멀티모달 상황인지 미들웨어 기반의 홈앤(HomeN) 매니저 시스템

안세열, 박성찬, 박성수, 구명완, 정영준, 김명숙
KT 미래기술연구소

HomeN manager system based on multimodal context-aware middleware

Seyeol Ahn, Sung-Chan Park, Seong-Soo Park, Myung-Wan Koo, Yeong-Joon Jeong, Myung-Sook Kim
Advanced Tech. Lab, KT
E-mail : {syahn, rapport, soopark, mwkoo, yjjung, mskim}@kt.co.kr

Abstract

The provision of personalized user interfaces for mobile devices is expected to be used for different devices with a wide variety of capabilities and interaction modalities. In this paper, we implemented a multimodal context-aware middleware incorporating XML-based languages such as XHTML, VoiceXML. SCXML uses parallel states to invoke both XHTML and VoiceXML contents as well as to gather composite multimodal inputs or synchronize inter-modalities through man-machine I/Os. We developed home networking service named "HomeN" based on our middleware framework. It demonstrates that users could maintain multimodal scenarios in a clear, concise and consistent manner under various user's interactions.

I. 서론

웹응용 프로그램을 수행하기 위한 단말과 환경은 점점 복잡해지고 있으며, 사용자들은 시스템에게 보다 강력한 인터페이스를 요구하게 되었다. 특히 모바일 환경의 단말과 응용 프로그램은 사용자에게 보다 직관적, 환경 적응적이고 선택적인 기계와의 상호작용을 필요로 하게 되었다. 이에 따라 사용자에게 다수의 모달리티가 제공돼야 하는 상황에 이르렀다.

전통적으로 멀티모달 시스템에 대한 연구는 다양하게 제시되어 왔다. 그 중 SALT(Speech Application Language Tags)라는 마크업 언어를 사용하는 하나의 제안[1]은 HTML안에 음성인식, 합성 관련 태그를 정의하였기 때문에 동시에 사용 가능하고 멀티모달 처리가 가능한 형태이다. 그러나 이 경우 SALT는 음성 마크업을 자체적으로 정의하여 기존의 단일 모달용으로 제작되었던 VoiceXML[2]을 더 이상 사용할 수 없다. 현재는 단일 모달리티용으로 제작된 기존의 형식을 바꾸지 않고 재사용할 수 있는 구조로 가고 있다[3].

이 논문에서는 멀티모달 상황인지 미들웨어(이하 미들웨어)를 이용하여 홈네트워크 서비스(이하 홈엔)에 적용하였다. 본 미들웨어는 상태 차트 XML (SCXML 또는 State chart XML) [4] 대화 모델링 언어를 도입하였는데 SCXML은 대화 매니저(Interaction Manager)에서 동작한다. SCXML과 XHTML, VoiceXML은 서로 느슨한(loosely-coupled) 관계를 갖도록 하여 단말과 서버간의 독립성을 꾀하였다.

대화매니저는 SCXML로 작성된 시나리오 스크립트를 불러온 다음 XHTML, VoiceXML과 같은 XML 마크업 언어를 처리하는 모달리티 콤포넌트를 구동한다. 또한 대화매니저는 사용자와 시스템 간의 상호작용을 하는 데 있어 이들 모달리티 콤포넌트를 제어, 감독할 뿐 아니라 사용자로부터 전달된 순차 또는 동시 멀티모달 입력을 처리하거나 외부 모듈과 통신한다.

먼저 VoiceXML 혹은 XHTML 마크업 언어를 이용하여 모달리티 컴포넌트의 기능을 수행하는 시나리오와 SCXML 마크업 언어를 이용하여 이들을 제어,

감독하는 멀티모달 시나리오를 제작한다. 제작된 스크립트에 의해 모달리티 컴포넌트는 멀티모달 서버에 플러그인 방식으로 동작하고 미리 제작된 시나리오에 따라 멀티모달 서비스가 미들웨어 상에서 제공된다.

