

홈 네트워크 환경에서 음성인식기반 사용자 인터페이스를 통한 가전기기 제어 시스템 구현¹⁾

김연우*, 장현수*, 김구수**, 엄영익*

*성균관대학교 정보통신공학부

**동양대학교 정보통신공학부

e-mail:{daroobil, jhs4071, yieom}@ece.skku.ac.kr, gusukim@dyu.ac.kr

Implementation of Home Appliance Control System with Speech Recognition based User Interfaces in Home Network Environments

Younwoo Kim*, Hyun-Su Jang*, Gu Su Kim**, and Young Ik Eom*

*School of Information and Communication,
Sungkyunkwan University

**School of Information and Communication,
Dongyang University

요 약

컴퓨팅 기술의 발전에 따라 유비쿼터스 시대로의 이행이 가속화되고 있다. 이에 따라 홈 네트워크 분야에 대한 연구와 상용화를 위한 노력이 활발해지고 있다. 이와 더불어 가전기기들의 종류는 다양해지고 복잡해지면서 사용자들의 가전기기 이용에 있어 사용법을 익혀야하는 어려움이 있다. 이러한 문제점을 해결하기 위한 일환으로 디지털 장치들을 편하게 사용하기 위한 멀티 모달 사용자 인터페이스가 요구되고 있다. 본 논문에서 네트워크 가전기기 제어가 가능한 홈 네트워크 미들웨어인 UPnP를 사용하여 VoiceXML을 통한 음성인식기반 사용자 인터페이스와 디지털 장치 제어 시스템을 제안하고 구현한 후 실험하였다.

1. 서론

최근 프로세서와 네트워크의 발전으로 홈 네트워크를 위한 가전기기들의 기술 발전이 가속화되고 있다. 이러한 기술의 발전으로 다양한 홈 장비들, 예를 들어 TV, 에어컨, DVD플레이어, 전구, 냉장고 등이 가정 네트워크 접속이 가능해지고, 사용자는 이러한 장비들을 이용하여 다양한 유비쿼터스 서비스를 제공받을 수 있게 된다. 이러한 네트워크 가전기기로 구성된 시스템을 홈 네트워크 시스템이라 한다[1].

사용자의 홈 네트워크 접속 환경이 풍부해짐에 따라, 사용자의 위치에 상관없이 공통적으로 대응할 수 있는 사용자 인터페이스가 요구되고 있다. 기존의 키보드, 마우스를 이용한 사용자 인터페이스 대신 사용자 친화적인 멀티모달 사용자 인터페이스가

필요하다.

본 논문에서는 XHTML+Voice와 UPnP가 결합된 멀티모달 사용자 인터페이스를 제공하는 홈 네트워크 시스템을 제안하고 구현하였다.

본 시스템은 UPnP를 이용하여 홈 네트워크를 구성하고, 장비와 환경에 따라 적응적인 UI를 생성하는 대신[2], 범용적이고 플랫폼 독립적인 웹을 이용한 사용자 인터페이스를 제공한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 음성인식 사용자 인터페이스 기반 가전기기 장치 제어 시스템에 쓰이는 기술들에 대해 소개한다. 3장에서는 제안 시스템의 구조와 구성에 대해 설명하고, 4장에서는 시스템의 구현과 실험 과정을 소개한다. 5장에서는 구현된 시스템에 대한 평가와 결론을 내린다.

2. 관련 연구

1) 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음. IITA-2006-(C1090-0603-0046)

2.1 UPnP(Universal Plug and Play)

UPnP는 TCP/IP와 웹을 기반으로 가정과 사무실 등에서 장비들 간의 데이터 전송과 제어가 가능한 P2P 네트워크이다[3]. UPnP는 각 장비들이 분산적으로 활동하며, 오픈 네트워크 환경을 기반으로 한다.

UPnP는 장비와 플랫폼에 독립적이며, 강력한 확장성을 가지고 있다. 익히 알려진 인터넷 기술들(IP, TCP, UDP, HTTP, XML 등)을 사용하기 때문에 구현과 이해가 쉬우며, 벤더들은 기본 장비의 기반이 되는 범용 프로토콜을 제공한다.

2.2 XHTML

XHTML (Extensible Hypertext Markup Language)은 HTML과 XML이 결합된 언어다. [4] XHTML은 XML과서에게는 XML로 해석되면서, 구형 브라우저에게는 HTML로 해석된다. XHTML은 코딩 규칙이 XML만큼 엄격하기 때문에 XHTML은 다음의 장점들을 가진다.

