

# IIS를 이용한 VoiceXML 실험 환경 구현

권형준, 김정현, 홍광석  
성균관대학교 정보통신공학부

## An Implementation of VoiceXML Test Environment Using IIS

Hyung-Joon Kwon, Jung-Hyun Kim, Kwang-Seok Hong  
School of Information and Communication Engineering, Sungkyunkwan University

### 요 약

유비쿼터스 컴퓨팅에서 중요한 기술 중 하나로 평가되는 음성인식 및 합성기술은 인간과 컴퓨터의 상호 작용에 있어 가장 편리하고 보편적인 방법이다. 음성인식 및 합성기술을 이용한 인간과 컴퓨터 상호작용 기반의 애플리케이션의 개발을 위해 음성 확장성 생성 언어(VoiceXML)을 이용하면 음성 인식 및 합성에 관한 전문 지식이 없어도 애플리케이션 제작을 쉽게 할 수 있다는 장점이 있어서 음성인식 및 합성기술의 인프라 구축과 저변 확대를 목적으로 일부 국내 업체들은 VoiceXML을 이용한 음성 애플리케이션을 제작하고 실험할 수 있도록 VoiceXML 실험 환경을 제공한다. 본 논문에서는 기존에 공개된 실험 환경을 소개하고, 다양한 실험 환경 제공을 위해 기존에 있던 Linux기반의 실험 환경과는 다른 Windows NT기반의 IIS(Internet Information Service)를 이용한 VoiceXML실험 환경을 제안하고 구현하였다. 그 결과 ASP(Active Server Page)와 ADO(ActiveX Data Object)를 이용한 VoiceXML 음성 애플리케이션 실험이 가능한 환경을 구축하였고, 사용자 평가 결과 제안한 방법이 유효하다는 것을 확인하였다.

### I. 서 론

음성인식 및 합성기술이 유비쿼터스 기반의 미래기술로서 매우 중요하다는 지적이 여러모로 보고된 바 있다. 다보스 포럼 및 MIT대학은 21세기 10대 유망 신기술로 지정하고 있으며, 뉴욕타임지는 2002년도 미국 정보통신을 이끄는 10대 기술로 선정하였다. Gartner Group 2001년도 보고서도 음성정보처리 기술을 향후 10년 12대 핵심기술로 선정하고 있으며, 특히 미국, 유럽 등 선진국은 음성인식 및 합성기술을 21세기 정보화 사회의 핵심 기술로 규정하고 관련 기술개발 등에 대규모 투자를 하고 있다. 국내 정보통신부에서는 IT839중 3대 인프라를 BCN, USN과 함께 소프트웨어 인프라로 정하고, 중요한 한 분야로서 HCI분야를 핵심 전략 산업으로 육성할 계획을 세우고 있다.[1]

마우스나 키보드를 쓰지 않고, 사람과 이야기 하는 것처럼 컴퓨터와 의사소통을 할 수 있는 애플리케이션을 설계한다고 생각했을 때, 이러한 음성인식 및 합성기술을 이용한 애플리케이션의 제작을 쉽게 해주는 언어로서

VoiceXML을 고려할 수 있다. 불과 몇 년 전까지는 이러한 기능의 애플리케이션을 개발하는 데에 있어서 상당히 전문적인 지식과 장비가 필요했으나, XML 기반의 VoiceXML은 전문 개발자가 아닌 사람도 쉽게 다룰 수 있는 장점이 있다.[2]

국내 일부 음성인식 및 합성기술 관련 업체들은 음성 애플리케이션 개발자 및 음성인식 및 합성기술에 관심을 갖고있는 사용자들을 위하여 웹 사이트로 이루어진 VoiceXML실험 환경을 제공하고 있지만, 대부분 Linux기반의 실험 환경을 제공하여 동적인 VoiceXML문서의 작성을 위한 Server Side Language로 ASP를 활용할 수 없다. 본 논문에서는 기존의 실험 환경과 더불어 음성 애플리케이션 개발 및 사용자 인프라 구축과 저변 확대를 위하여 Linux기반의 VoiceXML실험 환경 일변도의 현재 국내 실험 환경을 인지하고, ASP와 ADO를 이용할 수 있는 장점 및 특징을 가지고 있는 Windows NT기반의 IIS를 이용한 새로운 VoiceXML실험 환경을 제안하고 실제 실험 환경을 구축한다.

