람의 음성의 특징을 추출하고 분석하여 미리 입력된 인식 목록에서 가장 근접한 결과를 찾아내는 최첨단 소프트웨어 기술과 그 기술을 실현하는 하드웨어 기술로 구분할 수 있다.

본 논문에서는 다양한 음성서비스를 제공하는 VoIP 기술과 목소리로부터 의미를 알아내는 음성인식기술을 접목하는 분야를 살펴 음성 서비스의

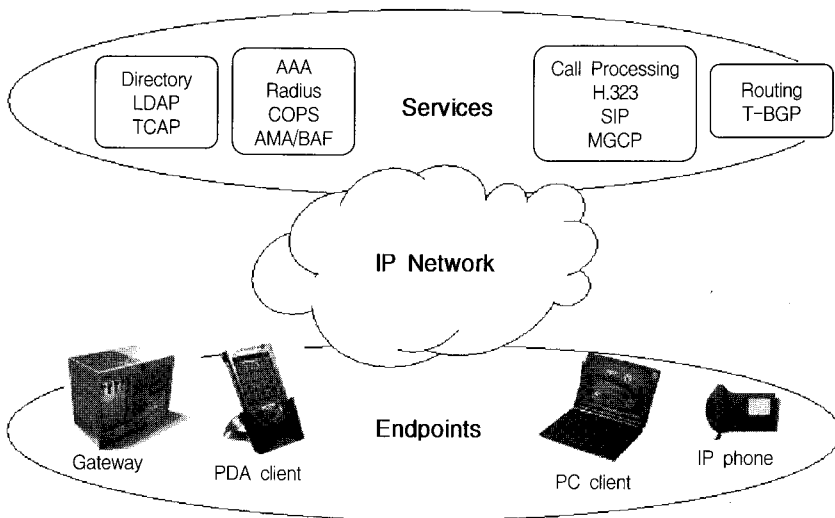
동향과 전화로 입력되는 사람의 음성을 인식하여 이를 구분하여 각기 다른 전화번호로 연결시켜 주는 종류의 음성 서비스를 위한 솔루션으로서의 가능성을 살펴보고자 한다.

2. VoIP 서비스 및 응용 범위

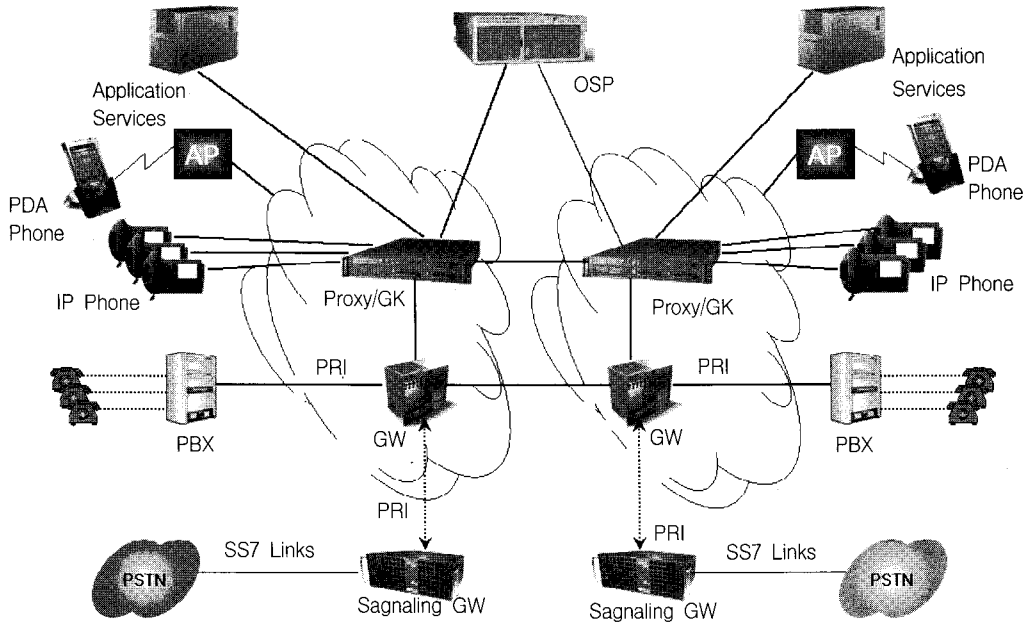
2.1 VoIP 서비스 개요

VoIP란 IP 네트워크 상에서 음성 및 팩스 데이터를 전송하는 기술이다. VoIP를 이용한 가장 대표적인 솔루션은 인터넷 전화로 시내통화요금으로 장거리 전화나 국제 전화를 할 수 있는 통화 요금을 절감할 수 있는 기술로 각광받고 있다. VoIP 시스템의 기본 구성 요소는 [그림 1]과 같다.

