VoiceXML을 이용한 음성 DB 수집 시스템 구현

An Implementation of Speech DB Gathering System Using VoiceXML

김 동 현* 노 용 완** 홍 광 석***
Kim Dong-Hyun Roh Yong-Wan Hong Kwang-Seok

요 약

음성 DB는 음성학, 음성인식, 음성합성등을 연구할 때 가장 기본적으로 필요한 요소이다. 음성 DB의 양과 질이 개발하고자하는 시스템의 성능을 좌우한다고 할 정도로 음성 DB의 중요성 및 역할은 막중하다. 최근 음성포탈을 비롯한 다양한 전화 서비스 기술의 발달로 인하여 전화 음성 DB 수집의 필요성이 시급한 상황이다. 기존 CTI 분야의 IVR 애플리케이션 전화 음성 DB 수집 시스템은 C/C++언어나 전용 개발 도구를 사용하여 이루어져 왔으며, 이로 인하여 각 응용서비스간 자원의 재활용이 어려운 실정이며 많은 인력과 시간을 필요하다는 문제점을 가지고 있다. 그러나, VoiceXML의 전화 음성 DB 수집 시스템은 XML에 내포된 태그형식을 갖는 언어로써 쉽고, 간단한 문법체계를 가지고 있어 조금만 노력을 기울이면 손쉽게 작성할 수 있어 인력과 시간을 절약할 수 있는 장점을 가지고 있다. 또한 단지 웹서버에 연결된 DB의 내용만을 변경함으로써 다양한 전화 음성 DB를 수집할 수 있는 장점을 가지고 있다. 본 논문에서는 음성인식이나 음성합성 등 음성정보처리기술의 개발에 가장 중요한 요소인 음성 DB를 VoiceXML을 사용하여 전화 음성 DB를 수집하는 시스템을 소개한다.

Abstract

Speech DB is basically required factor when we are study for phonetics, speech recognition and speech synthesis and so on. The quantity and quality of speech DB decide the efficiency of system that we develop. therefore, speech DB has an extremely important factor. Recently, development of the various telephone service technique such as voice portal, it is actual condition where the necessity of collection of telephone speech DB. The existing IVR application telephone speech DB collection system used C/C++ language or the exclusive developement tool. Thus it is the actual condition where the recycle of each application service for resources is difficult and have a problem of many labors and time necessity. But, VoiceXML is a language having tag form ipredicated in XML, which has easy and simple grammar system. Therefore, if we make a few efforts we could draw up easily, it has a merit reducing labors and time. Also, VoiceXML has many advantages of various telephone speech DB gathering because of changing contents of DB.

In this paper, we introduce telephone speech DB gathering system which is the most important factor for development of speech information processing technique.

Keyword : VXML, voiceXML, 음성 DB

1. 서 론

우리말의 공학적인 응용을 위해서는 그 기반이 되는 요소 기술로써 음성인식 및 합성으로 대표되는 언어처리기술의 연구가 필요하다[1]. 이러한 음성 및 언어처리기술의 연구를 위해 가장 먼저 확보되어야 할 것이 음성, 언어 및 각종 사전DB 등 국어 정보 베이스이다[2]. 음성 DB는 음성관련 개발자들이 언제든지 재사용이 가능하도록 부가적인 정보와 도큐먼트가 갖추어져 있으며, 컴퓨터로 읽을 수 있는 형태로 구성된 음성자료의 모음이다[3]. 오늘날에 전화는 가장 편리하고 값싼 단말기라고 할 수 있으며, 전화를 사용하여 많은 유용한 정보를 얻을 수 있다. 이러한 전화의 장점으로 인하여 현재 음성 입출력 기술을 이용하여 음성 포

* 준 회 원 : 성균관대학교 대학원 정보통신공학부 석사
super1621@hotmail.com(제 1저자)
** 준 회 원 : 성균관대학교 대학원 정보통신공학부 박사
elec1004@hotmail.com(공동저자)
*** 정 회 원 : 성균관대학교 정보통신공학부 교수
kshong@skku.ac.kr(공동저자)
[2004/03/05 투고 - 2004/04/09 심사 - 2004/10/07 심사 완료]

털 서비스 및 다양한 종류의 전화 서비스를 제공하여 실생활에 많은 편리성과 생산성을 제공해 주고 있다. 웹 포탈 서비스의 원조격인 야후사가 "전화로 거는 야후"라는 이름의 음성 포탈 서비스를 제공한데 이어, 경쟁사인 라이코스도 한층 진보된 음성 인식 기술을 무기로 "전화로 거는 웹서핑"이라는 새로운 개념의 음성 포탈 서비스를 개발 하였고, 온라인 비디오 회사인 Big Star 회사는 고객들이 영화를 사기 위해 전화로 사이트에 접근시키는 NetByTel.com의 음성 응답 서비스와 결합하여 서비스를 제공하고 있다. 이외에도 보이시안닷컴에서는 유.무선을 통해 메일, 일정관리, 주소록 등 개인정보 서비스와 주식시세, 날짜 컨텐트를 운영하며, 보체웹닷컴은 실시간 뉴스, 교통정보, 주식, 경매, 법률정보 등 각종 서비스를 제공한다. 이렇게 음성포탈 서비스는 다양한 분야에 적용되어 많은 이점들을 제공하기에 이를 활용한 서비스 개발이 증가 하고 있다. 따라서 전화 음성 인식, 합성 기술 연구시 시스템의 객관적 성능평가와 알고리즘 개발등의 모든 연구 개발자들이 사용할 수 있는 전화 음성 DB가 필요한 상황이다[4].

기존의 IVR(Interactive Voice Response)과 같은 플랫폼의 음성 DB 수집 시스템은 C/C++언어나 전용 개발 도구를 사용하여 이루어져 왔으며, 이로 인하여 각 응용서비스간 자원의 재활용이 어려운 실정이이었다. 즉, 응용 서비스의 내용이 변경되어지거나 시스템이 바뀌게 되면 다시 프로그램 해야하거나, 적절한 API로의 수정이 필요하다. 이와같은 사항은 응용 프로그램의 개발 시간과 각 서비스 시스템간의 호환성에 문제점이 발생하였으며, 시나리오 변경 요구시 전체적인 프로그램을 재 수행하는 문제점이 발생하였다. 또한 DB 연동을 위해 별도의 기술을 필요로 하는 매우 비효율적인 문제점을 갖고 있다. 이러한 문제점을 해결하고 전화나 음성을 통한 인터넷 컨텐츠의 이용 및 제작이 용이하도록 고안된 언어가 VoiceXML(Voice eXtensible Markup Language)이다[5].