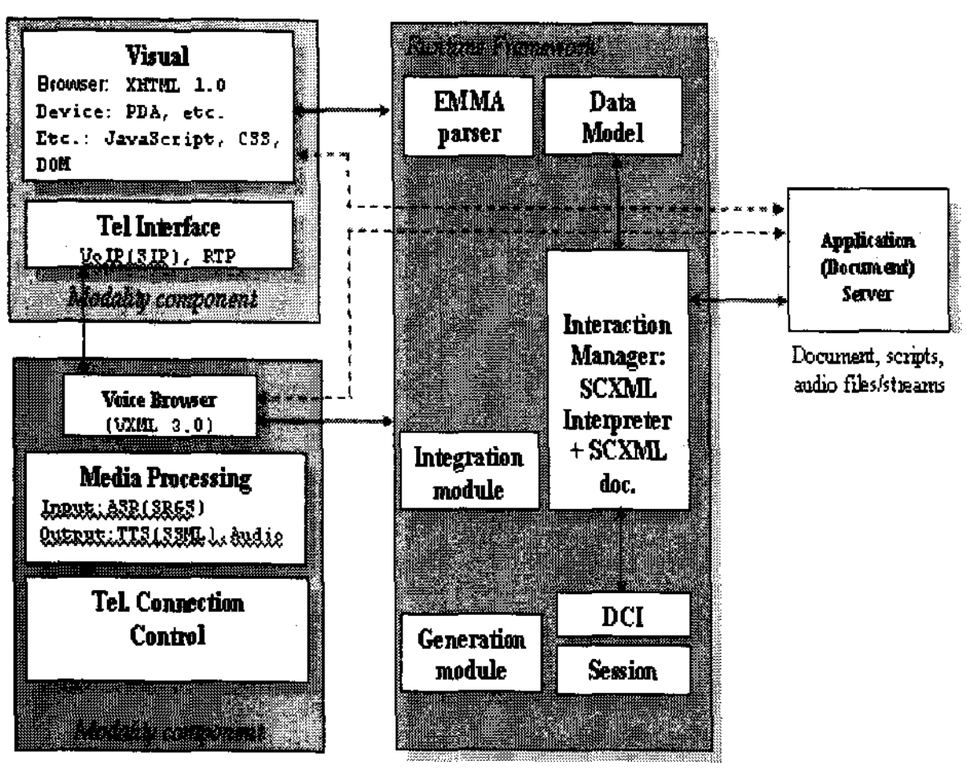
우리는 본 미들웨어를 이용하여 홈앤 매니저 시스템에 멀티모달 인터페이스를 추가하였다. 홈엔 서비스는 미들웨어의 상황인지보다는 멀티모달 인터페이스에 주안점을 두었다.

본 논문의 구성은 다음과 같다. 2장에서는 본 미들웨어의 구성요소를 기술하고 3장에서는 SCXML로 동작하는 멀티모달 통신 방법을 설명한다. 4장에서는 홈엔 매니저 서비스를 소개하고 5장에서 결론을 내린다.

II. 미들웨어

2.1 개요

그림 1에 도시된 바와 같이 미들웨어는 크게 모달리티 콤포넌트가 동작하는 클라이언트, 서버 그리고 외부 모듈(응용 프로그램 서버)로 구성된다.



<그림 1> 멀티모달 상황인지 미들웨어 구조

2.2 클라이언트

클라이언트는 멀티모달 응용 프로그램을 구동할 수 있는 휴대용 단말 또는 PDA다. 비음성 입출력의 경우 XHTML 1.0 브라우저에서 처리한다.

PDA에서 발생한 입력중 음성의 경우는 통상적인 분산형 DSR 인식 방법을 따랐다. 마이크 입력을 분석한 뒤에 음성 특징 추출기로부터 특징 데이터를 얻는다. 단말에서 끝점추출과 특징추출의 전처리를 담당하여 그 결과를 음성 서버로 전송하게 된다. 음성 서버는 후처리를 담당하여 ASR에서 인식 결과를 얻은 후 EMMA포맷으로 바꾸어 런타임 프레임워크의 인터랙션 매니저로 전달한다.