- XHTML은 XML을 따른다. 그러므로 XHTML 문서는 표준 XML 도구를 이용하여 읽고, 수정하고, 검사할 수 있다.
- XHTML문서는 브라우저에게 HTML문서보다 쉽게 해석된다.
- XHTML문서는 HTML DOM보다 여러 방면에서 응용 가능하다.

2.3 VoiceXML

VoiceXML은 자동 음성 인식(Automated speech Recognition)과 텍스트 음성 변환(Text To Speech)을 사용하기 위한 음성 사용자 인터페이스 생성용 XML이다[5].

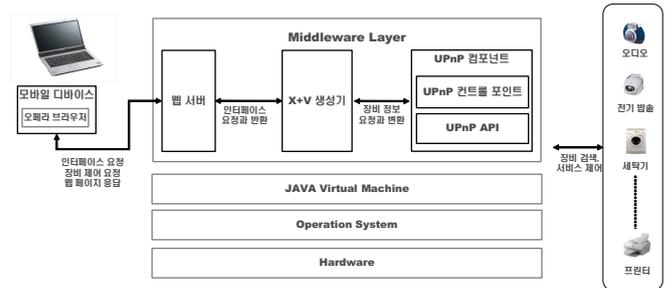
VoiceXML은 XML과 attribute format에 기반하고 있다. 최상위 태그는 <vxml>로, 이것은 VoiceXML임을 알리는 태그다. <vxml>태그의 내부는 폼이라는 대화상자로 나뉘어 있다. 폼은 <form>태그로 표현되며, id라는 속성을 갖는다. 그리고 id를 호출함으로써 커서를 폼에 위치시키도록 인식시킬 수 있다.

폼 안에는 여러 아이템들이 있다. 아이템들은 필드 아이템과 컨트롤 아이템으로 나뉜다. 필드 아이템은 요청자가 값을 채워 정보를 얻어내는 아이템이며, 컨트롤 아이템은 음성을 인식하지 않는, 작업과 상관없이 구분을 위해 존재하는 아이템이다.

3. 제안 시스템

3.1. 음성인식기반 사용자 인터페이스를 사용한 디지털장치 제어 시스템

본 논문에서 제안하는 음성인식 디지털장치 제어 시스템의 구성도는 그림 1과 같다. 이 시스템은 사용자가 UPnP 장치를 제어할 수 있도록, 브라우저에게 XHTML과 VoiceXML(이하 X+V)를 이용한 멀티모달 사용자 인터페이스를 제공하고 응답을 통해 UPnP장치를 제어한다.



(그림 1) UPnP기반 음성인식 디지털장치 제어 시스템

이 시스템은 Presentation Layer, Middleware Layer, JAVA Virtual Machine, operating System 4개의 부분으로 구성되어 있다.

Presentation Layer는 사용자로부터 장비 제어 요청을 받고, 요청 결과를 알려주기 위해 브라우저와 통신하는 계층이다. Presentation Layer는 X+V를 지원하는 오페라 브라우저가 담당한다.

Middleware Layer는 본 시스템의 중심이 되는 층으로, JAVA Virtual Machine 위에서 X+V 인터페이스 페이지를 Presentation Layer에게 제공하고, 요청에 따라 장치를 제어하는 계층이다.

JAVA Virtual Machine은 운영체제와 장비에 독립적인 미들웨어 구성을 위한 계층이다.

Operating System은 하드웨어 등의 자원을 관리하고, JAVA Virtual Machine에게 여러 자원들을 제공한다. 이 중 시스템의 중심이 되는 Middleware Layer는 웹 서버, X+V 생성기, UPnP 컴포넌트로 구성되어 있다.

- UPnP 컴포넌트는 홈 네트워크를 구성하는 UPnP 장비와의 통신을 담당한다. UPnP 컴포넌트는 UPnP 장치들을 탐색하고, 발견한 장치들을 등록한 후, 사용자의 명령에 따라 장비들에게 액션을 요청한다.

- X+V 생성기는 UPnP 컴포넌트로부터 UPnP장비에 대한 정보를 받아, 장비를 제어하는 XHTML 인터페이스 페이지를 생성한다. 이 인터페이스는 VoiceXML을 구현하고 있기에 음성인식이 가능하

다.

- 웹 서버는 사용자의 오페라 브라우저와 통신한다. 오페라 브라우저가 장비의 목록이나, 장비의 서비스 제어를 요청하면, X+V 생성기에게 웹 페이지 생성 요청을 하고, 페이지를 브라우저에게 반환한다.

3.2 음성인식

음성 인식은 VoiceXML을 이용해 이루어진다. 웹에서의 VoiceXML 인식은 XML과 HTML의 특징을 동시에 가진 XHTML을 이용하여 이루어진다.