본 논문에서는 위에 언급된 문제점을 해결하기 위해 VoiceXML을 사용하여 음성 DB를 수집하는 시스템을 구현하였다. VoiceXML을 사용하면 음성 DB 수집시 VoiceXML 문서가 시스템으로부터 독립적으로 존재하고 있기 때문에 음성 DB 수집 시나리오의 내용을 수정하거나 변경사항이 생기면 VoiceXML 문서만 수정하거나 변경하면되므로 새로운 기능을 첨가하는데 쉽다. 또한 다른 전화 음성을 수집할 경우에 DB목록만 변경하면 되므로 응용 서비스에 맞게 다양한 전화 음성을 손쉽게 수집할 수 있다[5]. 또한 음성 DB 수집시 전화기에 나오는 발성목록을 듣고 따라하면 되기 때문에 사용자는 시선을 집중하지 않고 편하게 전화음성 DB를 수집할 수 있다. 마지막으로 음성 DB를 수집하면서 실시간으로 녹음되는 음성을 확인후 저장되게 구현되어 있어 기존의 음성 DB 수집후 저장된 파일을 확인하지 않아도되므로 인력과 시간의 낭비를 줄일 수 있는 장점을 제공한다. 본 시스템으로 수집된 음성 DB는 주식 상장사와 주식 거래에 관련된 문장 통합하여 1568개 음성 DB 목록을 선정하였고, 20명의 음성 DB를 수집하였다.

본 논문의 구성은 다음과 같다. 2장에서는 Voice XML의 전반적인 사항에 대해서 알아보고 3장에서는 유무선 전화 사용자와 정보제공자를 연결해 주는 VoiceXML Gateway에 대해 설명한다. 4장에서는 VoiceXML을 이용하여 음성 DB를 수집하는 시스템을 나타내었으며 5장에서는 실험결과, 마지막으로 6장에서는 본 논문의 연구결과를 요약하여 제시한다.

2. VoiceXML

VoiceXML(Voice eXtensible Markup Language)은 웹에서 음성 입력과 출력을 지원하기 위한 목적으로 XML로 설계한 마크업 언어이다[5].

VoiceXML은 AT&T, IBM, Lucent Technology, Motorola 등 4개 기업에 의해 설립된 Voice XML Forum에서 제안하여 W3C(Word Wide Consortium)에서 승인을 받았다. 1999년 8월에

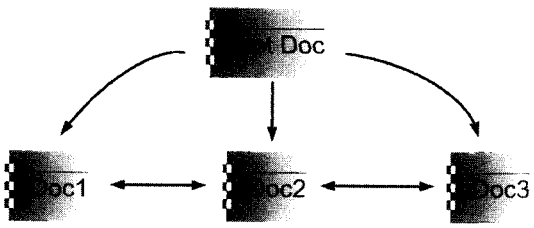
VoiceXML forum에서 버전 0.9를 발표한 후, 2000년 3월 보완하여 버전 1.0을 제안하였다. 2000년 5월에 VoiceXML spec 1.0이 발표되었고, 2001년 8월에 VoiceXML 2.0 Working Draft가 발표되었다. VoiceXML은 대화형 음성 애플리케이션 개발을 위해 고안된 XML 문서 형식의 일종으로 문서의 형식에 맞추어 작성된 VoiceXML 문서는 음성 애플리케이션에서 dialog 진행 방식을 지정하는 시나리오 역할을 한다. VoiceXML을 이용하면 합성음의 출력, 오디오 파일의 발송, 음성 입력의 인식, DTMF(Dual Tone Multi Frequency) 입력의 인식, 하이퍼 링크에 의한 타 서비스 연결, 호(Call)의 전달 및 차단과 같은 음성 응답 시스템 제공등 다양한 서비스를 쉽고 빠르게 제공할 수 있는 환경을 제공한다.

2.1 VoiceXML 문서구조(Document Structure)

2.1.1 시나리오 구성 방식에 따른 분류

- 단일 문서(single document application) : 서비스 시나리오의 구성은 하나의 문서로 이루어져 있다.
- 멀티 문서(multi document application) : 루트 문서에 여러개의 하위 문서들이 서로 정보를 공유하여 하나의 시나리오를 구성하며, 하위 문서들은 루트 문서의 다이얼로그를 사용할 수 있다.

멀티문서의 형태는 그림 1과 같이 하나의 애플리케이션에 여러개의 문서로 구성되어 있다.



〈그림 1〉 멀티 문서

- 서브 다이얼로그(subdialog) : 서브 다이얼로그는 일종의 서브루틴 개념으로 일반적으로 자주 사용되는 다이얼로그를 독립된 문서로 만들어 여러 문서에서 다이얼로그를 재사용할 수 있는 환경을 제공해 준다.

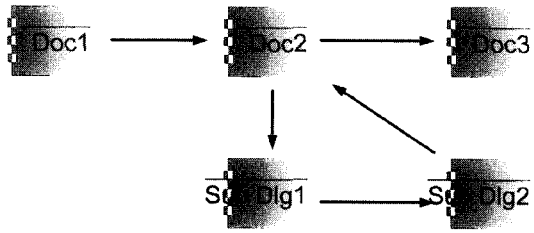
서브 다이얼로그는 다이얼로그 중간에 끼어드는 다이얼로그로써 프로그램에서의 함수와 같이 필요할 때 삽입하여 사용 가능하다. 서브 다이얼로그는 그림 2와 같이 구성되어 있다.