## II. 관련 기술

### 2.1. 음성 인식 및 합성

인간과 컴퓨터 사이의 커뮤니케이션 방법은 유비쿼터스 컴퓨팅의 중요한 요소인데, 현재 가장 현실성 있게 각광 받고 있는 방법은 인간이 사용하는 가장 보편적이고 편리한 정보전달의 수단인 음성이다. 인간의 음성을 인식한 컴퓨터는 음성합성기술을 이용하여 응답한다.[3] 음성합성은 음성의 어느 기술보다 상용화에 더 근접된 기술이며, 각종 텍스트 정보를 음성으로 변환하는 서비스에 광범위하게 응용할 수 있다.[4]

### 2.2. VoiceXML

일부 기업에서는 수 년 동안 텔레포니와 인터넷의 기술을 통합하는 연구에 매진했고, 이러한 목표를 이루는 과정에서 마크업 언어들을 만들어 냈다. 그 중에 하나인 VoiceXML은 AT&T, Motorola, IBM등이 독립 단체인 VoiceXML Forum을 함께 구성하면서 만들어낸 가장 훌륭한 성과물이다.

VoiceXML은 값비싼 하드웨어 비용이나 난해한 개발 언어 환경에서 벗어나게 해 주는 언어로서, 음성인식 및 합성에 전문지식이 없는 개발자도 음성 애플리케이션의 개발이 가능하도록 설계되었고, 더욱 동적인 VoiceXML 페이지를 위하여 Server Side Language를 혼용할 수 있다.[2]

### 2.3. Internet Information Service

IIS는 Windows NT 커널을 기반으로 동작하며, WWW, FTP, NNTP, SMTP등을 포함하는 인터넷 서비스를 제공하고 ISAPI 표준을 제안하는 마이크로소프트 웹 전략의 근간으로서 사용의 용이성, 신속성, 이식성, 보안성 및 확장성을 추구하여 디자인 되었으며, 많은 사람들에게 의해서 어떠한 경쟁 제품보다 우수한 것으로 인정되고 있다.[5]

### 2.4. Active Server Page

ASP는 서버 측면 애플리케이션의 기본 골격을 제공하여 보다 쉽게 동적인 웹 애플리케이션을 개발할 수 있는 Server Side Language이다. ASP문서는 HTML문법과 XML문법을 모두 포함할 수 있다.

웹 서버가 ASP문서에 대한 HTTP요청을 받으면, 메모리상에 정적인 HTML정보와 스크립팅에 의해 생성된 HTML정보가 결합된 가상의 출력 HTML파일이 생성되어 클라이언트에게 전송하게 되는 구조이다.[5] 이러한 특징을 이용하여 동적인 VoiceXML문서를 생성한다.

### 2.5. Computer Telephony Integration

컴퓨터가 전화 통화를 관리하고 텔레포니 서비스를 PC와 서버 등 기타 다른 컴퓨팅 장비에 통합할 수 있도록 해주는 소프트웨어와 하드웨어 컴포넌트들의 집합을 의미한다. 서버를 기반으로 하는 전화통신과 음성 애플리케이션을 통합하여 IVR 시스템을 통해 클라이언트의 음성 및 DTMF 입력을 받아들이고, 데이터베이스를 탐색하며, 다시 음성으로 클라이언트에 정보를 전달한다.[6]

이미 콜 센터와 같은 CTI를 응용한 서비스는 클라이언트를 대상으로 한 기업의 기본적인 서비스로 자리 잡았으며, 인건비 절감의 이득을 가져다주고 있다.