VoiceXML과 XHTML의 결합을 X+V라 한다. 이 X+V페이지는 오페라 브라우저로 인식 가능하다.

X+V 생성기는 UPnP 장비 정보로 X+V 음성인식 페이지를 생성한다. 생성된 X+V 페이지에는 프롬프트 기능과 필드 기능이 구현되어 있다.

프롬프트 기능은 인수 입력 창을 클릭 했을때, 해당 인수 입력 창에 설명을 TTS(Text To Speech) 기능을 사용하여 음성으로 출력해주는 기능이다.

필드 기능은 인수 입력 창이 선택된 상태에서 음성을 입력받으면 ASR(Auto Speech Recognition)기능을 사용하여 텍스트 값을 입력받는 기능이다.

3.3 시스템 동작 흐름도

사용자가 인터페이스를 요청하는 동작, 인터페이스를 통해 액션을 호출하는 동작은 그림 2와 같다.



(그림 2) 시스템 동작 흐름도

장비의 인터페이스 생성은 사용자의 음성인식 웹 페이지 방문으로 시작된다. 웹 브라우저는 웹 서버에게 음성인식 웹 페이지를 요구한다. 웹 서버는 X+V 생성기에게 X+V 페이지 생성을 요구한다. X+V 생성기는 UPnP 컴포넌트로부터 받은 정보로 X+V 인터페이스를 생성하여, 웹 서버에게 넘겨준다. 최종적으로 웹 서버는 웹 브라우저에게 생성된 인터페이스를 보여준다.

액션 호출은, 사용자가 위 동작에서 생성한 인터

페이스에 응답하며 시작된다. 웹 브라우저는 웹 서버에게 웹 페이지를 요구한다. 이때 인자로 호출하는 액션과 관련 값들이 전송된다. 웹 서버는 UPnP 컴포넌트에게 호출 액션, 관련 인자를 넘겨준다. UPnP 컴포넌트는 목표장비를 호출 한 후, 반환 값을 가져온다. 반환 값에 따라 X+V 생성기는 결과 페이지를 생성하고, 웹 서버에게 넘겨준다. 최종적으로 웹 서버는 웹 브라우저에게 호출 결과를 보여준다.

4. 구현 및 평가

UPnP기반 음성인식 디지털 장치 시스템을 구현하기 위한 환경은 다음과 같다. 운영체제는 Windows XP, 사용 언어는 Java, CPU는 Intel Pentium, 웹 브라우저는 오페라, IDE는 Eclipse, 그 외 사용 라이브러리에 CyberLink for Java[6], Apache Xerces (XML Parser)[7]등을 사용했다.

4.1. UPnP 컴포넌트 구현

UPnP API는 CyberLink[6]를 이용하였다. 이 API를 이용해 구현된 UPnP 컨트롤 포인트는 홈 네트워크를 구성하는 UPnP 장비와의 통신을 담당한다.

UPnP 컨트롤 포인트는 SSDP (Simple Service Discovery Protocol)를 통해 UPnP장비들을 탐색한 뒤, 이 후 X+V 생성기가 인터페이스 페이지를 생성할 때 필요한 장비에 대한 자료들을 가져온다. UPnP장비의 description.xml을 통하여, 그림 3과 같은 서비스 설명 파일을 찾고 분석하여 장비의 서비스와 액션에 관한 정보를 가져온다. 이 정보들은 미들웨어에 기록된 후 X+V 생성기가 요청할 때 정보를 제공해준다. 이때 XML 파일을 분석 하는데 필요한 XML 파서는 Apache Xerces[7]를 사용하였다.

```

<action>
  <name>SetPower</name>
  <argumentList>
    <argument>
      <name>Power</name>
      <relatedStateVariable>Power</relatedStateVariable>
      <direction>in</direction>
    </argument>
    <argument>
      <name>Result</name>
      <relatedStateVariable>Result</relatedStateVariable>
      <direction>out</direction>
    </argument>
  </argumentList>
</action>
<stateVariable sendEvents="yes">
  <name>Power</name>
  <dataType>boolean</dataType>
  <allowedValueList>
    <allowedValue>0</allowedValue>
    <allowedValue>1</allowedValue>
  </allowedValueList>
  <allowedValueRange>
    <maximum>123</maximum>
    <minimum>19</minimum>
    <step>1</step>
  </allowedValueRange>
</stateVariable>
<stateVariable sendEvents="no">
  <name>Result</name>
  <dataType>boolean</dataType>
</stateVariable>
</serviceStateTable>
    
```