2.1.2 실행 순서에 따른 분류

- Computer directed forms : 다이얼로그(<form>)에 있는 form item은 미리 정의된 순서에 따라 실행된다.
- Mixed initiative forms : 컴퓨터와 사용자 모두 대화의 진행을 능동적으로 변경할 수 있는 다이얼로그 모델로써, form 레벨의 grammar를 필요로 한다. 사용자의 응답에 따라 다이얼로그가 실행되는 순서가 변경될 수 있으며, 여러개의 다이얼로그 form item을 한번에 처리할 수 있다.

2.2 VoiceXML Elements

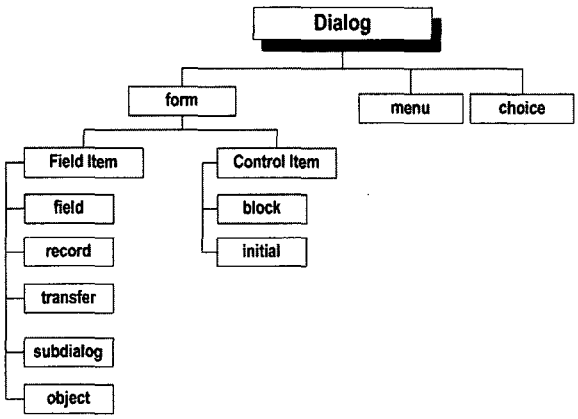
VoiceXML 규격 1.0에는 모두 47개의 요소가 정의되어 있다. 표1은 VoiceXML에서 제공하고 있는 엘리먼트를 기능을 기준으로 구분하여 정리하였다[6].



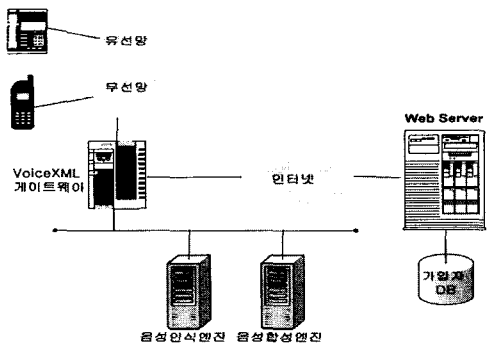
〈그림 2〉 서브 다이얼로그

〈표 1〉 VoiceXML 엘리먼트의 기능별 분류

분류 항목	요 소
문서	<vxml>, <meta>
다이얼로그	<form>, <menu>, <choice>
프롬프트	<prompt>, <enumerate>, <reprompt>
필드	<field>, <option>, <var>, <initial>, <block>, <assign>, <clear>, <value>
이벤트	<catch>, <error>, <help>, <link>, <noinput>, <nomatch>, <throw>
출력	<audio>, <break>, , <emp>, <pros>, <sayas>
입력	<dtmf>, <grammar>, <record>
호 제어	<disconnect>, <transfer>
흐름 제어	<if>, <elseif>, <else>, <exit>, <filled>, <goto>, <param>, <return>, <subdialog>, <submit>
기타	<object>, <property>, <script>



〈그림 3〉 VoiceXML의 다이얼로그 엘리먼트 계층구조



〈그림 4〉 VoiceXML 게이트웨이

VoiceXML에는 폼과 메뉴라는 두 종류의 다이얼로그가 있고, 폼은 사용자의 입력을 받아들이는 필드 아이템과 사용자의 입력을 받아들이지 않는 제어 아이템으로 분류된다.

그림 3은 이러한 VoiceXML의 다이얼로그 엘리먼트 계층구조를 보여준다[7].

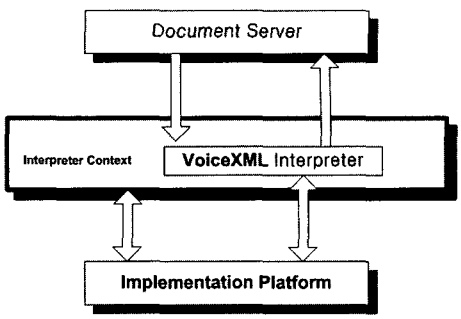
3. VoiceXML Gateway

VoiceXML Gateway는 전화망과 인터넷망을 연결해주는 역할을 하고 HTML의 웹 브라우저처럼 웹 서버에 URL을 전송하고 웹 서버가 보낸 Voice XML 문서를 분석하고 랜더하는 기능을 수행한다.

그림 4의 VoiceXML 게이트웨이의 구성을 보면 사용자가 유.무선 전화로 콜을 하면 전화망을 타고 콜이 VoiceXML 시스템으로 전해지고 시스템내부에서 전화의 콜 신호를 받아들여 인터프리터를 동기화시켜 인터넷망을 통해서 웹서버로부터 VoiceXML 문서를 요청한다[8]. 웹서버에서는 해당 VoiceXML 문서를 찾아서 VoiceXML 시스템으로 문서를 전송하면 해당 문서를 인터프리터에서 파싱하고, 시나리오를 진행시켜서 사용자에게 TTS를 통해 만들어진 합성음성을 전달한다.

3.1 VoiceXML 해석기

VoiceXML 해석기는 서버가 보낸 VoiceXML



〈그림 5〉 VoiceXML Architecture 모델

문서를 해석하고 사용자와의 상호작용을 수행하기 위해 실행 플랫폼을 제어하는 기능을 한다. 그림 5는 VoiceXML의 구조적 모델을 나타내었고, 각각의 역할에 대한 설명은 다음과 같다.

3.1.1 Document Server

웹 서버를 이용하여 HTTP 클라이언트 응용 프로그램에 해당하는 URI(Uniform Resource Identifier) 형태로 전송되는 VoiceXML Interpreter가 요청한 문서나 자원을 인터프리터에게 전송하는 역할을 한다.

3.1.2 VoiceXML Interpreter

VoiceXML 문서를 적재하고 그 내용을 해석해 실행하는 역할을 담당하고 VoiceXML 실행 환경의 가장 핵심적인 요소이다. VoiceXML 문서로 표현된 음성애플리케이션 시나리오를 해석하여 다이얼로그, 문법, 이벤트, 오디오출력, 콜 제어, 오디오 입력, 흐름 제어와 관련된 47종의 각 태그에 설정된 기능에 따라 문서 실행의 순차적 흐름을 제어하고, 음성 입출력 내용을 결정해 음성 플랫폼에 필요한 명령을 내린다. 또한 문서 서버를 통하여 필요한 자원을 다운로드하거나 다른 문서로 전이하는 등 VoiceXML 문서 실행을 총괄적으로 제어하는 역할을 한다.

