

# VoiceXML을 이용한 VUI 지원 웹브라우저 개발

## (Development of a Voice User Interface for Web Browser using VoiceXML)

예 상 후<sup>†</sup>      장 민 석<sup>\*\*</sup>  
(SangHoo Yea)    (MinSeok Jang)

**요 약** 현재의 웹정보들은 주로 HTML로 기술되어 있으며, 이러한 정보를 얻기 위해 사용자들은 마우스와 키보드와 같은 입력장치를 사용한다. 이와 같이 기존의 GUI 환경은 인간의 가장 자연스러운 정보 획득 수단의 하나인 음성을 지원하지 못하고 있다. 이러한 문제를 해결하기 위해 음성 인터페이스를 가진 여러 제품들이 개발되고 있다. 하지만 이들은 상호대화성이나 기존 웹환경을 수용한다는 측면에서 부족한 면을 가지고 있다. 본 논문에서는 현재 무르익어 가는 음성인식 기술과 XML의 파생언어인 VoiceXML을 이용하여, 기존의 인터페이스 환경을 XML 기반의 대화형 음성인터페이스 환경으로 대체하고자 한다. 이를 통해 기존의 인터페이스 환경을 수용한 VUI(Voice User Interface) 환경을 사용자에게 제공할 수 있다. 기존의 환경을 수용하기 위해 “XML Island” 기술을 이용하여 VoiceXML 문서를 HTML 문서에 포함시키며, 대표적인 정보획득화면인 메뉴, 게시판, 검색 엔진에 대한 대화형 음성 시나리오를 제안하고 있다.

**키워드** : 확장성표기언어, 음성확장성표기언어, 음성인식, 음성합성, 음성유저인터페이스, 다이얼로그

**Abstract** The present web informations are mainly described in terms of HTML, which users obtain through input devices such as mouse, keyboard, etc. Thus the existing GUI environment have not supported human's most natural information acquisition means, that is, voice. To solve the problem, several vendors are developing voice user interface. However these products are deficient in man-machine interactivity and their accommodation of existing web environment. This paper presents a VUI(Voice User Interface) supporting web browser by utilizing more and more maturing speech recognition technology and VoiceXML, a markup language derived from XML. It provides users with both interfaces, VUI as well as GUI. In addition, XML Island technology is applied to the browser in a way that VoiceXML fragments are nested in HTML documents to accommodate the existing web environment. Also for better interactivity, dialogue scenarios for menu, bulletin, and search engine are suggested.

**Key words** : XML, VoiceXML, Speech Recognition, Speech Synthesis, VUI(Voice User Interface), Dialog

### 1. 서 론

1991년 CERN의 WWW 서비스가 개발된 이후 인터넷은 폭발적인 성장을 이루어왔다. 하지만 현재 인터넷 정보들은 주로 HTML로 기술되어 있으며 이는 응용 및

통합에 대한 어려움으로 인해 그 한계를 드러내고 있다. 데이터와 스타일의 통합적인 구성으로 인해 효율적인 정보처리를 힘들게 하고 있다. 이러한 단점을 보완하고자 문서의 구조적 정보 및 내용을 정의하는 XML(eXtensible Markup Language)이 등장하게 되었다. 이는 SGML의 대표적인 단점들을 대부분 보완하였고, 특별한 단점이 발견되지 않고 있어 전자문서를 사용하는 대부분의 응용분야에서 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

· 본 연구는 산학협동재단의 2004년도 학술연구비 및 (주)티엔엔솔루션의 matching fund 지원에 의해 수행되었음

† 비 회 원 : (주)대신아이텍 사원  
sanghoo\_yea@hotmail.com

\*\* 정 회 원 : 군산대학교 컴퓨터정보학과 교수  
(corresponding author)  
msjang@kunsan.ac.kr

논문접수 : 2002년 10월 31일  
심사완료 : 2004년 12월 29일

생성되어지고 있으며, VoiceXML 또한 그러한 맥락에서 생성된 언어이다[1].

현재 VoiceXML은 VoiceXML 포럼에서 AT&T, IBM, 루슨트 테크놀로지, 모토로라 등 정보통신 분야의 4개 거대 기업체를 중심으로 표준화가 진행중이며 2002년 4월에 ver 2.0이 표준으로 인정되었다[2]. 이와 관련하여 현재 뚜렷한 응용 프로그램들이 나와 있는 상태는 아니나 현재 각 음성정보 통신업체를 비롯한 IT업체에서 VoiceXML을 대비하기 위한 각종 솔루션들을 개발하고 있으며, 이는 곧 다가올 VoiceXML의 파장효과가 클 것이라는 전망을 낳기에 충분하다. 하지만 지금까지의 VoiceXML은 전화기를 통해서 음성정보 서비스를 제공하기 위한 방편으로 주로 개발이 되어졌다. 하지만 본 논문에서는 이를 이용하여 현재의 웹 GUI 위주의 인터페이스에서 VUI(Voice User Interface)를 제공할 수 있도록 하고 있다. 특히 웹브라우저의 경우 국내의 음성인식관련 업체에서 개발하여 배포하고 있으나, 이는 브라우저의 메뉴를 음성으로 동작하게 하는 것과 HTML의 앵커태그를 음성으로 실행하도록 하는 데 그치고 있다. 따라서 실제 음성브라우저를 사용할 경우 단순히 사용자는 기존의 GUI를 음성으로 대체한 환경에서 사용을 해야 했기 때문에 어색한 동작을 보여주고 있다. 본 논문에서는 이러한 점을 해결하기 위해 VoiceXML을 이용하여 음성인터페이스의 시나리오를 제공하고, 이를 이용할 수 있는 VoiceXML VUI 브라우저를 설계/구현하고 있다.

## 2. 국내·외 VoiceXML 기술개발 현황

VoiceXML 연구가 진행되는 동안에도 관련 업체에서는 이를 사용하기 위한 각종 응용 솔루션들을 개발하고 있는 상태이다. 특히 IBM사는 자사의 WebSphere서버[3]의 Voice Toolkit[4]을 이용해 VoiceXML을 서비스할 수 있도록 개발했으며, MS사는 VoiceXML이라 명명하지는 않지만 VoiceXML의 성격을 반영하는 XML을 이용하여 자사의 상품 등에 적용하고자 노력하고 있다. 이외에도 뉴앙스, 스피치웍스 등 음성 서비스 업체에서 또한 자사의 플랫폼 및 서버 등을 개발/업그레이드하여 VoiceXML을 수용하고자 하고 있다.