음성 데이터나 팩스 데이터를 인터넷으로 전송하는 데는 일반적인 전화기나 팩스를 통해 기업 PBX 또는 ISP가 직접 운영하는 게이트웨이(Gateway)를 통해 전달되는 방법과 IP 전화기나 IP 팩스, PC 상의 클라이언트 프로그램을 통해 전송되는 방법이 있다. VoIP를 이용한 서비스를 위한 네트워크의 구성은 [그림 2]와 같다.



[그림 1] VoIP 시스템 기본 구성 요소[9]



[그림 2] VoIP 네트워크 구성[8]

2.2 VoIP 응용 범위

VoIP가 응용되는 서비스는 다음과 같은 다양한 종류가 있으며, 이외의 다양한 서비스와 결합하여 응용할 수 있는 범위는 무한하다.

2.2.1 저렴한 장거리 전화로 인식되는 인터넷 전화

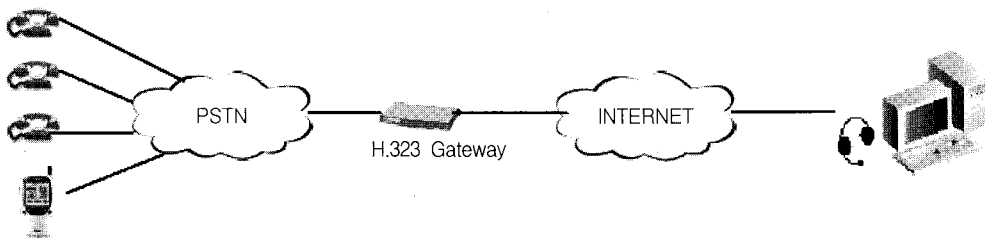
1) PC to Phone

무료전화로 널리 알려진 다이얼패드와 같이 PC에 프로그램을 설치하고 인터넷을 통해 해당지역의 게이트웨이(Gateway)에 연결하여 장거리 전화를 거는 형태이다. 헤드셋을 이용해야 하는 불편

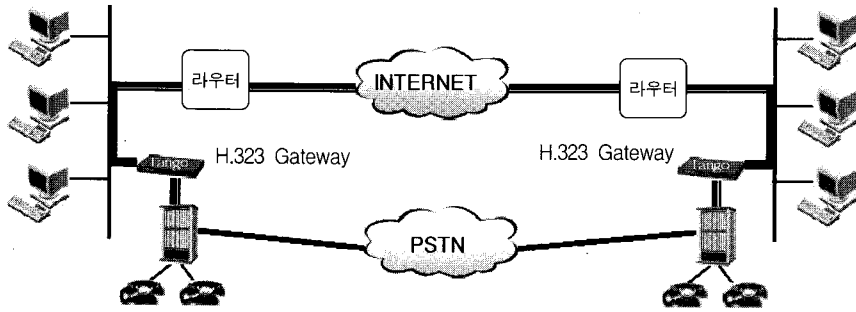
이 있어 IP 전화기를 이용하거나 1 Port Gateway에 일반 전화기를 연결하여 사용하는 형태로 발전했다.

2) Phone to Phone

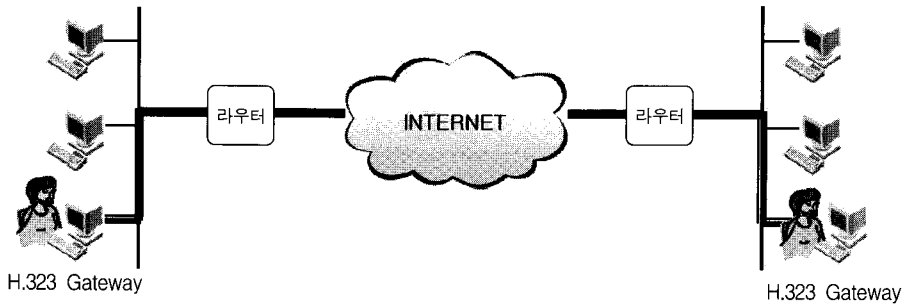
일반전화(가정용 전화, 휴대전화, 공중전화 등)를 사용하여 가까운 인터넷 전화 사업자의 게이트웨이(Gateway)에 접속하여 인터넷을 통해 다시 해당지역의 게이트웨이(Gateway)로 연결하여 전화통화가 연결되는 형태이다. 국제전화서비스업체를 통한 서비스가 대부분 이런 형태이다. 전화기에서 많은 번호를 눌러야 하는 불편이 있어 자동으로 게이트웨이(Gateway)에 연결해주는 오토 다이얼