3.1.3. VoiceXML Interpreter Context

웹 서버와 HTTP를 통해 데이터를 주고 받는 HTTP 통신을 담당하고, VoiceXML 해석기가 VoiceXML 문서를 해석하게 하고, VoiceXML 해석기와는 독립적으로 실행 플랫폼과 상호작용하는 역할을 한다.

3.1.4 Implementtation Plaform

VoiceXML Interpreter Context에 의해 제어되며 하드웨어와 소프트웨어를 모두 포함하며, 전화 수신 기능, 전화호 전환 기능, 음성인식 기능, 음성합성 기능, 음성과 오디오 재생 기능, 음성과 오디오 녹음 기능 등을 수행한다.

3.2 음성인식

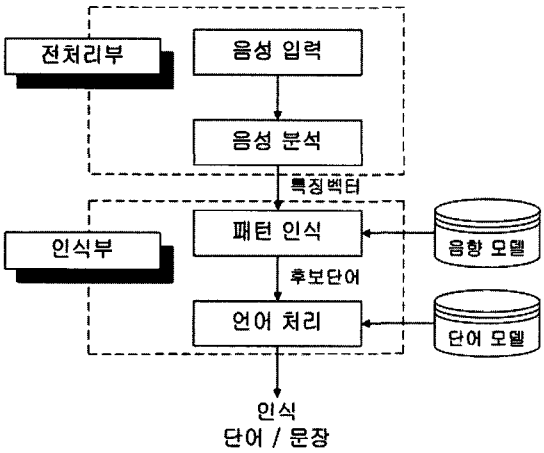
음성인식이란 음성에 포함된 음향학적 정보로부터 음운, 언어적 정보를 추출하여 이를 기계가 인지하고 반응하게 만드는 일련의 과정을 말한다[9]. 음성은 가장 쉽고 자연스러운 의사 전달의 수단인 동시에 음성의 입력 및 전달 과정에서 고가의 장치가 필요 없다는 장점을 가지며 man-machine 인터페이스의 응용으로 다양한 분야에서 그 효용성을 인정 받고 있다[10].

3.2.1 인식의 분류

음성인식은 사용 기술과 응용 분야에 따라 여러 가지 기준으로 분류할 수 있다. 인식 주체에 따라 화자 종속(Speaker Dependent), 사용자 그룹에 대해 음성 속성을 표현한 다중 화자(Multi Speaker), 화자 독립(Speaker Independent), 화자 적응(Speaker Adaptation) 등이있다. 음성인식의 단위에 따라 짧은 단어 사용이나 간단한 음성 제어 등의 경우에 사용되는 고립단어(Discrete Word) 방식과 연속 음성을 이용하는 연속 음성(Continuous Speech) 방식이 있다. 또한 인식의 어휘량에 따라 수배 어휘이내의 단어를 인식하는 소용량 어휘(Small Vocabulary) 시스템과 수만 단어까지 지원하는 대용량 어휘(Large Vocabulary) 시스템이 있다. 새로운 어휘를 추가하는 방법에 따라 분류를 하면, 인식을 위한 새로운 어휘에 대해 항상 등록과 학습을 해야 하는 어휘 종속(Vocabulary Dependent) 시스템과 등록된 어휘로부터 특징을 추출하여 새로운 어휘를 생상하는 어휘 독립(Vocabulary Independent) 시스템으로 분류할 수 있다[11].

3.2.2 음성인식 시스템

그림 6의 음성인식 시스템의 구성도를 보면 음성인식 시스템은 크게 전처리부와 인식부로 나눌 수 있다. 전처리부에서는 사용자가 발성한 음성으로부터 인식에 필요한 특징 벡터를 추출하고, 인식부에서는 음성 데이터베이스로부터 훈련한 기준 패턴과의 비교를 통해서 인식 결과를 얻게 된다. 복잡한 구조의 음성을 인식할 때에는 언어모델을 이용한 언어 처리 과정을 통해 최종 인식 결과를 출력한다[12].



〈그림 6〉 음성인식 시스템의 구성 요소

3.3 음성합성

Text 정보를 실제 인간의 음성과 유사한 명료성과 자연성을 가진 음성으로 합성하여 정보를 전달하는 것이다.

3.3.1 음성합성 단위

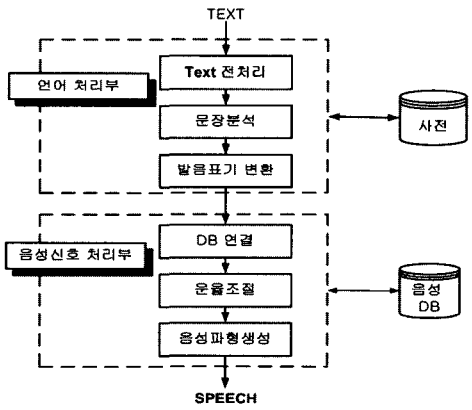
음성 합성에 필요한 음성 단위로는 음소, 음절, 반음절, 다이폰, 단어, 문장 등이 있다. 단어나 문장 단위는 음성의 연결 특성을 가지고 있기 때문에 고품질 합성음을 생성 할 수 있으나 무제한 합성을 위해서는 가능한 모든 음성을 발성이 가능하도록 해야 하는데 이는 실제로 불가능하다. 따라서 단어나 문장 단위는 제한된 범위의 합성에 유리하다. 음소는 음성의 가장 기본적인 요소이므로 작은 데이터로 무제한 합성이 가능하지만 음성의 연결 특성을 잘 표현하지 못하게 되므로 합성음의 음질이 떨어지는 단점이 있다. 반음절과 다이폰은 이러한 단점들을 보완해주기 때문에 음성 합성 단위로 주로 이용 되었다[13]. 그러나 최근 음성합성은 자연스럽게 발성된 문장 음성데이타를 대용량으로 녹음, 이로부터 합성단위를 추출하여 사용하는 기술이 널리 적용되고 있다[14].