VoiceXML은 XML의 서브셋이므로 파서 등은 기존의 XML 파서를 이용해서 개발함으로써 개발비용을 절감할 수 있다. 현재 VoiceXML을 지원하는 파서들은 대부분 외국 제품이며 국내의 파서들은 아직 공개가 되어있지 않은 상태이다. VoiceXML 규격은 버전 2.0을 발표한 이후 계속 표준화가 진행중이며, 국내 기업체 및 대학에서 이와 관련된 각종 기반 기술을 개발하고 있다. 하지만 이러한 연구들은 아직까지 인터프리터 또는 파

서 개발에 그치고 있어 좀더 응용할 수 있는 기반기술을 개발해야 하는 실정이다. 또한 W3C에서도 본 연구와 같이 VoiceXML을 웹 환경에서의 인터페이스로 사용하기 위해 2001년 12월부터 논의 중이다[5].

국외에서는 VoiceXML 기술관련 연구로 이를 좀더 효과적으로 이용하기 위한 "Dialog & Grammar"에 관한 연구와, 각종 플랫폼에서의 효과적인 VoiceXML 구축방법 등이 주축을 이루고 있다[6]. 이를 요약하면 국내에서는 아직까지 기반기술에 중점을 맞추고, 국외에서는 이미 실용화 단계의 작업을 하고 있는 상황이다. 또한 국내에서 파서 개발 및 단순 응용을 주로 목표로 하는 반면에 국외에서는 본 연구의 목표와 같이 VoiceXML을 웹 환경에서 인터페이스로 사용하기 위해 W3C에서 2001년 12월부터 "XHTML+Voice"[7]라는 이름으로 표준화에 대해 논의 중이다[2]. 이와 본 논문의 차이점은 현재의 웹 환경에 XML Island 기법을 이용함으로써 현재 환경을 변경하지 않고 VoiceXML을 지원할 수 있다. 반면에 XHTML+Voice는 XHTML에 VoiceXML을 병행하도록 하는 것이다. 그림 1은 XHTML+Voice 문서의 예를 나타내고 있다.

이밖에도 W3C의 멀티모달(multimodal) 인터페이스와, MS사 주도의 SALT[8,9] 등 좀더 편리한 인터페이스를 제공하기 위한 연구가 이루어지고 있다. 이중 멀티모달 인터페이스는 무선 전화기의 사용을 편리하게 하기 위해 WAP(Wireless Application Protocol)등에 음성인터페이스를 제공하여, 사용자의 편의를 돕는 방법이며, SALT(Speech Application Language Tags) 관련 연구에는 2001년 설립된 SALT 포럼에서 MS를 중심으로 전 세계 70여 개 음성전문 기업이 참여하고 있으며, 본 논문에서 지향하는 궁극적인 목표와 일치한다고 볼 수 있다. 하지만 SALT의 경우 XML 기반의 그들 자체의 표준화 및 관련 기술들을 논의하고 있으며, 본 논문에서는 음성 인터페이스 제공 방법에 있어 VoiceXML 포럼에서 제시한 VoiceXML Ver 1.0의 서브셋을 이용한다는 점이 다르다.

## 3. VoiceXML 구조

VoiceXML은 1999년 8월 ver 0.9를 발표 후 2000년 3월 ver 1.0으로 보완하여 발표하였으며, 2001년 10월에 ver 2.0이 발표된 이후 계속해서 보완해 나아가고 있다. 이는 오디오, 음성인식, DTMF, 음성입력, 전화 등 음성기반의 사용자 인터페이스를 통해 정보에 접근하는 방법을 제공하고 있으며, 이를 이용하여 기존의 음성서비스 업체들은 적은 비용으로 새로운 음성서비스를 제공할 수 있으며, 기존 ISP 업체에서는 VoiceXML 게이트웨이만을 설치함으로써 Voice ISP 서비스를 제공할 수

```

<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+Voice //EN"
  "xhtml+voice1.0.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:vxml="http://www.w3.org/2001/voicexml20"
  xmlns:ev="http://www.w3.org/2001/xml-events">
  <head>
    <title>What You See Is What You Can Say</title>
  <!-- first declare the voice handlers. -->
  <vxml:form id="voice_city">
    <vxml:field name="field_city">
      <vxml:grammar src="city.srgf" type="application/x-srgf"/>
      <vxml:prompt id="city_prompt">
        Please choose a city.
      </vxml:prompt>
      <vxml:catch event="help nomatch noinput">
        For example, say Chicago.
      </vxml:catch>
    </vxml:field>
  </vxml:form>
  <vxml:form id="voice_hotel">
    <vxml:field name="field_hotel">
      <vxml:grammar src="hotel.srgf"
        type="application/x-srgf"/>
      <vxml:prompt id="hotel_prompt">
        Select your hotel
      </vxml:prompt>
      <vxml:catch event="help nomatch noinput">
        For example, say Hilton.
      </vxml:catch>
      <vxml:filled>
        <vxml:prompt>
          You have chosen to stay at the
          <vxml:value expr="field_hotel"/>.
        </vxml:prompt>
      </vxml:filled>
    </vxml:field>
  </vxml:form>
  <!-- done voice handlers. -->
</head>
<body>
  <h1>What You See Is What You Can Say</h1>
  .....

```

그림 1 XHTML+Voice의 모습

도 있다. VoiceXML을 이용하여 서비스를 제공할 경우, 음성 서비스 제작자는 음성 입력력에 관련한 기술적인 사항을 알 필요 없이 VoiceXML 문서를 통해 음성 시나리오에만 집중할 수 있도록 한다[10-12]. 이의 구현 모델은 VoiceXML 문서를 제공하는 각종 응용프로그램과 데이터베이스, VoiceXML 게이트웨이, 음성 인식기/합성기 등으로 구성되어 음성서비스를 제공하게 된다[13]. 전화기를 이용해 VoiceXML 게이트웨이를 통해서

기존의 응용서비스와 연결되어 각종 음성 서비스를 제공받게 된다. 이때 음성인식 기술과 TTS(Text To Speech)기술을 이용하여 사용자에게 정보를 제공하며, 오디오 파일을 기본적으로 제공한다[14].

기존의 ARS 등에 비해 VoiceXML을 이용하면 훨씬 더 풍부한 서비스가 손쉽게 구축된다. 기존의 경우 새로운 아이디어를 음성 서비스와 인터넷 홈페이지에 적용하려는 경우 시스템을 수정해야 하지만, VoiceXML을 활용할 경우 서비스 내용을 문서 형태로 작성하기 때문에 서비스 구축/변경이 훨씬 더 용이하다. 이러한 장점들은 실시간으로 음성서비스를 제공할 수 있게 하는 토대가 된다.

#### 4. 웹 환경에서 VoiceXML을 이용한 음성인터페이스 활용방안

본 절에서는 음성인식을 이용한 각종 서비스/프로그램과 음성을 지원하는 웹브라우저에 대해서 알아보고, 그 문제점을 해결하기 위한 방안을 제시한다.

현재 음성 응용프로그램은 다음과 같이 크게 네 가지로 나눌 수 있다.