[그림 3] PC to Phone 방식[11]



[그림 4] Phone to Phone 방식[11]



[그림 5] PC to PC 방식[11]

러를 장착한 전화기가 출시되어 있다.

3) PC to PC

PC의 VoIP 클라이언트끼리 음성 통화를 하는 형태이다. 최근에는 음성채팅과 전화통화를 위해 메시저에 VoIP 클라이언트가 탑재되는 경우가 많아지고 있다.

2.2.2 웹기반 콜센터

VoIP 기술은 전자상거래의 발전과 병행하여 웹기반 콜센터로도 발전하였다. 웹기반 콜센터의 도입은 먼저 기존에 있는 콜센터 하부구조를 확장하는 방법, ASP 서비스 제공자로부터 콜센터 서비스를 임대하는 방법, 그리고 새로운 콜센터를 만드는 방법으로 세 가지 중에서 선택이 가능하다.

2.2.3 UMS

PSTN(Public Switched Telephone Network)을 통해 다양한 형식의 메시지를 동시에 처리할 수 있다.

2.2.4 전자결재 솔루션

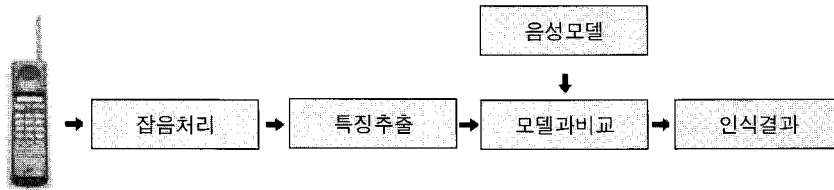
전자결재 시스템은 비실시간 매체 이용으로 인해 발생하는 비실시간성, 비집축성, 비인간성, 설명 불충분 문제를 실시간 문자/음성/화상 통신기술을 이용하여 해결한다.

이외에도 다양한 서비스를 생각해 볼 수 있는데, 요즘은 음성 뿐 만 아니라 주로 PC상에서 메시저를 이용한 화상 통화 서비스가 늘어나고 있다. 이와 같은 서비스를 살펴볼 때 VoIP 기술은 점점 더 많은 다른 서비스와 접목하여 새로운 서비스를 창출함으로써 서비스의 효과를 더욱 극대화 할 수 있다.

3. 음성인식

3.1 음성인식 기술의 정의

음성 처리 기술의 하나로 문자 대신 사람의 음성을 입력하여 음성의 특징을 추출하고 분석하여 미리 입력된 인식 목록에서 가장 근접한 결과를 찾아



[그림 6] 음성인식 방식[5]

그 의미를 알아내는 것을 음성인식기술이라 한다.

3.2 음성인식기술의 종류

음성인식 기술은 그 분류기준에 따라 여러가지 종류로 나뉜다. 음성 인식의 대상, 발음의 형태, 인식 대상 어휘수 등에 따라 세부적으로 분류된다.

3.2.1 음성 인식의 대상(화자)에 따른 분류

음성인식 기술은 음성 인식의 대상으로 삼는 화자에 따라 화자독립과 화자종속 인식 기술로 분류된다. 먼저, 화자종속 시스템은 특정 화자의 음성을 인식하기 위한 시스템으로, 현재 휴대폰에 탑재되어 사용되는 보이스 다이얼링(voice dialing) 시스템이 대표적인 예이다. 화자종속 시스템에서는 일반적으로 시스템의 사용 전에, 사용자의 음성을 저장, 등록시키고, 실제 인식을 수행할 때는 입력된 음성의 패턴과 저장된 음성의 패턴을 비교하는 패턴 매칭(pattern matching) 기법이 사용된다. 반면, 화자독립 시스템은 불특정 다수 화자의 음성을 인식하기 위한 것으로, 화자종속 시스템에서와 같이 사용자가 시스템의 동작 전에 음성을 등록시켜야 되는 번거로움이 없다. 화자독립 시스템은 다수화자의 음성을 수집하여 통계적인 모델을 학습시키고, 학습된 모델을 이용하여 인식을 수행하게 된다. 따라서, 각 화자의 특징적인 특성은 사라지고 각 화자 간에 공통적으로 나타나는 특성이 부각된다. 같은 어휘를 대상으로 같은 양의 학습 데이터를 사용한다면 대체적으로 화자종속 시스템의 성능이 화자독립 시스템보다 높게 나온다. 그러나, 화자종속 시스템의 경우, 음성이 등록된 화자 이외의 사람이 시스템을 사용한다면, 인식률은 크