3.3.2 음성합성 방식

음성합성 방식은 실제 조음 기관을 모델링하여 구현한 조음합성 방식과 합성단위의 포먼트 성분을 연결하여 합성하는 포먼트(formant)합성 방식, 음성신호를 분석하여 필요한 파라미터를 저장하여 합성하는 방식인 LPC 합성방식, 미리 녹음된 합성단위를 연결하여 합성음을 만드는 PSOLA 합성 방식으로 나눌 수 있다.

3.3.3 음성합성(TTS) 시스템

문장-음성 변환(TTS) 시스템은 입력 문장을 언어 처리를 통해 음성 위주의 기호열로 재구성한 다음 이를 합성해 내는 시스템이다. 일반적인 TTS시스템은 그림7과 같다.



〈그림 7〉 음성합성 시스템의 기본 구조

그림 7의 음성합성 시스템의 기본 구조를 보면 합성음을 만들어주는 입력으로 TEXT를 입력하면 그 TEXT에 대해 언어 처리부와 음성신호 처리부를 거쳐 합성음을 만들어 내는 전체적인 구조를 가진다.

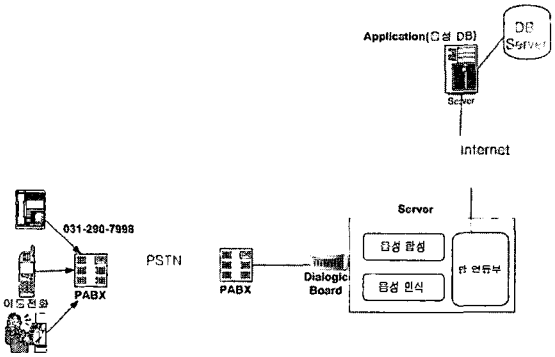
4. VoiceXML 기반 음성 DB 구축 시스템

VoiceXML 기반 음성 DB 구축 시스템은 음성인식과 음성합성 엔진을 사용하여 사용자와 말을 주고받듯이 음성 DB를 수집하는 대화형 음성 DB 수집 시스템이다.

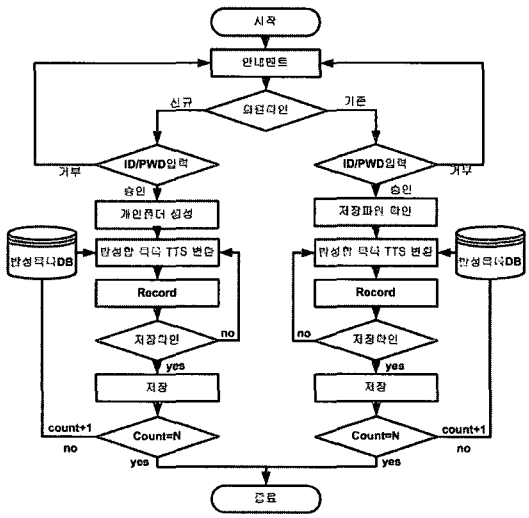
4.1 전체 시스템

그림 8과 9는 본 논문에서 설계한 시스템의 전체 블록도와 흐름도를 나타내고 있다. 그림8과 9를 보면 DB 사용자가 전화를 사용하여 콜하면 PABX와 전화망을 거쳐 Dialogic Board가 콜신호를 받아들인다. 그후 인터프리터를 동기화시켜 인터넷망을 통해 웹서버로부터 VoiceXML문서를 요청한다. 웹서버에서는 해당 VoiceXML 문서를 찾아서 VoiceXML Gateway로 문서를 전송하면 해당 문서를 인터프리터에서 파싱하여 안내멘트가 필요한 경우 TTS를 통해 사용자에게 전달된다. 사용자는 안내멘트를 듣고 자신의 ID와 비밀번호를 입력한다. User 데이터베이스로부터 ID와 비밀번호를 검색후 데이터베이스에 등록된 사용자일 경우 Data 데이터베이스에서 발성목록을 가져와 TTS로 변환하여 사용자에 들려준다. 사용자는 전화기에서 나오는 소리를 듣고 발성하면 발성된 소리를 사용자에게 들려준후 저장의 여부를 사용자에게 물어본다. 마지막 단계로 사용자는 저장여부를 결정한다. 이러한 일련의 과정을 통해서 전체적인 음성 DB를 수집하는 시스템이다. 다음은 각각의 모듈을 통하여 자세하게 설명하겠다.

그림 9에서 기존회원과 신규회원을 따로 구분하여 나타낸 이유는 사용자의 편리를 제공하고, 정확한 녹음 파일을 받기 위해서 구분하였다. 사용자가 전화로 한번에 1568문장을 받기에는 시간을 너무 많이 소비하여 지루함을 느끼고 집중력을 떨어뜨려 정확하지 못한 음성 파일을 녹음할 수 있다. 이러한 경우에 사용자는 전화를 끊고 수회에 걸쳐서 지금까지 녹음받은 이후의 리스트부터 녹음을 할 수 있도록 변수를 설정해 두었다. 지금까지 받은 발성 DB 리스트는 자동 저장되고 다음에 기존회원으로 로그인을 할때 저장된 리스트 이후부터 음성을 녹음할 수 있게 구현하였다. 이는 자신의 컨디션에따라 녹음받는 파일의 양을 자신이 조절할 수 있게하여 사용자에게 편리성과 정확한 녹음 파일을 받을 수 있는 환경을 제공한다.