1. ARS 등의 전화서비스
2. 음성엔진을 이용한 PC 응용프로그램
3. VoiceXML을 이용한 전화서비스
4. VoiceXML을 이용한 PC 응용프로그램

ARS 전화 서비스는 현재 각종 설문조사나 이벤트, 벨소리 등을 각 통신업체에서 경쟁적으로 구축하고 있는 상황이며, 음성엔진을 이용한 PC 응용프로그램은 현재 음성인식/합성관련 업체에서 자사의 상품을 홍보하는 수준의 프로그램이 제작되어 있는 정도이다. 또한 현재 국내의 VoiceXML 관련 기업에서는 기존의 ARS 시스템의 유지/보수비용의 절감을 위한 방편으로 VoiceXML을 이용한 ARS 시스템을 개발, 판매하고 있다. 마지막으로 VoiceXML을 이용한 PC 응용프로그램은 PC 환경에서의 파서 및 개발도구 정도만 개발되어져 있다[15,16]. 본 논문의 범위는 4에 해당되며, 이는 VoiceXML을 VUI로 사용하는 웹브라우저의 설계/구현을 의미하며, 이때 VUI를 구성하는 VoiceXML 다이얼로그 디자인까지 포함한다.

기존의 음성인식 브라우저는 크게 세 가지로 나뉜다. 첫 번째는 브라우저의 메뉴 기능을 음성인식 기능을 통해서 수행하도록 하는 것이며, 두 번째는 웹페이지의 앵커 태그를 인식하여 이를 발음하거나, 앵커태그의 순서에 따라 번호를 부여한 후 이를 발음할 때 문서의 전이가 이루어지게 하는 방법을 사용하는 것이다. 마지막으로 현재 개발되어 있는 대부분의 음성인식 브라우저는

위의 두 가지 기능을 모두 가지고 있다.

하지만 이들은 사용자와 웹사이트간의 상호대화성(interactive) 인터페이스를 제공하지 못하고, 단순히 링크를 음성으로 클릭함으로써 본연의 VUI 특징을 살리지 못하고 있다. 이는 웹사이트에 VUI를 제공할 수 있는 표준화된 도구가 마련되어 있지 않기 때문이다. 따라서 본 논문에서는 이러한 시각과 음성 인터페이스간의 차이를 인지하고 실제 웹페이지를 작성할 때 개발자가 자신의 웹 특성에 맞는 다양한 방법의 인터페이스 기반을 VoiceXML을 통해 마련하고자 한다.

VoiceXML은 XML의 서브셋이므로 현재의 XML 파서와 음성인식/합성 모듈을 이용하여 구현하고 있다. 본 논문에서 제시하는 VoiceXML을 이용한 음성 인터페이스는 그림 2와 같은 구조로 동작한다.

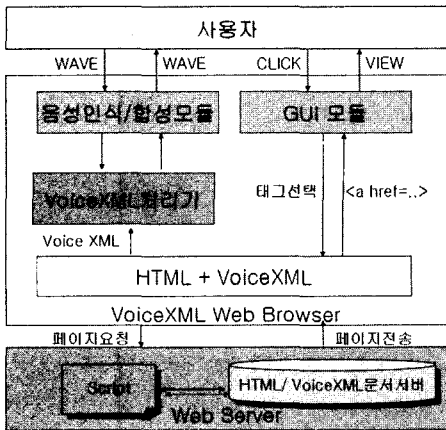


그림 2 웹 환경에서의 VoiceXML활용 방안 구성도

VoiceXML VUI 브라우저는 웹 서버로부터 스크립트와 DB에 의해 구성된 HTML 문서를 전달받아 HTML 내부에 XML Island로 첨가된 VoiceXML 문서를 추출하여 이를 음성인식기/합성기를 통해 사용자와 적절히 반응하여 VoiceXML 문서처리를 수행한다. 또한 일반적인 HTML 브라우저에서 사용되는 앵커태그는 기존의 방법과 동일하게 마우스를 통해 브라우저를 수행하며, 필요에 의해서 VoiceXML에 기술된 앵커태그에 한해서 음성으로 전이를 수행한다. 이렇게 함으로써 현재의 환경에 최소한의 수정을 가해 사용자의 접근을 용이하게 함과 동시에 VoiceXML 구문을 익힌 초보개발자로 하여금 다양한 방법으로 웹페이지를 개발할 수 있도록 한다. 이런 VoiceXML 문서는 다음의 그림 3에서와 같은 모습으로 HTML 문서상에 포함되어 VoiceXML VUI 브라우저에서 처리가 이루어진다.

XML Island는 HTML 내에 의미를 갖는 데이터를

```

<HTML>
<TITLE>VoiceXML 을 이용한 웹 VUI환경제공</TITLE>
<HEAD>음성을 이용할 수 있습니다.</HEAD>
<BODY>
  <XML ID="vxml">
    <vxml version="1.0">
      <form>
        .....
      </form>
    </vxml>
  </XML>
</BODY>
</HTML>

```

그림 3 VoiceXML을 HTML에 삽입하여 사용하는 모습

삽입하여 이를 HTML 상의 각종 스크립트로 제어하여 좀 더 풍부한 응용을 할 수 있도록 도와주는 기술이다. 이것은 실제 HTML을 브라우저할 경우에는 보여 지지 않는 내용으로, VoiceXML을 지원하는 웹페이지는 기존의 사용자들을 포함할 수 있게 된다.

본 논문에서는 이를 이용하여 웹에서 가장 많이 사용되는 세 가지 시나리오를 바탕으로 원활한 음성 인터페이스를 만들기 위한 VoiceXML 다이얼로그를 제시한다. 우선 현재의 웹사이트의 구성을 분류하면 다음과 같다.

첫째, 사용자가 웹페이지에서 메뉴를 살피고 선택하는 시나리오이다. 현재의 포털사이트 같은 경우 많은 메뉴와 현란한 그래픽으로 자주 방문하는 사용자가 아니라면 그 사이트의 특징을 파악하기 어려우며 특정 메뉴를 찾아가기에도 상당히 어렵다. 이는 단순한 열거식 메뉴 구성과 제한적인 표현방법에서 비롯된다. 하지만 본 논문에서 제시하는 VoiceXML을 이용한 메뉴구성 방식을 적용할 경우 다음 같은 이점을 있다. 음성은 GUI 환경에서의 마우스클릭 또는 키보드 입력과 병행하여 “예(yes)” 또는 “아니오(no)” 등의 음성입력 방법으로 자연스럽게 사용자를 웹페이지 제작자의 의도대로 동작을 이끌어 갈 수 있다. 이는 현재 포털업체 또는 쇼핑몰 등에서 사용자의 구매를 촉진하는 많은 방법들을 개발하고 있는 시점에서 하나의 방법이 될 수 있다. 또한 사용자에게 음성을 통한 광고 등을 보낼 수 있으므로 인터넷 광고시장까지 그 영역을 넓혀갈 수 있을 것이다.