게 저하된다. 따라서, 최근에는 화자독립 시스템을 구축하고 실제 사용할 때는 사용자의 음성에 적합하도록 인식 모델을 변형하는 기법들이 개발되고 있는데, 이를 화자적응 기술이라 하며, 이러한 시스템을 화자적응 시스템이라 부른다[1].

3.2.2 발음의 형태에 따른 분류

다음으로는 발음의 형태에 따라 고립어 인식 시스템과 연속어 인식 시스템으로 나뉜다. 고립어인식 시스템에서는 각 단어가 또박또박 발음되고 각 단어 사이에는 충분한 길이의 묵음구간이 존재한다고 가정한다. 따라서, 인식의 초점이 각 단어가 다른 단어와 얼마나 다른가에 있고 인접한 단어의 영향은 무시된다. 현재 휴대폰에서 널리 채용되는 보이스 다이얼링(voice dialing)이 고립단어 인식의 좋은 예라 할 수 있겠다. 이에 반해, 연속어인식 시스템은 문장 단위로 인식을 수행하는 시스템을 의미한다. 각 문장은 평상시와 같이 발음되며 특별히 단어 사이의 묵음은 첨가되지 않는다. 연속어의 경우, 한 단어의 특성이 인접한 단어의 발음에 의하여 영향을 받는데 이를 조음효과(coarticulation effect)라 부르며, 이러한 조음효과는 연속어 인식을 어렵게 만드는 큰 요소로 자리잡고 있다. 일반적으로 연속어 인식에서는 각 단어 모델을 연결하여 search space를 구성하고, Viterbi decoding 기법을 사용하여 단어의 시퀀스(sequence)를 찾아나가는 방법을 사용한다.

3.2.3 인식 대상 어휘수에 따른 분류

마지막으로는 인식의 대상이 되는 어휘수에 따라 소용량 시스템과 대용량 시스템으로 구분된다. 소용량 시스템은 어휘수가 수백개 이하인 경우로,

각 단어를 별개의 모델로 구성하는 것이 일반적이다. 반면, 대용량 시스템은 인식 대상 어휘수가 수천에서 수십만에 이르는 시스템으로, 이 경우 각 단어를 별개의 모델로 표현하는 것은 현실적으로 어렵다. 따라서, 각 단어를 음소와 같은 서브워드-유닛(subword-unit)으로 구성한 후, 각 서브워드-유닛(subword-unit)을 하나의 모델로 지정하는 기법을 사용한다.

음성인식 기술은 위에서 언급한 바와 같이 분류 기준에 따라 여러 종류로 나뉜다. 분류된 각 기술은 모두 고유의 장, 단점을 지니고 있기 때문에 어느 하나가 다른 것보다 우월하다고는 얘기할 수 없다. 따라서, 음성인식 시스템을 구현하려고 할 때는 적용분야의 특성과 구현될 시스템의 경제성을 고려하여 적절한 방식을 선택하여야 한다.

3.2 음성인식기술의 응용 분야

음성인식 기술에서 최근 관심의 초점이 되고 있는 분야는 Acoustic modeling, Key-Word Spotting, 화자적응, 환경보상 등이다. 현재 미국을 비롯한 선진국을 필두로 CMU, MIT, 영국의 캠브리지 대학을 비롯한 많은 대학과 연구소에서 음성인식에 관한 연구를 수행하고 있으며, 그 추세는 더욱 확장되고 있는 실정이며 매우 활발하게 전개되고 있다. 미국의 경우, IBM, AT&T, Lucent Technologies, 마이크로소프트 등의 컴퓨터 및 통신관련 기업체들은 독자적인 대용량 음성인식 시스템을 개발하고, 자사의 관련 제품에 응용하고 있다. 한편, 드래곤 시스템즈(Dragon Systems)를 비롯한 많은 소프트웨어 기업들은, 워드 프로세싱을 위한 음성인식 소프트웨어를 출시하고 있으며 그 성능은 날로 향상되고 있다[2].