〈그림 8〉 음성 DB 수집 블럭



〈그림 9〉 음성 DB 수집 흐름도

4.2 신규회원 모듈

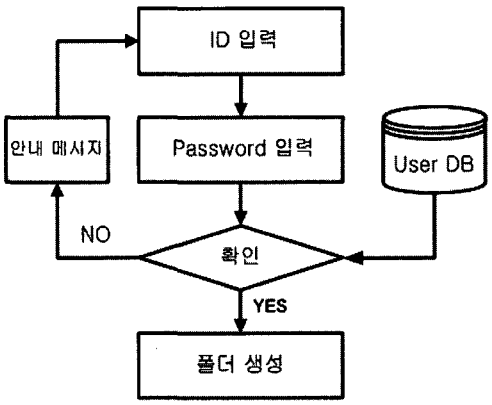
신규회원 모듈에서는 사용자 인증 과정과 사용자 ID명의 고유폴더를 생성하는 두개의 단계로 나타낼 수 있다. 사용자 인증 과정에서는 다른 임의의 사용자가 녹음하는 행위를 방지하기 위해서 등록된 사용자만이 녹음을 할 수 있게 사용자 인증 과정을 두었다. 사용자의 ID와 password는 전화상의 인식의 오류를 줄이고, DTMF 입력 기능을 사용하기 위해서 숫자의 조합만 유효하도록 설정해 놓았다. 사용자는 신규회원 가입시 자신이 원하는 아이디와 패스워드를 전화기의 DTMF로 입력하거나 차례로 발성하면 DB에 자동 저장하게 된다. 폴더 생성 단계에서는 사용자 ID명의 폴더가 생성된다. 다른 사용자와의 구별을 위해서 폴더 생성시 기존의 동일한 폴더가 있는지 확인하여 기존의 폴더가 존재하면 오류 메시지를 보내고 존재하지 않으면 사용자 ID명의 폴더가 생성되도록 설계하였다.

그림10은 신규회원 확인 흐름도를 나타내고 있다. 신규회원 사용자가 로그인을 할때 ID와 password를 입력한후 User DB에 기존에 등록된 사용자인지 확인하기 위해 UserDB를 검색한다. User DB에 등록된 사용자이지 않을 경우는 사용자 DB에 ID와 password가 저장되며 그 다음으로 사용자 ID명의 폴더가 생성된다. User DB에 등록된 사용자일 경우에는 ID와 password를 확인하라는 안내멘트가 나간후 다시 처음으로 돌아가 반복된 과정을 밟게 된다.

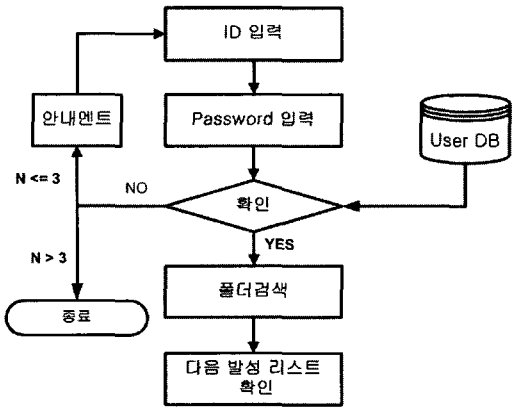
4.3 기존회원 모듈

기존회원에서는 사용자 DB에 저장된 사용자인지 확인하는 사용자 인증과정과 기존에 생성된 사용자명의 폴더를 검색하고 중단된 이후의 발성 리스트를 검색하는 단계로 나타낼 수 있다.

기존회원 모듈에서는 그림 11과 같이 사용자 인증 과정에서 ID와 password를 입력한후 user DB를 검색하여 신규회원 단계에서 등록한 사용자인지 확인하는 과정이다. ID나 password가 일치하지 않으면 3번의 반복과정을 거쳐 사용자 인증 과정을 수행할 수 있게 하였다. 3번 연속으로 일치하지 않으면 강제로 프로그램을 종료하도록 설정해 놓았다. 이렇게 설정해 놓은 이유는 신규회원 과정에서 user DB에 등록하지 않은 사용자이므로 인증되지 않은 사용자가 녹음하는 행위를 방지하기 위한 과정이다. 사용자 인증 과정을 거친후 다음으로 신규회원 과정에서 생성된 폴더가 있는지 검색하는 과정을 가진다. 신규회원 과정에서 반드시 사명자ID명의 폴더가 반드시 생성되기



〈그림 10〉 신규회원 확인 흐름도

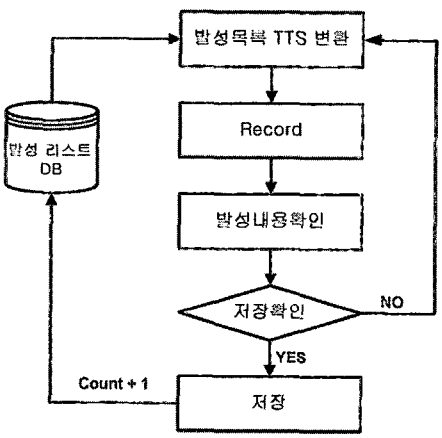


〈그림 11〉 기존회원 확인 흐름도

때문에 사용자 ID명의 폴더가 검색되지 않는다면 신규회원 과정을 올바르게 거친 사용자가 아니다. 이러한 경우에 처음부터 다시 시작하라는 안내메세지가 나간다. 사용자 ID명의 폴더가 검색되면 안내메세지는 사용자의 이름을 부른후 "xx님반갑습니다" 라는 멘트가 나온후 사용자 ID명의 폴더로 이동하여 다음 발성할 목록을 발성목록DB에서 가져와 발성할 목록이 TTS로 변환된다. 사용자 인증과정과 폴더 검색 기능의 목적은 user DB에 등록하지않은 사람의 녹음행위를 방지하고, 사용자 한 사람마다 자신의 ID명의 폴더가 하나만 생성하기 위해서이다. 그리고, 그 폴더안에 다른 사람의 음성DB가 저장되지 않고 자신의 음성DB만 저장되게 하기위해 이러한 단계를 두었다.

4.4 음성 녹음 모듈

그림 12와 같이 음성 녹음 모듈에서는 발성목록을 발성목록 DB로부터 가져와 TTS(Text-To-Speech)로 변환하여 사용자에 들려준다. 사용자는 전화기에서 나오는 발성목록 리스트 소리를 듣고 "삐"소리후 똑같이 발성을 하면 된다. 발성을 제대로 발성했는지 확인하기 위해 전화기에서 사용자가 발성한 소리를 들려주고 "저장하시겠습니까?" 라는 안내멘트가 나온다.