## III. 기존의 VoiceXML 실험 환경

일부 기업에서는 VoiceXML실험을 위해 회원 가입 후 개인이 만든 VoiceXML문서를 업로드 하여 실험할 수 있는 환경을 제공하며, 강력한 기능의 디버거와 개발의 편리함을 위해 텍스트 에디터를 자체적으로 제작하여 배포하고 있다.

보이스웨어의 음성 합성 및 인식 엔진을 사용하는 와이더덴닷컴(<http://vxml.widerthan.com>)은 회원 가입 시 부여하는 회원 번호를 이용하여 자신이 업로드 한 VoiceXML문서를 실험할 수 있다. 함께 제공되는 디버거는 직접 전화를 걸지 않아도 웹 상에서의 가상 실험을 통해 디버깅이 가능한 강력한 기능을 제공한다.

보이스인조이닷컴(<http://www.voicenjoy.co.kr>)은 한국전산원, KT, 삼보정보통신이 주관하는 홈페이지로서 와이더덴닷컴과 유사한 VoiceXML실험 환경과 더불어 VoiceXML문서 제작 전용 에디터를 프리웨어로 배포하고 있으며, VoiceXML 음성 애플리케이션 개발자들을 위한 공개 세미나를 주최한다.

와이더덴닷컴과 보이스인조이닷컴은 Linux기반으로서 PHP를 이용한 VoiceXML문서의 실험이 가능한 특징이 있는 반면에, Windows NT의 IIS 기반에서 동작하는 ASP를 활용할 수 없다. 또한, 데이터베이스를 이용한 VoiceXML 음성 애플리케이션의 실험을 위해서는 DBMS를 개인이 준비해야 하는 단점이 있다.

## IV. IIS를 이용한 VoiceXML 실험 환경 구현

제안하는 Windows NT의 IIS를 기반으로 하는 실험 환경은 ASP를 이용하여 제작된 동적인 VoiceXML문서의 실험이 가능하며, ADO 및 DAO를 이용하여 Microsoft Access로 디자인한 MDB파일에 접근할 수 있으므로 실험을 위해 간편한 데이터베이스의 연결이 가능한 장점을 가진다.

### 4.1. 개발 환경 및 동작 환경

서버의 중앙처리장치는 안정성을 극대화하기 위하여 듀얼프로세서를 탑재한 Intel Xeon을 사용하였고, 전화와 컴퓨터의 상호작용을 위한 CTI 보드는 한 번에 12채널을 수용할 수 있는 Intel Dialogic D/120JCT-LS를 사용하였으며, 음성인식 및 합성 엔진은 KT의 휴보이스를 사용한다. 표 1은 개발 환경 및 제안한 실험 환경이 동작하는 서버의 하드웨어 및 소프트웨어의 자세한 사양을 나타낸다.

표 1. 개발 환경 및 서버 사양

항목	설명
CPU	Intel Xeon 3.0Mhz
RAM	2GB
CTI Board	Intel Dialogic D/120JCT-LS PCI(12CH)
OS	Windows 2000 Professional Service Pack 4
IIS Ver.	5.0
ASR	KT Huvoice
TTS	KT Huvoice

### 4.2. 사용자 실험 방법

제안한 실험 환경을 이용하기 위해서는 회원 가입이 필요하다. 웹 상에서의 사용자 인증 후에는 원하는 파일의 개수만큼 VoiceXML문서 혹은 ASP문서 및 MDB파일을 업로드 할 수 있다. 전화를 걸어 실험 시에 업로드 할 파일을 지정하지 않고 진행하면 그림 1과 같이 과거에 업로드 한 파일의 목록을 출력한다. 이 단계에서는 파일을 삭제할 수 있으며, 간단한 텍스트 에디터가 내장되어 있어 VoiceXML문서를 웹 상에서 수정할 수 있다.