음성인식 기술의 응용분야를 살펴보면, 우선 주식정보, 전화번호 안내, 지역 정보 안내 등을 비롯한 데이터베이스 쿼리(database query) 분야를 들 수 있다. 이 분야는 원하는 정보를 얻기 위하여 취해야 하는 절차를 음성으로 대체하여 사용자의 편

의를 돕는데 그 목적이 있다. 특히, 근래 매스컴에서 화두가 되고 있는 보이스 포탈(voice portal)은 인터넷의 정보를 사용자가 전화를 통하여 검색할 수 있게 하는 분야로, 음성인식기술이 필수불가결이다. 다른 응용분야로는 컴퓨터와의 인터페이스 분야를 들 수 있다. 넓은 의미로는 음성인식 기술이 적용되는 모든 분야가 이 범주에 포함된다고 생각할 수 있으나, 여기서는 컴퓨터 운영체제의 활용 및 문서편집 분야만을 국한시키기로 한다. 또 다른 음성인식 기술의 응용분야로는 교육 및 엔터테인먼트 분야를 들 수 있다. 현재, 세계 굴지의 전화회사에서는 음성인식을 이용한 다양한 데이터베이스 쿼리(database query) 서비스를 운용 중이며, PC Windows OS 및 딕테이션(dictation)을 위한 소프트웨어가 상용화되어 있는 현실이다. 미래에는 음성인식 기술의 응용분야가 더욱 확대되어 사람들의 문화적 행태에도 커다란 변혁을 가져올 것이라 전망된다.

4. VoIP와 음성인식 통합 기술

VoIP는 단독으로 쓰이기보다는 다른 다양한 기술 및 서비스와 접목되었을 때 그 효과가 극대화된다. 여기서는 VoIP와 음성인식기술을 통합한 서비스의 개요에 대해서 살펴본 후 VoIP와 음성인식 기술을 통합한 적용 분야에 대해 알아보려고 한다.

4.1 VoIP와 음성인식 통합 기술의 개요

VoIP는 특히 음성인식 기술과 통합되었을 때 최대의 기능을 발휘할 수 있다. 꾸준한 발전을 거듭해 온 음성인식 기술 덕택에 이제는 좀더 자연어 처리가 용이하고 다양한 입력이 가능한 서비스 제공이 가능하게 되었다. 현재 유무선 전화 및 PC를 통한 인터넷 서비스에 있어서의 한계는 무엇보다도 제한된 입력 방식에 있다. VoIP에 음성인식 기술을 접목시킬 때 이 문제는 자연스럽게 해결될 수 있다.

4.2 VoIP와 음성인식 통합 기술의 적용 분야

현재 VoIP와 음성인식 기술을 통합한 새로운 서비스가 많이 등장하고 있다. 그 가운데 핵심적인 몇가지를 소개하면 다음과 같다.

4.2.1 음성 다이얼

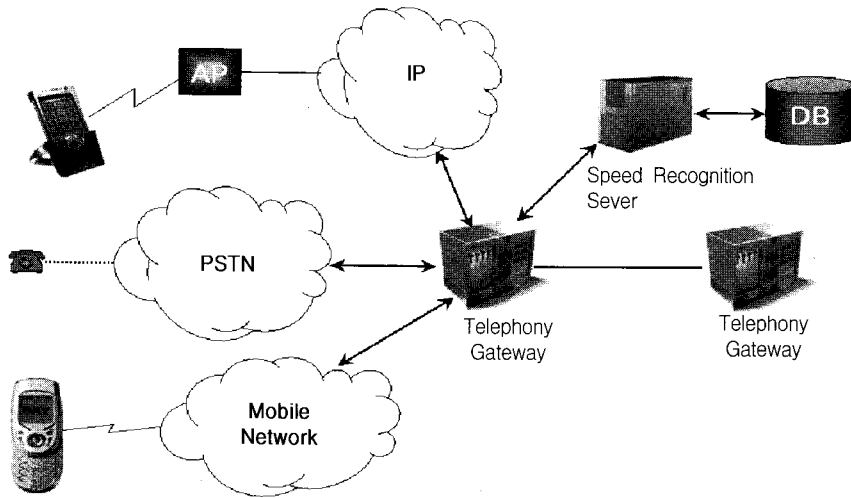
전화로 입력되는 사람의 음성을 인식하여 이를 구분하여 각기 다른 전화번호로 연결시켜주는 것으로 전화번호를 직접 입력하거나 이름 또는 부서를 입력하여 전화번호를 누른 효과를 낼 수 있다. 전화번호를 입력하는 방법은 입력된 음성을 그대로