〈그림 12〉 음성 녹음 모듈

사용자는 제대로 발성하지 못했다고 생각되면 "아니오"라고 대답하면 저장되지 않고, 발성목록 TTS변환단계로 되돌아가 같은 발성목록이 전화기를 통해 사용자에게 들려준다. 사용자가 "아니오"라고 대답하였을 경우에는 이와같이 반복적인 과정을 거치게 되지만, 제대로 발성하였을 경우에 "예"라고 대답하면 사용자의 발성이 사용자명의 ID폴더에 발성목록 번호명의 파일 이름으로 저장된다. 저장된후 발성목록 DB의 카운트가 증가하여 다음 발성목록으로 넘어간다. 기존의 음성 DB 수집시스템에서는 DB 수집후 제대로 발성된 DB인지 확인을 해야하는 단계를 가지고 있어 시간과 인력을 낭비하는 문제점을 가지고 있었으나, 본 시스템에서는 실시간으로 사용자가 발성한 소리를 확인하는 단계를 가지고 있어 시간과 인력의 낭비를 줄일 수 있는 장점을 가지고 있다. 또한, 저장단계에서 저장되는 파일이 다른 사용자가 녹음한 파일과의 섞이지 않도록 사용자의 ID명의 폴더에 저장되도록 구현하였다.

5. 실험 및 결과

5.1 개발 환경

시스템의 개발 환경은 윈도우 2000 professional을 사용하였고, 웹서버는 윈도우 NT전용 Web 서버인 IIS(Internet Information Server) 5.0을 사용하였다. 사용자 로그인 처리와 폴더의 생성과 파일의 이동을 위하여 ASP(Active Server Page)를 이용하여 시스템을 구축하고, 사용자가 발성한 목록을 DB로 저장하기 위해 Microsoft Access를 사용하였다. Call 제어를 위해서 Intel Dialogic 41JCT/LS를 사용하였고, Interpreter, 음성인식기와 합성기는 KT의 HUVOICE 1.0을 사용하였다.

5.2 전화 음성 DB 구축

전화 음성 DB 구축을 위한 수집대상 어휘는

〈표 2〉 전화 음성 화자 통계(지역별)

	남자	여자	계
서울.경기	4	2	30%
충청	2	2	20%
전라	2	2	20%
경상	4	2	30%
계	60%	40%	100%

〈표 3〉 전화 음성 화자 통계(연령별)

	남자	여자	계
20세-25세	4	4	40%
26세-30세	4	4	40%
31세-40세	4	0	20%
계	60%	40%	100%

증권 거래에 관련된 상장사 1558종목과 증권 거래에 관련된 문장 10문장을 선정하였다.

본 시스템의 발성화자 선정기준은 화자의 성별(남.여), 발성자의 주거 지역(서울.경기, 충청, 전라 , 경상), 화자의 연령 등의 균형을 고려하여 총 20명의 음성 데이터를 수집하였다.

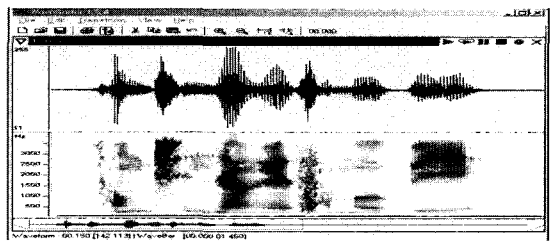
표 2와 표 3은 전화음성 DB 수집에 참여한 발성화자들에 대한 지역별 연령별 통계를 나타내고 있다.

〈표 4〉 본 시스템 구현사항 정리

기 능	목 적	구 현 방 법
음성을 이용한 발성목록 안내	자연스러운 발화유도 및 사용자 편이성 증진	VoiceXML을 이용한 발성목록생성
발성 화자별 파일 저장	DB 규모의증대로 인한 파일관리의 어려움 해결	ASP를 이용한 발성화자별 Fold 생성 및 분류저장
발성목록 및 미발성 목록에따른 발성음 안내기능	DB 구축시, 발성목록의 방대함으로인하여 다수에 걸쳐 DB구축 서버에 접속, 발성목록을 발성	ASP를 이용한 발성목록, 미발성목록 및 발성 저장장소의 구분

표 4는 본 시스템의 구현 사항을 간략히 정리하여 나타내었다.

그림 13은 본 시스템으로 "성신양회삼우비" 종목을 수집한 파일의 파형과 스펙트럼을 나타낸 그림이다. 8k 샘플링에 8bit μ-law PCM 방식을 이용하였으며 자동 수집된 음성 데이터는 본 연구실의 서버컴퓨터의 하드 디스크에 자동 저장된다. 음성 DB를 수집할 때 사용자는 발성할 목록을 직접 보지 않고 전화 음성을 통해서 목록을 들려주고 따라 발성하여 수집할 수 있어 사용자는 편하게 전화 음성 DB를 수집할 수 있다. 그러나, 일부 발성목록에서 자연스럽지 못한 합성음의 출력은 사용자가 전화음성을 듣고 확실한 발성내용을 확인하지 못하는 경우가 있다. 따라서 사용자는 정확하지 못한 전화음성일 경우에는 발성 리스트를 보고 발성해야 하는 단점을 가지고 있다. 그러나, 앞으로 더욱 자연스러운 음성합성기가 제공된다면 사용자는 발성리스트를 보지 않고, 전화기의 음성을 듣고, 전화음성 DB를 수집할 수 있어 발성자에게 더욱 편리한 시스템이 될 것이다.



〈그림 13〉 수집된 파일의 파형과 스펙트럼

본 시스템이 제공하는 주요 특징은 XML에 내포된 태그형식을 갖는 언어로써 쉽고, 간단한 문법체계는 조금만 노력을 기울이면 손쉽게 작성할 수 있어 인력과 시간을 절약할 수 있는 장점을 가지고 있다. 또한 단지 웹서버에 연결된 DB의 내용만을 변경함으로써 다양한 전화 음성 DB를 수집할 수 있는 장점을 가지고 있다.