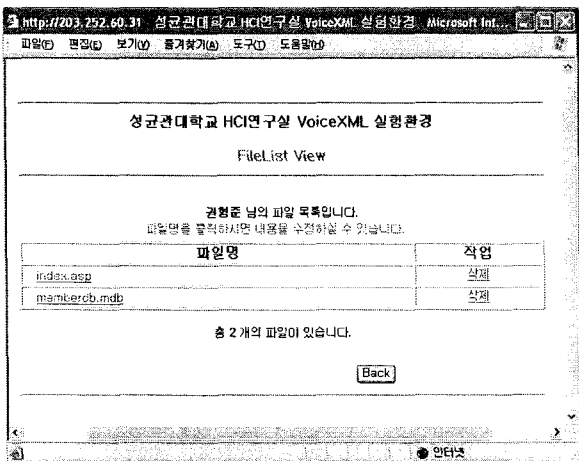


그림 1. 파일 목록 출력 화면

### 4.3. 서버 프로세스

사용자가 자신이 업로드 한 음성 애플리케이션의 실험을 위해 실험 전화에 연결하여 음성 및 주민등록번호로 사용자 인증 과정을 거치게 되면 자신이 업로드 한 VoiceXML 음성 애플리케이션이 실행됨을 확인할 수 있다. 이러한 과정에서 서버는 일련의 작업을 수행하게 되는데, 실험시의 서버 처리 과정을 그림 2에 순서도로 도시하였다.

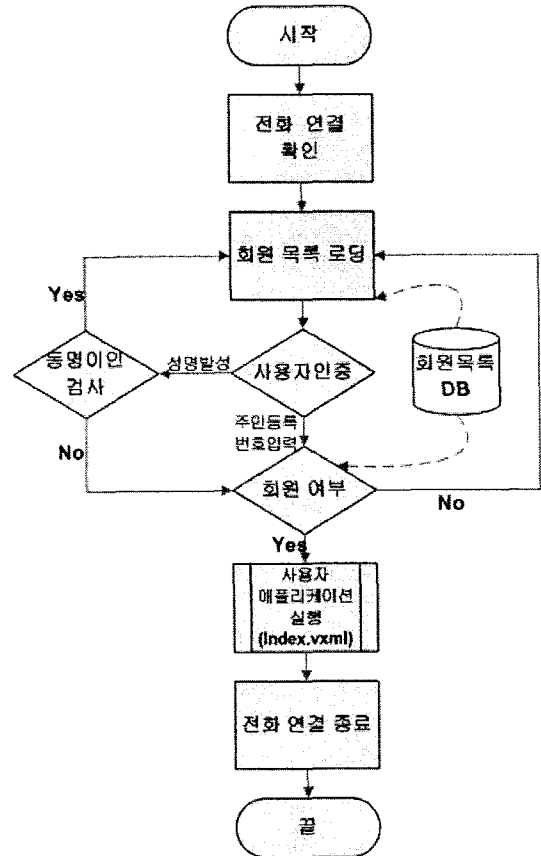


그림 2. 서버 처리 과정

## V. 실험 및 사용자 평가

### 5.1. 실험 내용

구현한 실험 환경의 신뢰성을 평가하기 위하여 ASP와 ADO를 이용한 동적인 VoiceXML 예제 파일을 작성하여 업로드 후 실험하였다. 실험에 사용한 VoiceXML 파일의 소스코드는 그림 3에 나타내었다. 간단한 예제이지만 Server Side Language와 VoiceXML을 혼용한 음성 애플리케이션 개발에 관심을 갖는 사람들에게 좋은 예제가 될 것이다.

```

<!-- db.mdb의
member 테이블 name 필드의 목록을 출력하는 예제 -->
<?xml version="1.0" encoding="euc-kr"?>
<vxml version="1.0">
  <form>
    <block>
      name 필드의 목록은 다음과 같습니다.
    <%
      set dbConn = server.createObject("adodb.connection")
      path = "DATA SOURCE="
      path = path + server.mappath(".") & "\db.mdb"
      conn = "provider=Microsoft.Jet.OLEDB.4.0;" & path
      dbConn.open(conn)
      sql = "select name from member"
      set rs = dbConn.Execute(sql)
      While not rs.eof
        Response.Write(rs("name"))
        rs.MoveNext
        if not rs.eof then Response.Write(" , ") end if
      Wend
    %>
    입니다.
  </block>
</form>
</vxml>
    
```