로 텍스트로 변환하여 전화번호를 알아내는 경우이고, 이름을 입력하는 방법은 입력된 음성을 텍스트로 변환 후 해당하는 전화번호를 찾아 반환하는 경우이다. 예를 들면, 내선번호를 알 수 없을 때 직접 부서명이나 찾는 사람 이름을 말하면 자동으로 연결해주는 방식을 생각해 볼 수 있다.

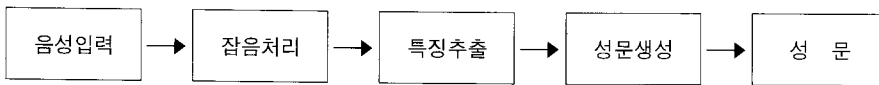
4.2.2 사용자 인증

음성인식 기술 중 화자인증을 이용하면, 국제 전화 서비스업체의 사용자로 가입시 발부되는 PIN과 비밀번호를 누르는 불편을 덜 수 있다.

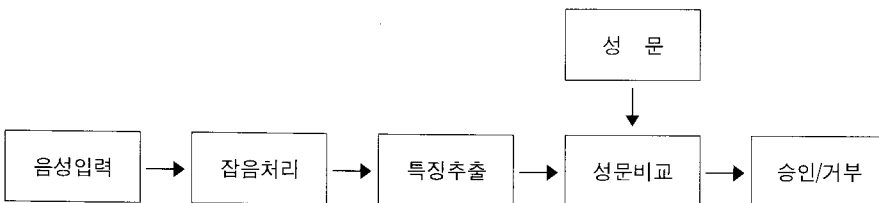
화자인증이란 각 개인들이 지니고 있는 음성의



[그림 7] VoIP+음성인식 네트워크 기본 구성[10]



[그림 8] 화자 등록



[그림 9] 화자 인증

특징, 즉 손가락의 지문과 같은 성문을 이용하여 본인 여부를 인증할 수 있는 일종의 생체 측정을 통한 기술로서, 음성을 암호로 사용하므로 불법 도용이 불가능하다[11].

전화를 걸때 사용자 인증을 위해 많은 번호를 눌러야 하는 방식을 개선할 수 있으므로 향후 많이 쓰이게 될 서비스이다.

4.2.3 UMS

PSTN을 통해 음성이나 문자 등 메시지를 전달함에 있어 통신환경과 시간에 구애받지 않고 발신자와 수신자 모두가 전자우편, 전화, 팩스 등 원하는 형태로 접근할 수 있도록 구현할 수 있다.

UMS에서는 음성인식기술과 음성합성기술을 사용하며 음성명령인식기능 외에도, 음성메시지 문자변환기능으로 음성으로 남긴 메시지를 보이스 메일이 아닌 텍스트 메일과 팩스로도 전달할 수 있다. 또 보이스 스캔 기능으로 음성 사서함이나 보이스 메일을 스캔하여 중요한 내용만 확인해볼 수도 있다[6].