6. 결 론

현재 음성 입출력 기술을 이용한 음성포털 및 다양한 전화 서비스를 제공으로 실생활에 많은 편

리성과 생산성을 제공해 주고 있다. 이러한 다양한 응용분야에 따라 각기 다른 형태의 음성DB의 구축이 필요하다. 기존 CTI 분야의 IVR 애플리케이션 전화 음성 DB 수집 시스템은 C/C++언어나 전용 개발 도구를 사용하여 이루어져 왔으며, 이로 인하여 각 응용서비스간 자원의 재활용이 어려운 실정이며 많은 인력과 시간을 필요하다는 문제점을 가지고 있다. 그러나, VoiceXML을 이용하면 음성 DB 수집시 VoiceXML 문서가 시스템으로부터 독립적으로 존재하고 있기 때문에 음성 DB 수집시 음성 DB 수집 시나리오의 내용을 수정하거나 새로운 기능을 추가하고자 할때 VoiceXML 문서만 수정하면되므로 새로운 기능을 첨가하는데 쉽다. 또한, XML에 내포된 태그형식을 갖는 언어로써 쉽고, 간단한 문법체계를 가지고 있어 조금만 노력을 기울이면 손쉽게 작성할 수 있어 인력과 시간을 절약할 수 있는 장점을 가지고 있다. 또한 단지 웹서버에 연결된 DB의 내용만을 변경함으로 다양한 전화 음성 DB를 수집할 수 있는 장점을 가지고 있다. 또한, 음성 DB를 수집하면서 실시간으로 녹음되는 음성을 확인 후 저장되게 구현되어 있어 기존의 음성 DB 수집 후 저장된 파일을 확인하지 않아도되므로 인력과 시간의 낭비를 줄일 수 있는 장점을 가지고 있다. 마지막으로 사용자는 음성 DB 수집시 발성리스트를 보지 않고 전화기의 음성을 듣고 음성을 수집할 수 있어 사용자는 편하게 음성을 수집할 수 있는 장점을 제공한다. 본 시스템의 성능을 테스트하기 위하여 발성자의 개인차, 지역차 등을 고루 흡수할 수 있는 전화 음성을 수집하였다. 선정된 발화성화자 수는 20명이고, 수집대상어휘는 증권거래에 관련된 1568문장을 선정하여 전화 음성 DB를 수집하였다. 수집된 전화음성 DB를 분석한 결과 전화망의 노이즈로 인하여 깨끗한 음질은 아니지만 전화망 환경에 적절한 음성DB를 수집하였다. 현재 본 시스템으로 수집된 음성 DB는 증권거래에 관련된 음성 DB이지만 앞으로 다양한 응용분야에 맞게 전화음성 DB를 수집할 예정이다.

참 고 문 헌

[1] 이용주, "한국어 음성언어정보처리와 음성 데이터베이스", 한국어정보처리 소식 제 2권 특별기고, 1994.

[2] 김종교, "한국어 표준 전화 음성 데이터 베이스 구축", 한국음향학회 음성.음향 WORKSHOP, pp.5-9, 1995

[3] 이용주, 김상훈 , "대규모 공통 음성 DB 구축 현황" ,대한전자공학회지, vol.30, No.7, pp.749-757, 2003.

[4] Weibin Zhu, Wei Zhang, Qin Shi, Fangxin Chen, Haiping Li, Xijun Ma, Liqin Shen, "Corpus building for data- driven TTS System, Proceedings of 2002 IEEE Workshop on, pp.199-202, 2002.

[5] W3C, "Voice Extensible Markup Language VoiceXML"(Version 1.0), http://www.w3.org/TR/voicexml, 2000.

[6] Larson, "VoiceXML and the W3C speech interface framework", multimedia IEEE, vol.10, No.4, pp.91-93, 2003.

[7] 박섭형, "VoiceXML: 음성 웹 애플리케이션 구축을 위한", pp.45 , 한빛미디어, 2001.

[8] Quiane Ruiz , Manjarrez Sanchez, "Design of a VoiceXML gateway", Proceedings of the Fourth Mexican International Conference on, pp.49-53, 2003.

[9] 이건상, 양성일, 권영헌, "음성인식", pp.25, 한양대학교 출판부, 2001.

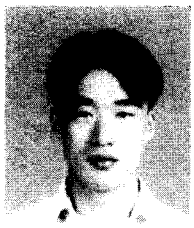
[10] Juang, B.H. , Tsuhan Chen, "The past, present, and future of speech processing", Signal Processing Magazine IEEE, Vol.15, No.3, pp.24-pp.48, 1998.

[11] 윤재선, "한국어 음성인식 Diction System의 구현", pp.9, 성균관대학교 박사학위 논문, 2001.

[12] 김동환, "효율적인 음성인식을 위한 SCHMM

의 성능 개선 방법", pp.2, 성균관대학교 석사학위 논문, 2001.
[13] 이현구, "한국어 문장-음성합성 시스템에서 감정표현과 발성속도 제어에 관한 연구", pp.11, 성균관대학교 박사학위 논문, 1999.
[14] Rutten.P, Coorman.G, Fackrell.J, Van Coile.B, " Issues in corpus based speech synthesis", IEE Seminar on, pp.1-pp.7, 2000.

◐ 저 자 소 개 ◑



김 동 현(Kim Dong-Hyun)
2003년 강원대학교 컴퓨터 정보통신공학부 졸업(학사)
2003년 ~ 현재 성균관대학교 대학원 정보통신공학부 재학(석사)
관심분야 : 음성인식, 음성합성, VoiceXML
E-mail : super1621@hotmail.com

노 용 완(Roh Yong-Wan)
2001년 남서울대학교 정보통신공학과 졸업(학사)
2003년 성균관대학교 대학원 정보통신공학부 졸업(공학석사)
2003년 ~ 현재 성균관대학교 대학원 정보통신공학부 재학(박사)
관심분야 : 음성인식, 음성이해, 신호처리
E-mail : elec1004@hotmail.com

홍 광 석(Hong Kwang-Seok)
1985년 성균관대학교 전자공학과 졸업(학사)
1988년 성균관대학교 대학원 전자공학과 졸업(공학석사)
1992년 성균관대학교 대학원 전자공학과 졸업(공학박사)
1990년 ~ 1993년 서울보건전문대학 전산정보처리과 전임강사
1993년 ~ 1995년 제주대학교 정보공학과 전임강사
1995년 ~ 현재 성균관대학교 정보통신공학부 교수
관심분야 : 오감인식, 융합 및 재현, HCI
E-mail : kshong@skku.ac.kr