

# 웹 환경에서 VoiceXML을 이용한 음성인터페이스 활용방안

장민석 방초균<sup>0</sup>

국립군산대학교 컴퓨터정보과학과  
msjang@kunsan.ac.kr, dain0127@hanmail.net<sup>0</sup>

## A Method For Utilizing Voice Interface in Web Environment Using VoiceXML

Min-Seok Jang Cho-Kyun Bang<sup>0</sup>

Detp. of Computer Information Science, Kunsan National University

### 요 약

현재의 웹 환경은 HTML로 구성이 되어있고 이로인해 하이퍼링크를 따라가기 위해 마우스 클릭을 통해 작업하는 GUI환경이 주를 이룬다. 하지만 이러한 방법은 인간이 가장 손쉽게 사용하는 음성과 비교해 볼 때 상당히 불편한 축에 속한다. 이를 해결하기 위해 현재 부각되는 음성인식 기술과 전화기를 통해 정보를 제공하고자 하는 XML의 파생인 VoiceXML을 이용하여 현재 HTML이 주류를 이루는 웹 환경을 VoiceXML을 이용한 음성인터페이스 환경을 마련하고자 한다.

### 1. 서 론

XML은 현재까지는 SGML의 대표적인 단점들을 대부분 보완하였고, 특별한 단점이 발견되지 않고 있어 전자문서를 사용하는 대부분의 응용분야에서 XML을 전자문서의 표준으로 적용하려는 움직임이 활발하다. 또한 XML의 유연한 확장성으로 인해 채널정의포맷 (Channel Definition Format, CDF), 화학마크업언어 (Chemical Markup Language, CML), Open Financial exchange(OFX) 등의 많은 파생언어가 생성되어지고 있다. VoiceXML 또한 그러한 맥락에서 생성된 언어이다 [1][2].

현재 VoiceXML은 VoiceXML 포럼에서 AT&T, IBM, 루슨트 테크놀로지, 모토로라 등 정보통신 분야의 4개 거대 기업을 중심으로 표준화가 진행중이며 올해 10월에 ver 2.0의 초안이 나온 상태이며 이 또한 곧 표준으로 채택될 것으로 전망된다 [2]. 이러한 VoiceXML은 현재 뚜렷한 응용 프로그램들이 나와있는 상태는 아니나 현재 각 음성정보 통신업체를 비롯한 IT업체에서 VoiceXML을 대비하기 위한 각종 솔루션들을 개발하기 위해 노력하고 있으며, 이는 곧 다가올 VoiceXML의 파장효과를 클 것이라는 전망을 남기에 충분하다. 하지만 아직까지 국내에서는 이렇다 할만한 연구가 미비한 실정이며, 이 또한 VoiceXML의 특성보다는 XML의 특성에 중점을 둔 연구들이 대부분이다. VoiceXML은 XML의 서브 셋이라 볼 수 있지만, XML과는 다르게 장치에 특성을 고려하는 특성을 가지고 있다. 이는 VoiceXML의 용도가 음성을 이용해서 일반 전화기 등을 통해서 정보를 획득하고자 하는 방법이기 때문이다. 지금까지의 VoiceXML은 기존의 전화기를 통해서 음성정보 서비스를 하기 위한 방편으로만 개발이 되어졌다. 물론 웹의 정보를 VoiceXML을 통해서 전화기로 전달받을 수 있는 방법이 개

발되어지고 있으나, 웹에서 사용되어지는 정보는 전화기를 이용한 사용보다는 데스크탑 또는 PDA 등의 단말기에서 더욱 많이 쓰이리라 생각되어진다. 특히 PDA등에서는 제한된 입력 인터페이스를 음성이 효과적으로 뒷받침해줄 수 있으리라 기대가 된다. 그러므로 앞으로의 연구는 전화기를 이용한 정보의 접근 또는 VoiceXML을 이용하지 않은 브라우저/ 음성 응용프로그램과는 달리 현재의 각종 응용프로그램 또는 웹 페이지 내에서 VoiceXML을 접목시켜 전화기뿐만 아니라 기타 단말장치에서도 편리한 정보의 접근/ 응용프로그램의 사용을 할 수 있도록 하고자 한다.

본 논문은 다음과 같은 순서로 구성되어있다.

2장에서 음성인식 기술과 그에 따른 음성 인터페이스에 관해 알아보고, 3장 VoiceXML 구조를 설명하며, 4장에서는 VoiceXML이 음성인터페이스로 사용하는 방법을 제시한다.