4.2.4 전화 안내 서비스

가장 널리 알려진 음성인식기술 응용 서비스로 음성안내에 따라 필요한 항목을 직접 말하거나 번

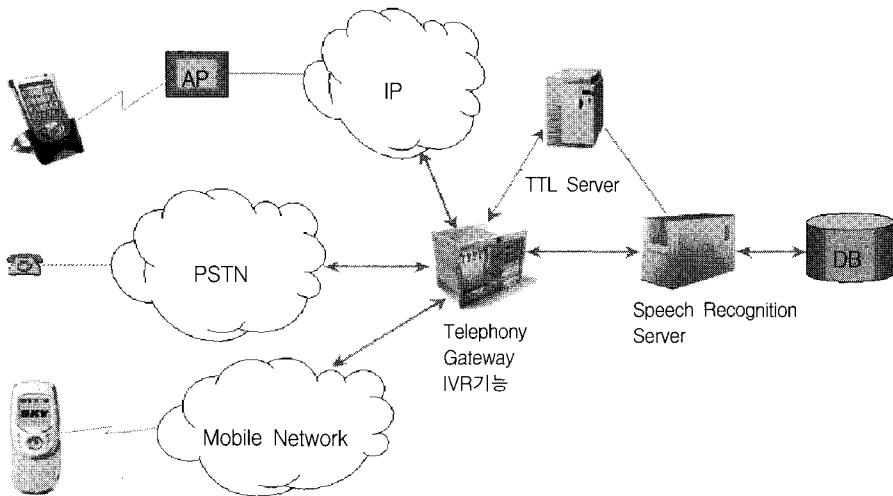
호를 누르면 해당정보를 들을 수 있게 되어 있다. 이미 철도 안내 서비스나 주가 안내 서비스 등이 운용 중에 있다. 철도 안내 서비스를 예로 들어보면, 안내에 따라 곧바로 “서울역”, “통일호” 등을 말하면 자동으로 인식해 열차시간, 잔여좌석, 철도요금 등 필요한 정보를 안내해 준다.

4.2.5 콜센터

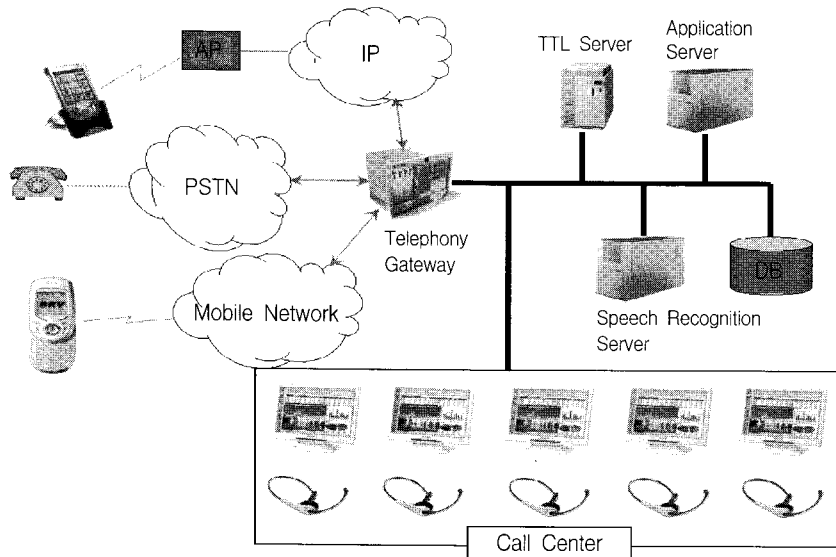
단순히 콜(음성)만 전달해 주는 것과는 달리 현재 시점까지의 고객 이력과 응대 내용을 전달해 줌으로써 다음 응대 요원이 보다 원활한 고객 서비스를 제공할 수 있도록 지원하는 콜센터로서 기존의 콜센터 기능에 음성인식기술과 음성합성기술이 포함되어 있다. 응대 요원이 응대할 수 없는 상황이거나 고정된 패턴의 문의 내용인 경우 음성인식기술로 문의 내용을 파악하고, 미리 데이터베이스에 저장된 내용을 가지고 TTS(Text To Speech) 서버를 거쳐 나오는 합성음성으로 대신 응답할 수 있다.

4.2.6 전화 통역

지금과 같은 속도로 음성인식기술이 발전한다면 향후 2~3년 이내에 외국인과의 통화를 통역해주는 서비스가 등장할 것으로 보인다. 외국인의 음



[그림 10] 전화 안내 서비스



[그림 11] 콜센터

성을 음성인식서버를 통해 텍스트로 변환한 후 번역 엔진을 거쳐 TTS 서버를 통해 우리말로 들을 수 있다. 우리말을 외국어로 바꾸는 경우도 마찬가지이며, 이미 음성 외국어를 통역해주는 PDA용 소프트웨어가 나와 있다[7].

5. 요약

지금까지 VoIP에 음성인식을 접목했을 때 만들 수 있는 서비스를 동향에 맞춰 살펴보았다. 최근 들어 음성인식 기술을 이용한 서비스나 상품들이 홍수처럼 쏟아져 나오고 있으며, 이제는 음성인식 기술이 GUI나 일반 DTMF를 이용하는 사용자 인터페이스(User Interfaces)를 대신할 수 있을 정도로 발전되었고 또 앞으로도 지속적인 발전이 있을 것이라 예상되므로 이제 시작된 VoIP와 음성인식의 접목은 수많은 다양한 종류의 새로운 서비스를 창출해낼 것으로 예상된다.