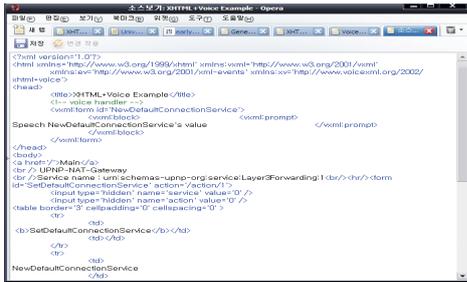
(그림 3) Service Description

UPnP 컨트롤 포인트는 사용자의 요청을 UPnP

장비에게 전달한다. 이때 전달되는 대상 액션 정보와 인자 값들은 웹서버를 통해 웹 브라우저로부터 받는다. 호출이 완료되면, 장비로부터 받은 반환 값을 웹서버를 통해 웹 브라우저로 보내준다.

4.2. X+V 생성기 구현

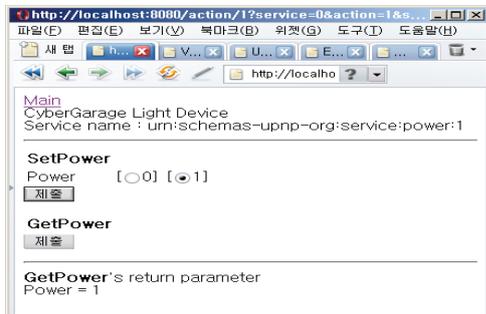
X+V 생성기는 UPnP 장비를 제어하기 위한 X+V 인터페이스 페이지를 생성한다. 생성된 X+V 인터페이스 페이지 소스는 그림 4와 같다.



(그림 4) 생성된 X+V 인터페이스 소스

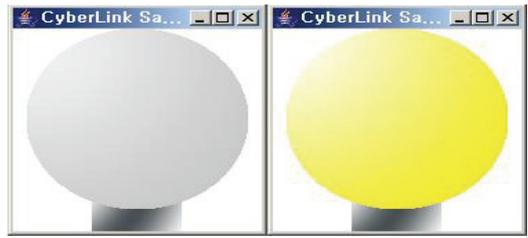
4.3. 웹 서버 제작

웹 서버는 웹 브라우저와의 웹 페이지 요청에 응답하는 역할이다. 웹 서버는 고유의 특수 태그를 가지고 있다. <UPNP_DEVICE_LIST>태그는 등록된 UPnP 장비 목록을 출력시키는 명령어다.



(그림 5) 액션 호출 UI

호출 URL을 /device/[장비번호]로 하면, 그림 5와 같이 해당 장비를 제어할 수 있는 페이지가 전송된다. 호출 URL을 /action/[장비번호]?service=서비스번호&action=액션번호&[인수 값]로 하면 입력 값을 바탕으로 UPnP 컴포넌트에게 장비에게 액션을 요청한다. 실험에 이용된 가상 UPnP장비는 PC상에 애플레이트된 것으로 그림 6과 같다.



(그림 6) 전등이 조작되어 작동되는 모습

5. 결 론

본 논문은 홈 네트워크 UPnP와 VoiceXML을 이용한 음성 인식 멀티모달 인터페이스를 제공하는 디지털 장치 제어 시스템을 제안한 후 구현하고 실험하였다.

음성인식기반 사용자 인터페이스를 이용한 디지털 장치 제어 시스템은 웹을 기반으로 운영체제와 하드웨어에 무관한 사용자 인터페이스를 제공하며, 자바 가상 머신을 이용한 홈 서버 구현으로 운영체제와 하드웨어에 독립적인 환경 구축이 가능하다.

하지만 VoiceXML을 지원하는 웹 브라우저가 한정적이기에, VoiceXML을 지원하는 유일한 웹 브라우저인 오페라 브라우저가 지원하지 않는 환경에서는 음성 인식이 불가능하다는 한계를 가지고 있다.

향후 자바 가상 머신을 이용한 플랫폼 독립적인 인터페이스 제공에 관한 연구를 통해 보완할 것이다.

참고문헌

- [1] A. Vilei, G. Convertino, and F. Crudo, "A New UPnP Architecture for Distributed Video Voice over IP," ACM 5th Int'l Conf., No. 2, 2006.
- [2] J. Nichols, B. Myers, and B. Rothrock, "UNIFORM: Automatically Generating Consistent Remote Control User Interfaces," SIGCHI Conf., pp.611-620, 2006.
- [3] UPnP, <http://www.upnp.org>
- [4] W3C, <http://www.w3.org/>
- [5] VoiceXML Forum, <http://www.voicexml.org/>
- [6] Satoshi Konno, "CyberLink for Java", <http://www.cybergarage.org/net/upnp/java/>
- [7] Apache Xerces (XML Parser), <http://xml.apache.org/>