둘째, 게시판을 통한 게시물을 읽는 시나리오이다. 웹 환경이 온라인 커뮤니티의 주된 활동무대가 된 이유 중의 하나가 게시판의 활성화이다. 기존의 텍스트 기반의 게시판보다 좀더 사용자에게 친밀감을 줄 수 있으며 이용법 또한 기존의 방법을 병행하여 좀더 편리하게 이용할 수 있다. 게시판의 특성은 실시간으로 갱신된다는 것이다. 이는 VoiceXML 문서 또한 실시간으로 제작되어야 함을 의미한다. 이는 CGI, ASP, PHP 등의 스크립

트 등을 통해 작성하여야 함을 말한다. 이때 주의할 것은 게시물의 선택을 위해 사용자 발음의 양을 줄이는 것과, 유일한 문장을 만들어 사용자에게 제시해야 하는 것이다. 가장 유용한 방법은 게시물 번호를 발음하거나, 해당 게시물의 행을 발음하는 방법을 통해 해결할 수 있다.

셋째, 검색엔진을 통한 정보검색 시스템 이용 시나리오이다. 검색엔진은 정보수집의 중요한 부분을 차지하고 있다. 현재 각 포털업체에서 제공하는 검색엔진이 경우, 검색결과를 텍스트로 보여주기 때문에 사용자는 검색된 문서들을 주의 깊게 보아야 하며 다른 작업과 병행하기 힘들다. 이때 VoiceXML 문서를 함께 사용자에게 전송함으로써 사용자에게 검색결과를 음성으로 알려준다면 사용자는 좀더 편리한 환경에서 정보를 획득할 수 있을 것이다.

### 5. VoiceXML VUI 브라우저 설계/구현

#### 5.1 VoiceXML VUI 브라우저 설계

현재까지 VoiceXML을 웹 환경에서 VUI로 활용할 수 있음을 알아보았다. 본 장에서는 실제 브라우저를 개발하여 그 효과를 실험하고자 한다. 위에서 언급한 세 가지 시나리오를 충실하게 수행할 수 있는 VoiceXML VUI 브라우저를 설계/구현한다. 현재까지의 VoiceXML 파서들은 개발업체의 IDE 환경에 통합되어 있어 이를 이용하여 응용프로그램을 만들 수 없는 상황이다. 따라서 여기서는 MSXML 3.0을 VoiceXML 파서로 이용하며, VoiceXML 처리기에서 VoiceXML 시나리오에 맞추어 사용자에게 VUI 환경을 제공한다. 이때 음성인식/합성기는 MS사의 Agent를 이용하여 VoiceXML 문서상의 각 엘리먼트 기능에 맞추어서 인식/합성기능을 수행한다. 이러한 처리기의 구조는 그림 4와 같다.

이의 주요 동작은 처리기 내에서 MSXML은 Voice-

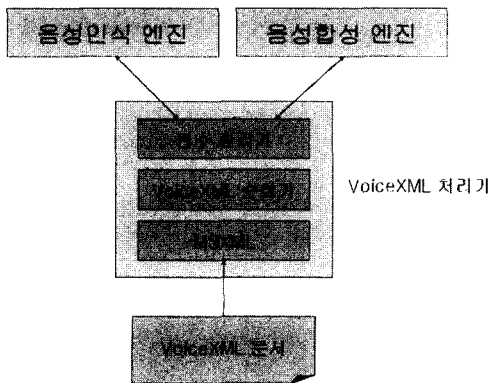


그림 4 VoiceXML 처리기 구조

XML 1.0 DTD를 기준으로 VoiceXML 검증을 하고, VoiceXML 순회기가 DOM을 이용하여 순회를 하며 VoiceXML 기능을 수행한다. 이때 변수 관리기는 VoiceXML에서 사용되어지는 변수들과 ECMAScript에 의해 처리되는 각종 추가적인 명령들의 정보를 관리하여 VoiceXML 순회기와 상호보완적으로 작업을 수행한다.

#### 5.1.1 VoiceXML VUI 브라우저에서 인식하는 엘리먼트

현재 공개되어 있는 VoiceXML은 전화기를 통신수단으로 이용하여 음성서비스를 제공하기 위해 규정되어 있기 때문에 본 논문에서와 같이 웹 환경을 대상으로 할 경우에는 필요 없는 엘리먼트가 상당수 존재한다. 따라서 우선 필요한 엘리먼트만을 추출하고자 한다. 표 1과 같이 엘리먼트를 추출하여 VUI 시나리오에서 사용하고 있다. 전화기의 특성을 고려한 엘리먼트 및 실제 시나리오 작성시 불필요한 엘리먼트들을 제거하였다. 이는 웹 환경에서 VUI로 사용할 경우 기존의 기술과 함께 쓰여야 하기 때문이고, 개발자들의 부담을 최소화하기 위해서다. 이렇게 함으로써 같은 기능을 수행하는 시나리오가 거의 같은 모습의 VoiceXML 코드로 작성되게 할 수 있었다. 이는 코드 작성 때 정형화된 코드를 제공하여 사용자들이 쉽게 이용할 수 있도록 한다.

표 1 추출한 엘리먼트

| 형태별 분류 | 엘리먼트                              |
|--------|-----------------------------------|
| 다이얼로그  | form, field, option, menu, choice |
| 분기     | goto                              |
| 조건문    | if, else, elseif, option          |
| 음성관련   | block, prompt                     |
| 연결     | link                              |
| 변수관련   | var                               |

#### 5.1.2 VoiceXML 처리기 작동과정

VoiceXML 처리기는 그림 5에서와 같이 로드된 문서

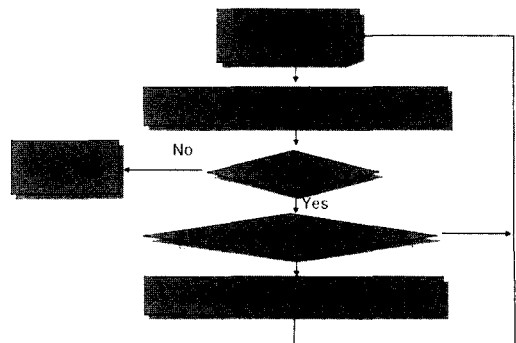


그림 5 VoiceXML 처리기 동작과정

를 파싱후 적절한 방법으로 사용자와 정보를 주고받으며 서비스를 제공한다.

이 과정은 처음으로 VoiceXML 문서를 파싱하여 문서의 무결성을 확보한 후 초기화 과정으로 로드된 문서의 변수 등의 정보를 초기화/저장한다. 다음으로 VoiceXML 문서가 사용자와 대화가 필요한지의 여부에 따라서 폼해석 알고리즘(Form Interpretation Algorithm)을 수행하거나 또는 다른 문서로의 전이를 수행한다. 폼해석 알고리즘은 그림 6에서처럼 수행한다[5,17].