### 2. 음성인식 기술/ 음성 인터페이스

우리가 컴퓨터나 기타 정보를 획득할 수 있는 장치를 사용할 때 불편한 점 중 하나는 마우스 또는 키보드 등을 통한 제한된 방법을 이용하여 명령들을 수행하는 것이다. 이를 해결하고자 음성인식 기술들이 개발되었으나 다음의 두가지 이유로 인해 널리 사용되지 못하고 있다. 첫째, 기술이 실용화하기에 충분한 수준으로 발전하지 못한 것이다. 응용분야에 따라 차이가 있으나 음성인식을 실용 제품에 사용하려면 인식률이 약 95% 이상 되어야 한다. 그러나 사람이 자유롭게 말하는 것을 95% 이상 인식하는 것은 현재로서는 불가능하다.

둘째로는 현재까지 개발한 기술을 실용시스템에 접목하는 기술이 부족하다는 점이다. 응용분야별 음성인식 요구사항이 무엇인지, 또 음성인식을 사용자 인터페이스로 사용할 때 어떤 방식을 사용해야 효과적인지 등의 연구

가 부족하다.

그리고, 이러한 음성인식기술들의 종류로는 화자중속, 화자독립 등이 있다. 우선 특정 사용자의 음성만을 인식하는 화자중속 음성인식과 여러 사람의 음성을 인식하는 화자독립 음성인식을 들 수 있다. 화자중속 음성인식은 화자독립 음성인식에 비해 인식률이 높아 실용화하기에 유리하다. 그렇기 때문에 현재 일부에서 응용제품들을 만들어 사용하고 있다. 또한, 음성인식의 단위에 따라 고립단어 및 연속음성인식으로 나눌 수 있다.

음성인식기술을 응용한 시스템들이 현재 많은 개발이 이뤄져 있는 상황이다. 마이크로소프트도 윈도우 OS에 곧 음성인식시스템을 포함할 예정이다 [3]. 특히 PDA는 키보드를 쓸 수 없어 음성인식기능이 절실히 요구된다. 애플은 뉴턴 PDA에 이를 이용하고 있으나 아직 인식률이 사용자의 요구 수준에 이르지 못하고 있다. 음성만을 사용하는 전화이용서비스는 114 안내시스템처럼 항상 사람이 처리해야 한다. 그러나 이는 막대한 인건비가 소요되므로 음성인식의 필요성이 절실한 분야 중 하나이다. 실제 세계 각국의 전화회사가 가장 많이 음성인식 연구에 투자를 하고 있으며, 음성 다이얼링 등 많은 응용 분야가 실용화되고 있다. 이 밖에 주요 전화회사들도 자동차 내의 음성인식, 공장자동화를 위한 명령어인식 등 음성언어 연구에 힘을 기울이고 있다.

### 3. VoiceXML 구조

VoiceXML은 작년 8월에 ver. 0.0를 발표후 올해 3월 ver. 1.0으로 보완하여 발표후 또다시 올해 10월에 ver. 2.0이 발표된 상황이다. 이렇게 빠른 속도로 표준화가 진행중인 언어는 없을 것이다. 이것은 그만큼 VoiceXML의 관심을 보여주는 예라 할 수 있겠다. 이런 VoiceXML은 오디오, 음성인식, DTMF, 음성입력, 전화 등 음성 기반의 사용자 인터페이스를 통해 정보에 접근하는 방법을 제공하고 있다. VoiceXML을 이용하여 기존의 음성서비스 업체들은 적은 비용으로 새로운 음성서비스를 제공할 수 있으며, 기존 ISP업체에선 VoiceXML GateWay만을 설치함으로써 Voice ISP 서비스를 할 수 도 있다. VoiceXML을 이용하여 서비스를 제공할 경우, 음성 서비스 제작자는 음성 입출력에 관련한 기술적인 사항을 알 필요 없이 VoiceXML 문서에 의해 음성 시나리오만에 집중할 수 있도록 해준다.

VoiceXML은 그림 3-1 처럼 구성되어 작동한다. 문서 서버로부터 VoiceXML문서를 읽어 이를 VoiceXML 인터프리터에 의해 해석이 된후 이를 응용 프로그램에서 이용하는 구조를 가지고 있다 [2].