EMMA는 음성, 펜, 키보드, 필기체 등 사용자의 입력을 받아 환경 정보를 메타 데이터에 실어 XML스타일로 정형화한 데이터 구조[5]로서 멀티모달 시스템에서 이질적인 콤포넌트간 데이터 교환을 가능하게 해준다.

비음성인 경우 XHTML브라우저는 키입력을 받아 XHTML브라우저에서 EMMA 포맷으로 변환하여 멀티모달 실행 프레임워크으로 전송한다. 여기서 전송 포맷은 상시 또는 비상시의 특정 통신 프로토콜에 국한되지 않고 1-1 통신이 가능한 임의의 네트워크에 의해 연결된다.

멀티모달 출력에서 음성의 경우는 음성 브라우저의 TTS로부터 인코더에서 음성콘텐츠를 인코딩하여 PDA로 전송하면 디코더에서 디코딩 과정을 거쳐 최종 합성음을 스피커로 내보낸다.

2.3 실행(runtime) 프레임워크

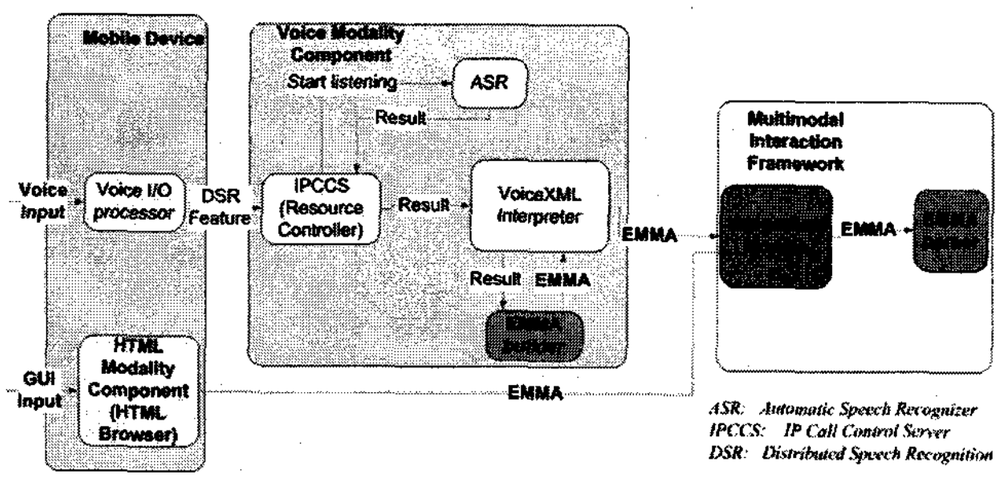
멀티모달 실행 또는 인터랙션 프레임워크는 어플리케이션의 시작, 종료등의 제어, SCXML의 해석, 그리고 모달리티 콤포넌트와의 통신을 담당한다. 모달리티 콤포넌트에서 전달된 원시 입력 데이터는 EMMA해석기에서 SCXML 실행 엔진에서 직접 다룰 수 있는 콘텍스트로 변환된다.

실행 프레임워크의 구성요소를 상술하면 데이터의 논리적 구조를 담당하는 데이터 세션 관리 모듈, 입출력을 관리하는 입출력 프로세서, 대화 흐름을 제어하는 SCXML 실행 엔진, EMMA해석기, SCXML해석기, 각종 환경 정보 또는 사용자 선호 정보가 DB로 저장되어 있는 DCI(Delivery Context Interface) [6]이다.