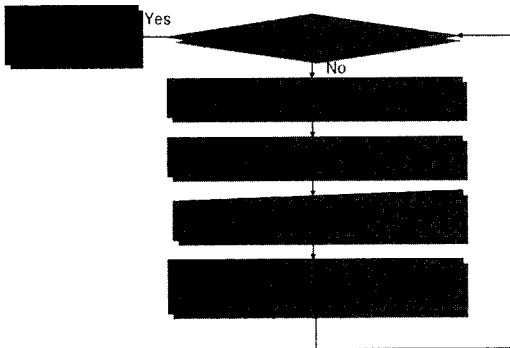


그림 6 폼해석 알고리즘

다중 폼을 갖고 있는 VoiceXML 문서는 모든 폼을 처리하거나 중간에 사용자의 입력에 따라 다른 폼 또는 새로운 VoiceXML 문서로의 전이를 하게 된다. 단일 폼 내에서는 <prompt> 엘리먼트에 의해서 사용자에게 질의 또는 정보를 알려준 다음 적절하게 다음 처리를 수행하다가 <filled> 엘리먼트에 의해서 사용자의 입력을 받아 유효성 검사 후 적절하게 다음 엘리먼트를 수행하게 된다.

5.2 구현

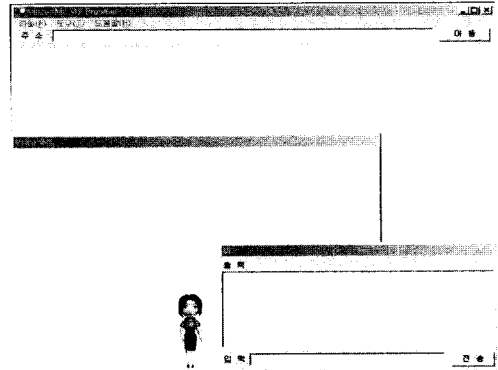
본 논문에서 구현한 VoiceXML VUI 브라우저는 VoiceXML 1.0을 충실히 반영하였으나, ECMAScript와 <grammar>와 관련된 부분은 VoiceXML과는 또 다른 표준안을 따라야 함으로 제외하였다. 표 2와 같은 환경에서 개발하여 실험하였다.

개발 언어로는 Java와 같이 플랫폼에 독립적인 개발 환경을 제공해 주는 닷넷 프레임워크상의 개발언어인

표 2 VoiceXML VUI 브라우저 개발 환경

| 구분         | 내용                                     |
|------------|--|
| 운영 체제      | Windows 2000 Server Service Pack 5.0   |
| 개발 언어      | C# (Dot NET Framework)                 |
| XML 파서     | MSXML 3.0                              |
| 음성인식/합성 엔진 | MSAGENT 2.0 (한국어 TTS: L&H TTS 3000 엔진) |

C#을 이용하였다[13]. MSAGENT 2.0은 Speech 5.0에서 영어/중국어를 인식할 수 있는 인터페이스를 마련해 주며, 추가적으로 다른 업체의 모듈을 이용할 수 있도록 되어 있다[18]. 한글 합성엔진으로는 L&H사의 한글 TTS 3000을 이용하였다. 그림 7은 구현한 브라우저의 초기 화면이다.



(a) VoiceXML VUI 브라우저의 각종 창



(b) VoiceXML VUI 브라우저의 기능메뉴

그림 7 구현된 VoiceXML VUI Browser

그림 7(a)에서의 브라우저는 크게 세 개의 창으로 구분된다. 우선 HTML을 나타내는 창이 있고, HTML에서 추출한 VoiceXML을 나타내는 창이 있으며, 마지막으로 사용자와의 음성 인터페이스를 시각적으로 나타내는 부분이 있다. 음성 입력은 입력창과 병행하여 작동하며 출력은 캐릭터에 의해 음성출력과 동시에 텍스트 출력을 병행하도록 하고 있다. 그림 7(b)의 메뉴 구성은 온라인 상의 웹사이트뿐만 아니라 사용자가 직접 작성한 웹페이지를 확인해 볼 수 있도록 파일메뉴의 하위메뉴로 로컬파일을 읽어올 수 있는 기능과, VoiceXML 문서의 구조를 볼 수 있는 VoiceXML 창과, 사용자의 음성입력을 받고, 음성출력을 하는 VUI 창을 선택할 수 있는 메뉴를 포함하고 있다. 이 기능으로 사용자에게 따라서 자신이 원하는 인터페이스 환경을 조절할 수 있다.

5.3 실험

우선 메뉴 선택 시나리오에 대해서 브라우저를 다음과 같은 순서로 실행한다. 게시판/검색엔진 시나리오에 대해서도 동일하게 적용한다.

1단계 : HTML 상에 음성 인터페이스가 쓰여질 부분을 추출하여 음성시나리오를 제작한다.

2단계 : 구성된 음성시나리오를 VoiceXML로 코딩하여 HTML에 첨가한다.

3단계 : 일반 브라우저/VoiceXML VUI 브라우저에서 작동여부를 확인한다.

5.3.1 음성 시나리오 제작

본 절에서는 VoiceXML이 웹사이트에서 VUI로 사용될 수 있도록 현재의 환경을 분석하고 이를 토대로 VoiceXML 다이얼로그를 설계한다.

그림 8은 메뉴 선택에 의해 페이지의 전환이 이루어지는 일반적인 HTML 문서이다. 여기에서 메뉴부분을 음성 인터페이스로 변환하였으며, 추가적인 문서의 정보를 음성으로 출력하도록 하였다. 문서정보를 음성시나리오로 변환시 단순히 문서의 내용을 음성으로 전환할 경우 사용자에게 오히려 혼란스러운 인터페이스를 제공하게 되므로 간단명료하게 변환시킬 페이지의 링크정보와 링크연결을 위한 단어의 선별이 필요하다. 그림 8, 그림 9는 이를 나타낸 그림이다.

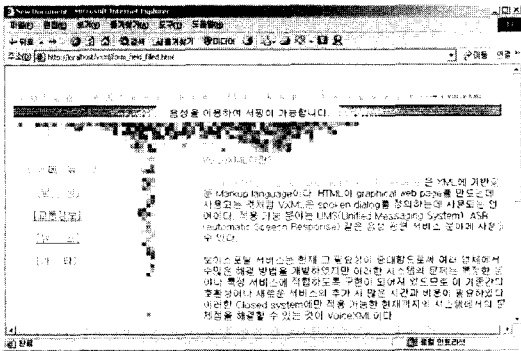


그림 8 VoiceXML을 추가할 HTML 문서

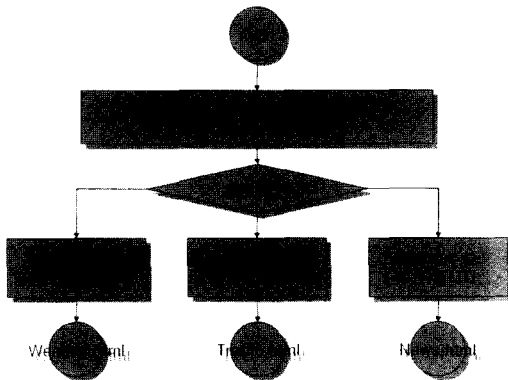


그림 9 음성 시나리오의 순서도

기존의 VoiceXML 서비스에서는 복잡한 구조로 표현을 해야 했으나, 본 논문에서 제시하는 VUI로 사용할

경우 대부분 그림 9에서와 같이 간단한 구조로 충분히 표현이 가능하다. 이는 시나리오가 HTML 페이지에 종속적으로 쓰이기 때문이다. 하지만 페이지의 정보를 음성정보로 전환하는 것에 목적을 둔다면 기존의 복잡한 VoiceXML 구조를 갖게 될 것이다.