본 논문의 그림들에서 유선전화, 무선전화, 무선인터넷 등 세 종류의 서비스가 계속 등장한다. 이렇게 세 종류의 서비스를 유지하는 이유는 현재 유무선 전화 및 인터넷 서비스 사업자에 관련되어

각각 다른 비즈니스 모델이 다음과 같이 형성될 수 있기 때문이다.

유선 전화망을 통한 인터넷 서비스는 유선망과 인터넷망을 연동시켜 주는 하드웨어를 개발하는 제조업체와 인터넷 정보를 제공해 주는 정보제공업자 및 서비스를 제공하는 통신사업자가 협력하여 부를 창출하는 비즈니스 형태이다.

무선 전화망을 통한 인터넷 서비스는 무선 전화 사업자들이 이미 무선 인터넷이라는 이름으로 다양한 정보 제공 서비스를 지속해왔다는 점에서 인터넷 정보를 제공해 주는 정보제공업자 및 무선전화 사업자 그리고 서비스 제공업자가 협력하여 매출을 올리는 비즈니스 형태이다.

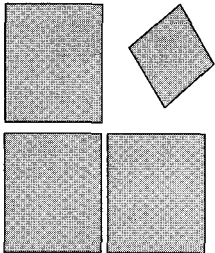
무선 인터넷 서비스는 일반 인터넷 서비스와의 차이는 없으며, 특별히 이동성을 강조한 서비스를 제공한다면 일반적인 인터넷 기반 정보제공자보다 경쟁력을 가질 수 있다.

VoIP는 단독으로 쓰이기보다는 다른 다양한 기술과 서비스와 합쳐졌을 때 그 효과가 커진다. 이제 음성처리기술, 특히 음성인식기술과 함께 사용되는 VoIP 기술의 응용 범위가 어디까지 확대될지 사뭇 기대되는 바이다.

참고 문헌

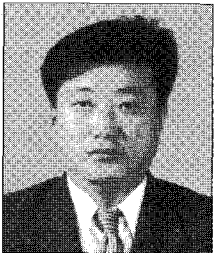
- [1] 김남수, 「음성인식 기술의 개요와 최근 동향」, 뉴미디어통신공동연구소, 1999.
- [2] 김남수, 「휴먼 인터페이스 기술」, Telecommunications Review, 제10권, 제1호(2000).
- [3] 김정환, 이윤철, 이동일, 「VoPN 기술 및 서비스 시장 동향」, 한국전자통신연구원, 2001.
- [4] 김창현, 「VoIP(Voice over Internet Protocol) 개요」, BizSpectrum(<http://www.BizSpectrum.com>), 2001.
- [5] 한국전산원, 「대화형 음성인식 서비스 모델 연구 및 시험」, 2000.
- [6] 뉴욕타임즈, www.nytimes.com/2002/04/25/technology/circuits/25NEXT.html.
- [7] 디지털타임즈, http://www.dt.co.kr/dt_srcview.html?gisaid=2002041602010151589001.
- [8] David Oran, *Understanding Voice over IP*.
- [9] Douskalis, *IP Telephony The Integration of Robust VoIP Services*.
- [10] http://www.saltforum.org/images/SALT_Arch.jpg.
- [11] <http://www.sysbas.com>.

◆ 저 자 소 개 ◆



오 재 삼 (brain@directcall.co.kr)

국민대학교 경영대학원에서 MBA를 취득하였고, 현재 국민대학교 BIT전문대학원 박사과정에 재학 중이다. 현재 모바일라인, 다이렉트콜 등의 대표이사로 재직 중이며 현 관심 분야는 VoIP, 별정통신, 음성인식 등의 분야이다.



윤 용 근 (yoon@kookmin.ac.kr)

The University of Texas at Austin에서 MBA를 취득하였고, Louisiana State University에서 박사학위를 취득하였다. 그리고 국민대학교 정보관리학과의 학과장, 정보관리학부의 학부장을 역임한 바 있고, 현재 국민대학교 경상대학 정보관리학부의 부교수로 재직중이다. 현 관심 분야는 데이터통신, GIS, G-CRM 등이다.