그림 3. ASP와 VoiceXML 혼용 예제

### 5.2. 실험 결과

그림 3의 소스 코드로 제작된 VoiceXML 문서를 업로드 하여 실행할 때의 사용자와 시스템간의 시나리오를 그림 4에 나타내었다. 동명이인이 있음을 가정하고 주민등록번호로 재입력을 받는 시나리오이다.

```

(전화 연결 시작)
System: 안녕하세요? VoiceXML 실험 환경입니다.
System: 주민등록번호를 입력하시거나, 이름을 말씀해 주세요.
User: 권형준
System: 같은 이름을 가진 사용자가 있으므로,
System: 주민등록번호를 입력해 주세요.
User: (주민등록번호 입력)
System: 어서오십시오.
System: 권형준님의 index.vxml 파일을 실행합니다.
System: name 필드의 목록은 다음과 같습니다.
System: A, B, C입니다.
(전화 연결 종료)
    
```

그림 4. 예제 실행 결과 시나리오

### 5.3. 사용자 평가 및 개선 방안

제안한 실험 환경의 성능 평가를 위한 항목을 선정하고 각 항목에 대한 최고점수를 각 문항당 1명에 5점 만점으로, VoiceXML 음성 애플리케이션 개발자 및 연구원 총 5명의 의견을 종합하여 표 2에 나타내었다. 설문 결과, 문제가 있는 소스 코드 실행 시의 불안정함 및 보안성에 문제가 있을 수 있다는 의견을 얻었으며 사용자 인터페이스의 편리함, 데이터베이스 연동과 내장 텍스트에디터와 관련된 내용에서 좋은 평가를 받았다.

표 2. 사용자 만족도

평가항목	A	B	C	D	E	합계
사용자 인터페이스	5	4	4	4	5	22/25
내장 텍스트에디터	4	4	5	4	4	21/25
데이터베이스 연동	5	4	5	4	4	22/25
서버 보안성	3	3	2	2	3	13/25
서버 안정성	3	3	3	3	2	14/25

※ 개인별 각 항목당 5점 만점 / 총점 25점

## VI. 결론

본 논문에서는 음성인식 및 합성기술의 인프라 구축과 지면 확대 및 국내 VoiceXML 실험 환경의 다양화를 위하여 IIS를 이용한 VoiceXML 실험 환경을 구현하였다. 그 결과, ASP와 ADO를 이용하여 작성된 동적인 VoiceXML 문서로 이루어진 음성 애플리케이션을 실험할 수 있는 환경을 구축할 수 있었다.

국내에 공개된 실험 환경들은 각자의 사양에 따라 모두 장단점을 가지고 있고 각자 특징들이 존재한다. 각 실험 환경들과 더불어, 함께 제공되는 디버거 및 텍스트에디터를 적절히 사용하면 VoiceXML 음성 애플리케이션 개발자 및 개발에 관심을 가지는 학생들에게 큰 도움이 될 것이며, 본 연구도 그에 일조하길 기대한다.

제안된 실험 환경은 보안성 및 안정성 실험을 거친 후 누구나 이용할 수 있도록 서비스 할 예정이다.

## 감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원 사업 및 2006년도 두뇌한국 21사업의 연구결과로 수행되었음(IITA-2005-C1090-0501-0019).

## 참고 문헌

- [1] 김병수 외, "유비쿼터스 환경의 음성언어기술 정책방향", 한국정보과학회지 제 24권 제 1호 통권 제 200호, 2006.
- [2] Eve Astrid Anderson 외, "Early Adapter VoiceXML", Wrox, 2001.
- [3] 이건상 외, "음성인식", 한양대학교 출판부, 2001.
- [4] 안병호 외, "TTS를 이용한 멀티미디어 통신서비스", 한국통신학회지 제 16권 5호, 1999.
- [5] Alex Homer 외, "Professional Active Server Page 3.0", Wrox, 2000.
- [6] William A. Yaberry, "CTI", AUERBACH, 2002.