5.3.2 VoiceXML 코딩/HTML에 추가하기

앞서 제작한 음성 시나리오를 바탕으로 VoiceXML을 코딩하면 여러 가지 방법이 있겠으나 본 논문에서는 그림 10과 같이 <filled> 엘리먼트를 이용하는 방법을 이용하였다.

```

<HTML>
<TITLE>.....</TITLE>
<HEAD>
<XML ID="vxml">
<vxml version = "1.0">
<form>
<field name="selection">
<prompt>날씨, 교통정보, 뉴스중 원하시는 서비스를
말씀해주세요.</prompt>
<filled>
<if cond = "selection == '날씨'">
<goto nextitem = "weather" />
<elseif cond = "selection == '교통'" />
<goto nextitem = "traffic" />
<elseif cond = "selection == '뉴스'" />
<goto nextitem = "news" />
.....
<link next = "weather.html"/>
</field>
<field name = "traffic">
<prompt>현재의 교통정보를 말씀드리겠습니다.
.</prompt>
<link next = "traffic.html"/>
</field>
<field name = "news">
<prompt>오늘의 주요 뉴스입니다.</prompt>
<link next = "news.html"/>
.....
    
```

그림 10 작성된 VoiceXML 코드

그림 10과 같이 <filled> 엘리먼트와 조건을 나타내는 엘리먼트인 <if> 엘리먼트를 이용하여 입력된 정보에 따라서 적절한 음성을 출력하며 또한 <link> 엘리먼트로 전이까지 이루어지도록 하고 있다. 그림 10은 사용자에게 다음과 같은 음성질의를 통해 문서의 전이와 사용자에게 적절한 정보를 제공한다.

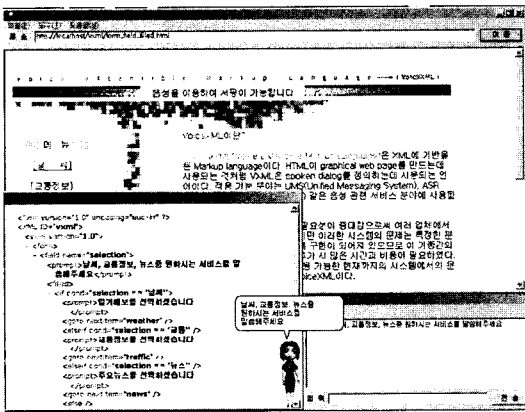
[컴퓨터] 날씨, 교통정보, 뉴스 중 원하시는 서비스를 말씀해주세요  
 [사용자] 날씨

[컴퓨터] 오늘의 일기예보를 말씀드리겠습니다(weather.html 로 이동).

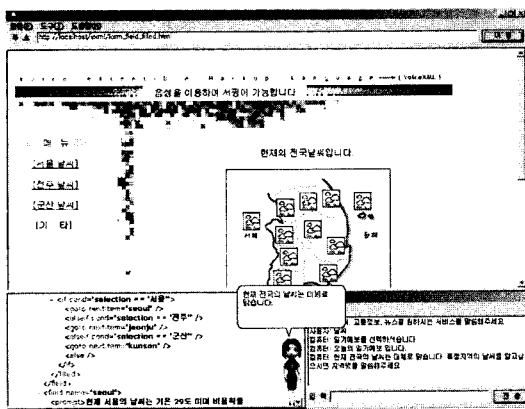
...

5.3.3 일반 브라우저/VoiceXML VUI 브라우저를 이용한 실험

본 절에서는 앞서 제작한 VoiceXML 문서를 포함한 HTML 문서를 이용하여 실제 VUI 환경을 마련할 수 있는지를 실험한다. 그림 8과 같이 브라우저되는 HTML 문서 내에 그림 10의 VoiceXML 문서를 포함하여 본 논문에서 제작한 VoiceXML VUI 브라우저를 이용하여 그림 11과 같은 결과를 얻었다.



(a) 초기 메인 화면



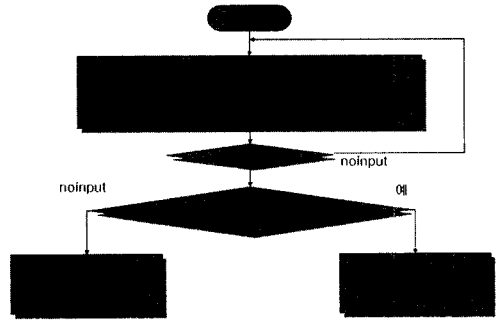
(b) 메인 메뉴에서 하위메뉴로 전이

그림 11 VoiceXML VUI 브라우저의 선택 메뉴 시나리오 결과

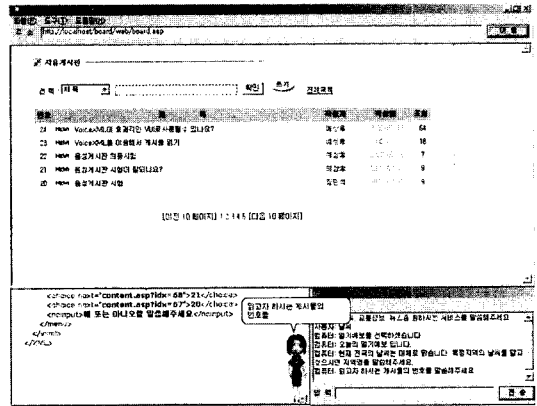
그림 11에 나타난 바와 같이 VoiceXML문서를 인식하여 앞서 작성된 음성시나리오와 동일하게 작동하고 있다.

5.3.4 게시판/검색엔진 실험

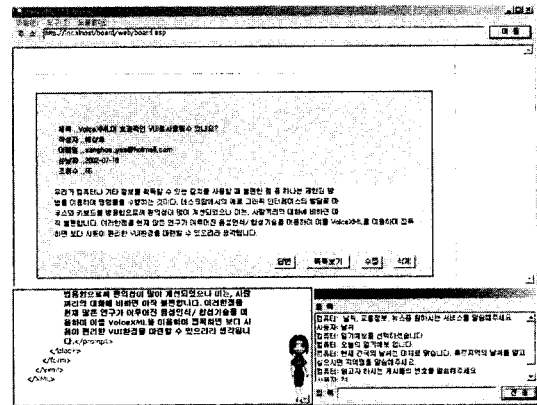
게시판의 경우 게시된 게시물에서 읽고자 하는 게시물의 번호와 게시물의 제목을 음성정보로 사용했으며, 검색엔진은 카테고리 검색하는 방법과 키워드로 검색하는 방법 두 가지를 모두 이용할 수 있도록 하였다. 그림 12는 기존의 웹 게시판에 VoiceXML을 이용한 VUI를 적용하여 음성을 통해 게시물을 읽는 방법을 구현한 것이다.