III. 멀티모달 통신

3.1. EMMA 생성과 해석

본 미들웨어는 멀티모달 입력에 대하여 XML코드로 변환된 EMMA를 이용하였다. 각종 메타데이터를 더하여 XML스타일로 변환된 EMMA는 멀티모달 실행 프레임워크에서 클라이언트 도메인으로부터 전달 가능한 콘텍스트 파라메터로서 취급된다. 모바일 단말의 각 클라이언트 도메인은 사용자 입력에 대하여 상호작용하는 채널을 제공하여 입력들은 병렬 혹은 순차적인 방법으로 처리된다. <그림 3>은 각 클라이언트 도메인, 즉 모달리티 콤포넌트에서 발생된 입력에 대하여 EMMA로 변환되어 멀티모달 인터랙션 프레임워크에 전달되는 과정을 보인 것이다.



<그림 3> EMMA 마크업을 이용한 사용자 입력의 생성과 해석

음성입력의 경우, 사용자로부터 발생된 원시 음성 신호는 모바일 응용 프로그램의 실행시에 음성 I/O 프로세서에서 처리된다. 원시 음성 신호로부터 DSR(Distributed Speech Recognition)엔진이 동작한다. 여기서 음성특징이 추출되면 RTP(Real-time transport protocol)에서 리소스 콘트롤러인 IPCCS로 넘겨지고 ASR서버는 음성인식 결과를 얻는다. 이어 VoiceXML 인터프리터는 인식 결과를 EMMA 빌더로 보내어 XML 형식의 EMMA데이터를 되돌려 받은 뒤 멀티모달 인터랙션 프레임워크로 전달한다.

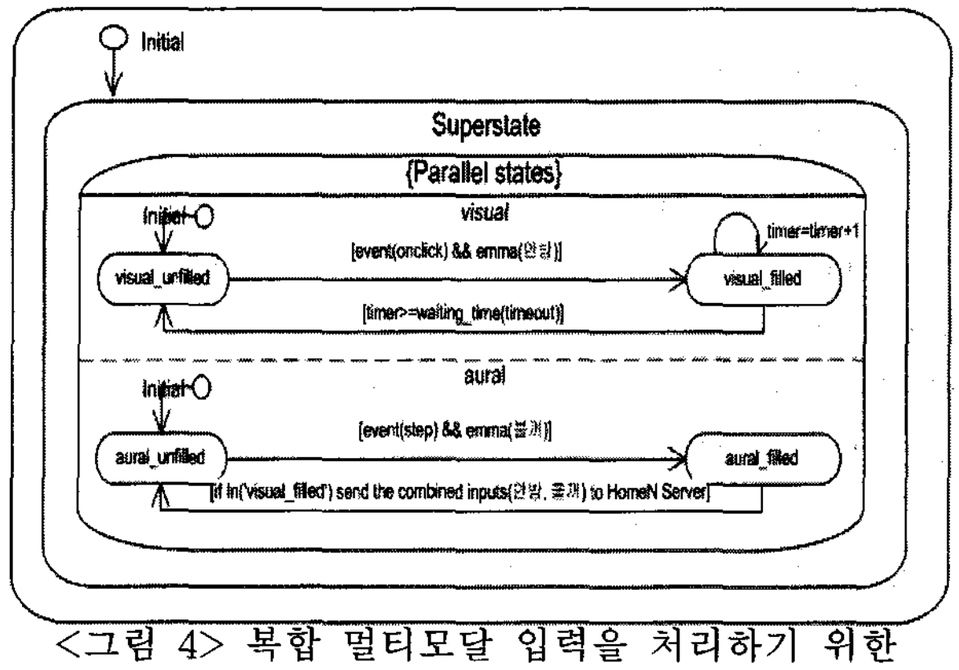
비음성 입력의 경우, XHTML브라우저의 GUI입력은 PDA브라우저에서 바로 EMMA 형식으로 변환되어 간단히 멀티모달 인터랙션 프레임워크로 전달된다.

3.2. 데이터 세션 관리 모듈

상이한 모달리티로부터 들어온 EMMA입력은 실행 프레임 워크의 통합 모듈에서 처리된다. EMMA 파서는 EMMA XML트리를 순회하면서 JEXL 표현언어의 해쉬맵(hash map) 형태의 속성과 값의 쌍으로 정렬된 리스트로 변환한다. JEXL 표현 언어로 변환된 사용자 입력은 SCXML 스크립트 안에서 자유롭게 기술이 가능하며 SCXML은 이들 값을 외부 콘텍스트로 간주되고 데이터 세션 관리모듈에서 관리한다.