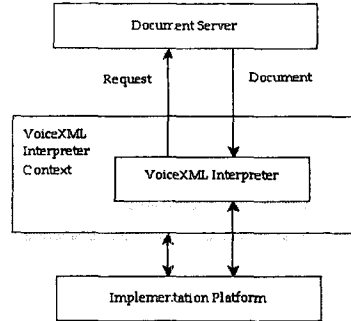


그림 3-1 VoiceXML 구현 모델

### 4. VoiceXML을 이용한 음성인터페이스 활용방안

기존의 응용프로그램에서 음성인터페이스를 사용할 때에는 각각의 메뉴 또는 기능들과 음성과 1:1 또는 제약적인 N:1로 대응하여 명령을 수행했다. 이런 방법들은 응용프로그램을 사용하는 사람 또는 개발자들에게 제한적으로 음성인터페이스를 구현해야 하는 불편함이 있었다. 이는 비단 일반 응용프로그램뿐만 아니라 인터넷 브라우저에서 또한 같은 방법으로 명령이 수행되어졌다. 게다가 현재 나와있는 음성웹브라우저는 실제 웹 문서간의 링크를 통한 이동시 HTML코드의 앵커태그(<A>)를 인식하여 이를 음성으로 발음하면 문서의 전이가 이루어지는 형식을 취하고 있다. 이는 GUI환경에서의 작업을 음성으로 단순 변환한 것이어서 실제 사용시 어색하게 작업을 수행한다. 이는 음성과 시각의 차이점을 간과한 이유와 적절한 방법의 페이지 전이를 할 수 있는 방법이 없었기 때문일 것이다. 그래서 본 논문은 이러한 시각과 음성 인터페이스간의 차이를 인지하고 실제 웹페이지 작성시 웹페이지 개발자가 자신의 웹 특성에 맞는 다양한 방법의 인터페이스 기반을 VoiceXML을 통해 마련하고자 한다.

VoiceXML은 XML의 서브셋이므로 현재의 XML파서와 음성인식/합성 모듈을 이용하여 구현할 수 있을 것이다. 본 논문에서 제시하는 VoiceXML을 이용한 음성 인터페이스는 다음과 같은 구조를 갖고 동작을 한다.

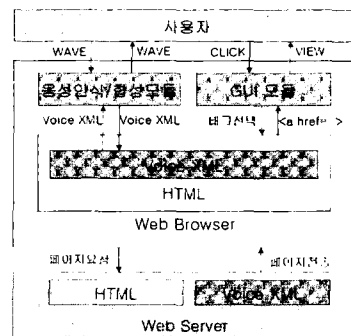


그림 4-1 웹 환경에서의 VoiceXML활용 방안 구성도

그림 4-1 에서 나타낸 바와 같이 음성 입출력의 방법에 대한 기술은 VoiceXML이 담당하며 일반적인 브라우저에서 사용되어지는 앵커태그는 기존의 방법과 동일하게 마우스를 통한 전이가 이루어지며, 필요에 의해서 VoiceXML에 기술된 앵커태그에 한해서 음성으로 전이를 수행한다. 이렇게 함으로써 현재의 환경에 최대한 수정 없이 사용되어질 수 있어 사용자의 접근을 용이하게 하며 또한 개발자들에게도 간단한 VoiceXML을 익힘으로써 다양한 방법으로 웹페이지를 개발할 수 있는 토대를 마련할 수 있으리라 본다. 본 논문에서 제시하는 VoiceXML을 이용한 음성인식 인터페이스의 작동 방법을 설명하면 다음과 같다.

HTML페이지에 삽입 되어있거나 또는 링크를 통해서 VoiceXML과 HTML을 읽어들이고 후 HTML에서 VoiceXML을 분리하여 이를 VoiceXML 파서와 TTS 등으로 사용자에게 음성을 통한 질의 또는 사용자의 음성 명령을 음성인식모듈에 의해서 VoiceXML의 전이를 발생시켜 사용자의 작업을 돕는다. 이러한 음성인터페이스를 위한 음성인식방법으로는 본 논문에서는 현재의 음성인식 기술을 감안하여 화자독립 고립단어를 선택하여 이를 VoiceXML에서 사용하고자 한다.