(a) 게시판 순서도



(b) 게시판 초기화면

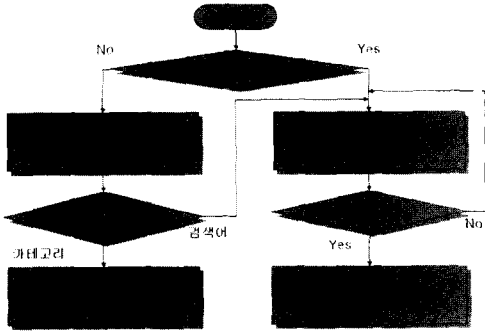


(c) 게시물 읽기

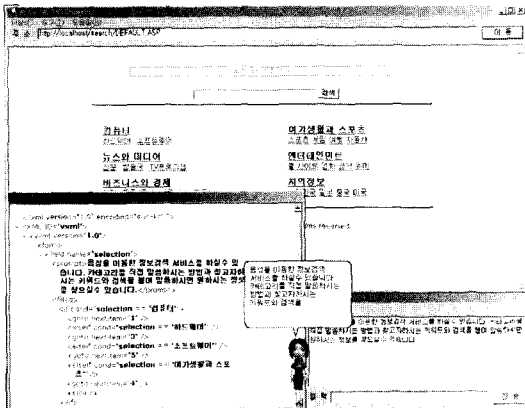
그림 12 VoiceXML VUI 브라우저의 게시판 시나리오 결과



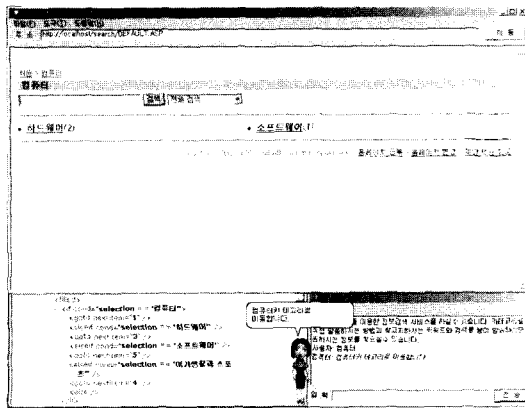
게시판에서 게시물을 선택할 경우에는 게시물의 번호를 이용하는 것이 가장 효과적이었다. 그림 12(a)는 게시판에서 제공하는 기능을 순서도로 나타낸 것이며, 브라우저로 실행된 게시판 (b)의 초기화면에서 게시물 중



(a) 검색엔진의 순서도



(b) 검색엔진 초기화면



(c) 카테고리 검색

그림 13 VoiceXML VUI 브라우저의 검색엔진 시나리오 결과

에서 사용자가 원하는 게시물의 번호를 음성 클릭함으로써 페이지가 (c)로 전이되어 게시물 내용이 TTS 엔진을 통해 서비스된다.

마지막으로 검색엔진의 경우에는 그림 13에서와 같이 카테고리별 검색엔진을 대상으로 VUI를 제공한다. 카테고리 검색엔진은 가장 기초적인 검색 엔진 중 하나로서 찾고자 하는 정보의 카테고리를 알고 있다면 다른 검색엔진에 비해 효율적인 검색을 할 수 있는 특징이 있다.

그림 13(a)는 처음 사용자에게 키워드 검색을 할 것인지 카테고리를 찾아가는 검색을 할 것인지를 묻는 절차의 순서도이며, 검색엔진 초기화면 (b)에서 컴퓨터 관련 카테고리를 선택한 후 하위 카테고리인 하드웨어를 선택하여 카테고리 내에 존재하는 정보들을 (c)처럼 획득할 수 있다. 브라우저의 전체적인 동작과정은 다음 그림 14와 같다.

5.4 VoiceXML VUI 브라우저의 비교

본 논문에서 구현한 VoiceXML VUI 브라우저와 기존의 음성 인식 브라우저의 차이점은 표 3과 같다. 본 논문에서 구현한 VoiceXML VUI 브라우저는 기존의 음성인식 브라우저[5,14,19,20]가 메뉴 등을 음성으로 작동케 하는 것과는 달리 웹상에서 GUI 환경과 병행하여 VUI를 추가적으로 지원하여 사용자에게 다양한 인터페이스를 지원하게 하고 있다. 또한 음성출력이 가능하여 사용자는 긴 글을 읽을 필요 없이 듣기를 통해 정보를 얻을 수 있다. 이는 뉴스와 같은 서비스에 활용하면 사용자는 라디오를 듣는 것처럼 다른 작업을 수행하면서 웹의 정보를 얻을 수 있다. 또한 XHTML+Voice 방식 [7]과 달리 XML Island 기법을 통해 기존 HTML 문서의 수정을 최소화하면서, VoiceXML 다이얼로그를 쉽게 재사용할 수 있다는 장점과 GUI방식과 VUI방식 인터페이스를 동시에 제공한다면 장점을 가지고 있다. 아울러 SALT[8,9]와 달리 본 VoiceXML VUI는 VoiceXML 국제 표준을 따르기 때문에 VUI를 제공하기 위해 VoiceXML 다이얼로그를 작성할 때 기준으로 삼을 수 있다. 마지막으로 본 논문에서 VoiceXML VUI 브라우저 구현시 제작한 파서는 다른 VoiceXML 응용프로그램 구현시 사용할 수 있도록 클래스로 만들어 놓음으로써 재사용성의 장점이 있다.

6. 결론 및 향후 연구과제

지금까지 국내·외에서 음성브라우저(VoiceXML 포함) 관련연구는 대부분 전화기를 이용한 시스템을 대상으로 연구되었으며, VoiceXML을 시뮬레이트하는 수준에 머물렀다. 본 논문에서는 정보획득의 주수단인 기존의 웹페이지에 VoiceXML을 적용함으로써 상호대화적 VUI 환경을 제공하는 방법을 제시하였다. 이를 위해 효

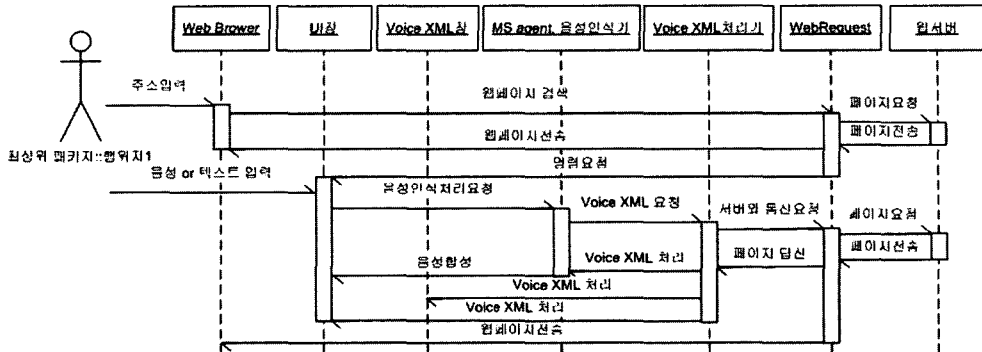


그림 14 VoiceXML VUI 브라우저의 동작과정

표 3 음성인식 브라우저 비교

| 구 분                     | 본 논문의 VoiceXML VUI 브라우저                       | 기존의 음성인식 브라우저(VoiceXML 미지원)           | 기존의 음성인식 브라우저(VoiceXML 지원)[5][14] | 비고  |
|-------------------------|---|---------------------------------------|-----------------------------------|---|
| 제품명                     | VoiceWeb V1.0                                 | Sayser(voicenic), VoiceBrowser(보이스웨어) | -                                 |   |
| HTML 지원 여부              | HTML, VoiceXML 동시 지원                          | HTML 지원                               | VoiceXML만 지원                      | HTML 문서에 XML Island 기법을 활용함으로써 VoiceXML 적용 용이     |
| 페이지 이동                  | -VoiceXML 적용 다이얼 로그를 통해 제공<br>-HTML의 하이퍼링크 방식 | -음성 메뉴 적용<br>-HTML의 하이퍼링크 방식          | -VoiceXML 적용 다이얼 로그만 제공           |   |
| 파 서                     | 자체 클래스로 제공                                    | X                                     | -                                 |   |
| 음성출력                    | ○   | X                                     | ○                                 |   |
| 이용 단말기 (경유망)            | PC(인터넷망)                                      | 전화기(PSTN망)                            | PC(인터넷망)                          |   |
| 상호대화성                   | ○   | X                                     | ○                                 | VoiceXML을 통한 Dialog 제공(메뉴 선택, 게시판, 검색엔진 시나리오 제공)  |
| 기존 HTML 정보에 VUI 적용의 용이성 | 용이  | 미적용                                   | 미적용                               | 웹페이지에 VoiceXML 쉽게 추가가능. 후자는 HTML 정보 대폭 수정함으로써 가능. |
| 개발 비용                   | 저렴  | 고비용                                   | -                                 |   |
| 유지비용                    | 저렴  | 고비용                                   | -                                 | 간단한 VoiceXML 문서 제작을 통해 용이. 후자는 전문가의 도움 필요.        |
| 서버부담                    | 적음  | 많음                                    | -                                 | 음성 인식/합성기가 클라이언트에 탑재.                             |
| 인터페이스 제공 유형             | GUI + VUI 병렬제공                                | VUI 제공                                | VUI 제공                            |   |
| 표준화                     | VoiceXML 1.0 지원                               | X                                     | VoiceXML 0.9                      |   |
| 기 타                     | 공개된 음성인식/TTS 엔진 이용                            | 전용 음성인식 엔진을 이용                        | 미연급                               |   |

과적인 VUI를 제공하기 위해서 가장 보편화된 메뉴선택, 게시판, 검색엔진 시나리오에 적용하는 방법을 제시하였다. 이는 기존 방식과는 달리 개발자가 직접 Voice-

XML 문서를 제작함으로써 웹사이트 방문자에게 능동적인 VUI를 제공할 수 있는 토대를 제공하고 있다. 하지만 본 연구 범위의 밖에 있는 JSGF 해석기와

Grammar 관련 연구는 아직까지 해결되지 않았다. VoiceXML을 좀 더 효율적으로 사용하기 위해서는 위의 두 가지가 병행 연구되어야 할 것이다.

### 참 고 문 헌

- [1] 홍용택, XML과 VoiceXML의 효율적인 통합방안 제시, 군산대학교 석사학위논문, 2002년 2월.
- [2] VoiceXML 포럼, <http://www.voicexml.org>
- [3] IBM, <http://www-4.ibm.com/software/webservers/appserv>
- [4] IBM, <http://www-4.ibm.com/software/speech/enterprise/vtoolkit.html>
- [5] 김경란, VoiceXML 기반 음성 브라우저의 설계 및 구현, 성신여자대학교 전산학과 석사학위논문, 2001.
- [6] VoiceXML 포털사이트, <http://www.voicexmlplanet.com>
- [7] 월드와이드웹 컨소시엄, <http://www.w3.org/Submission/2001/13/>
- [8] T.Raman, "Cascaded speech style sheets," in Proc. Syxth International World Wide Web Conference(M. R. Genesereth and A. Patterson, eds.), Santa Clara, CA, pp.109-117, 1887.
- [9] Kuansan Wang, "SALT: A SPOKEN LANGUAGE INTERFACE FOR WEB-BASED MULTIMODAL DIALOG SYSTEMS," ICSLP, 2002.
- [10] Peter j. Danielsen, "The Promise of a Voice-Enabled Web," IEEE Computer, VOL.33, NO.3, pp.104-106, Aug. 2000.
- [11] 마이크로 소프트웨어, 2000년 9월.
- [12] about XML, 박재호 저, 2001년 1월.
- [13] 마이크로소프트 닷넷, <http://www.microsoft.com/net/>
- [14] 김경란, 홍기형, "VXML 편집기와 음성 브라우저의 설계 및 구현", 한국정보과학회 춘계학술발표 논문집, 제27권 제1호, pp.414-416, 2002년 4월.
- [15] 보이스엔조이, <http://www.voicenjoy.co.kr>
- [16] 미디어포드, <http://www.mediaford.co.kr>
- [17] 신현경, 추상구문트리 기반의 VoiceXML 해석기 설계 및 구현, 숭실대학교 컴퓨터학과 석사학위논문, 2001년 6월.
- [18] 마이크로소프트 스피치, <http://www.microsoft.com/speech/>
- [19] 보이스닉, <http://www.voicenic.co.kr>
- [20] 보이스웨어, <http://www.voiceware.co.kr/>



### 예 상 후

2001년 군산대학교 컴퓨터정보학과 학사. 2003년 군산대학교 컴퓨터정보학과 이학석사. 2003년 4월~현재 (주)대신아이텍 사원. 관심분야는 웹기반기술(XML), 머신비전, 모션제어



### 장 민 석

1989년 연세대학교 전자공학과 공학사  
1991년 연세대학교 전자공학과 공학석사(컴퓨터통신). 1997년 연세대학교 전자공학과 공학박사(프로토콜공학). 관심분야는 웹기반기술(XML), 프로토콜공학, 소프트웨어공학