현재 구현되어 있는 음성인식브라우저는 크게 두가지 기능이 있다. 앞서 언급했듯이, 브라우저자체의 메뉴기능을 인식하여 사용자에게 편의를 제공하며, 또다른 것은 HTML페이지의 앵커태그를 인식하여 이를 유일한 단어로 만든 다음 이를 음성으로 인식하여 페이지의 전환이 이루어지고 있다. 하지만 이런 방법들의 불편한점은 웹페이지를 제작할때의 의도는 마우스 또는 키보드를 이용하여 서핑이 이루어진다는 가정에서 만들어지기 때문에 실제 음성인식브라우저에서 앵커태그를 인식하여 서핑할 때 어색한 느낌이 들며 이는 마우스를 이용할때 보다 오히려 불편한 느낌이 들고있는 실정이어서 사용자들의 의견을 받고있다. 이러한 문제점들을 해결하기 위해 본 논문에서는 몇가지 시나리오를 바탕으로 음성인터페이스를 설계하려 한다.

첫째, 사용자가 웹페이지에서 메뉴를 살피고 선택하는 시나리오.

단순한 열거식 메뉴구성과 제한적인 표현방법으로 기인한다. 그러므로 본 논문에서 제시하는 VoiceXML을 이용하여 메뉴구성을 할 경우 다음과 같은 이점을 얻을 수 있다. 음성은 시각보다 자연스럽게 사용자의 동작을 이끌어 갈 수 있다.

둘째, 게시판을 통한 게시물 읽는 시나리오

게시판의 특성은 실시간으로 갱신된다는 것이다. 이는 VoiceXML 문서 또한 실시간으로 제작되어야 함을 의미한다. 이는 CGI, ASP, PHP 등의 스크립트 등을 통해 작성하여야 함을 말한다. 이때 주의 할 것은 게시물의 선택을 위한 발음의 양을 줄이는 것과, 유일한 문장을 만들어 사용자에게 제시를 해야 한다. 이때 가장 유용한 방법은 게시물 번호를 발음하거나, 해당 게시물의 행을 발음하는 방법을 통해 해결 할 수 있다.

셋째, 검색엔진을 통한 정보검색을 하는 시나리오

현재는 결과를 사용자에게 보여줄 때 텍스트 기반으로만 보여주기 때문에 사용자는 검색된 문서들을 주의 깊게 보아야 하며, 다른 작업과 병행하기 힘든 상황이다. 이때 VoiceXML문서를 함께 사용자에게 전송함으로써 사용자에게 검색결과를 음성을 통해 알려주기 때문에 사용자는 좀더 편리한 환경에서 정보를 획득 할 수 있을 것이다.

앞서 설명한 인터페이스 설계방법을 고려하여 그림과 같이 VoiceXML문서를 HTML에 삽입하였다. VoiceXML은 XML서브셋이므로 기존의 XML기술을 그대로 사용할 수 있다 [4]. 본 문서는 XML 아일랜드 방법을 통해 HTML상에서 VoiceXML문서를 포함하게 하였다. 이는 다른 응용기술로도 이용할 수 있을 것이다.

```
<HTML>
<TITLE>.....</TITLE>
<HEAD>.....</HEAD>
<BODY>
<XML ID="voice">
<?xml version="1.0" encoding="cuc kr"?>
<DIALOG>
<STEP NAME="init">
<PROMPT>안녕하세요 기상청입니다. 날씨를 알고자 하신곳의 지명을
만남해 주십시오.</PROMPT>
<HELP>지명은 다음과 같습니다. 동두천, 백령도, 서울, 인천, 수원, 강화
중 한곳을 말씀해 주십시오.</HELP>
<INPUT TYPE="OPTIONLIST" NAME="location">
<OPTION VALUE="동두천" NEXT="#동두천단어">동두천
</OPTION>
<OPTION VALUE="백령도" NEXT="#백령도단어">백령도
</OPTION>
<OPTION VALUE="서울" NEXT="#서울단어">서울</INPUT>
</STEP>
<STEP NAME="동두천단어">
<PROMPT>강일의 동두천 날씨는 현재기온 18.5도, 풍향 S, 풍속 2.0
입니다.</PROMPT>
</STEP>
<STEP NAME="백령도단어">
<PROMPT>강일의 백령도 날씨는 현재기온 23.74도, 풍향 S, 풍속 2.0
입니다.</PROMPT>
</STEP>
</DIALOG>
</xml>
</XML>
```

그림 5-2. VoiceXML을 HTML에 삽입하여 사용하는 모습

참고문헌

- [1] World Wide Web Consortium, <http://www.w3c.org>
- [2] 홍용택, "XML과 VoiceXML의 효율적인 통합방안 제시", 군산대학교 교육대학원 석사논문, 2001.10
- [2] VoiceXML Forum, <http://www.voicexml.org>
- [3] 마이크로소프트 스피치, <http://www.microsoft.com/speech/>
- [4] Professional XML, WROX