3.3. SCXML의 해석과 상태 기계 모델링

SCXML 도큐먼트가 SCXML 파서에 의해 해석되면 자바 오브젝트 모델이 생성된다. 사용자 입력에 따라 결정적(deterministic) 방식으로 상태기계가 동작한다. SCXML이 사용자에게 두 가지 모달리티로 입력을 허용한다면 두 개의 병렬 상태가 생성되어 활성화된다. 본 시스템의 병렬상태는 XHTML과 VoiceXML 모달리티 컴포넌트로 맵핑되었다.



<그림 4> 복합 멀티모달 입력을 처리하기 위한 SCXML 다이어그램

3.4. 복합(결합) 멀티모달

복합 멀티모달은 보조 결합 멀티모달이라고도 하며 자세한 정의는 [7]에 나와 있다. 사용자가 "안방"을 클릭하면서 "불꺼"라고 발성하는 시나리오를 가정해 보자. 사용자로부터 얻어지는 입력은 장소 + 동작의 복합 멀티모달 입력이다. 각 입력된 부분 정보가 해석되고 그에 따라 모달리티의 상태가 전이하면 어떤 상태에 존재하는가에 따라 입력값이 통합되거나 초기화된다.

3.5. 멀티모달의 입력 처리

<그림 4>는 대화 매니저에서 수행되는 SCXML의 동작을 다이어그램으로 나타낸 것이다. 멀티모달 복합 입력 동작을 상술하면 다음과 같다. 멀티모달 실행 프레임워크 XHTML과 VoiceXML의 입력 상태를 나타내는 2개의 병렬상태 visual과 aural을 모델링하여 활성화한다. XHTML과 VoiceXML 브라우저가 접속되면 사용자 입력대기 상태가 된다. 각 모달리티에서 사용자 입력이 발생하면 조건 전이에 의해 그 이벤트의 속성이 부분정보인가 완전정보인가를 판별한다. 그 정보가 완전정보이면 다른 모달리티의 입력 대기 없이 병렬상태를 종료하고 다른 쪽 모달리티에 동기화 신호를 보낸다.

만약 XHTML 입력 이벤트의 속성이 위치정보("안방")라면 이 입력은 부분정보로서 visual_unfilled에서 visual_filled의 상태로 전이한다. 이 때 안방의 속성값은 SCXML 로컬 변수에 저장된다. 마찬가지로 입력 수신 대기 상태의 다른 쪽(음성) 모달리티에서 타임아웃 이내에 부분정보인 음성입력 "불꺼"를 수신하면 aural_unfilled 상태에서 aural_filled의 상태로 전이한다. 이 때 병렬 상태의 모든 자식 상태는 각각 visual_filled와 aural_filled에 존재한다. 이때 aural의 엿보기(In) 기능에 의해 SCXML은 병렬 상태를 초기화하면서 각 부분정보들을 통합하여 홈엔 서버에 전송한다. 만약

입력 이벤트가 발생하지 않고 타임아웃이 발생하면 값이 채워지지 않은 상태로 초기화된다.

3.5. 멀티모달 출력 처리

멀티모달 출력은 간단하다. 모달리티 수만큼 병렬 상태를 초기화한 후 음성 합성기 또는 디스플레이 장치 등에 구동 메시지를 보낸다. 병렬 상태의 속성상 메시지 전송은 동시에 이루어지며 출력 상황은 사용자가의 인터럽트 이벤트 혹은 타임아웃 이벤트가 발생할 때까지 지속된다.

Ⅳ. 멀티모달 홈엔(HomeN) 서비스

4.1. 홈엔 서비스 개요

홈엔((HomeN)이란 KT에서 시범서비스 중인 초고속 인터넷 기반 홈네트워크 서비스이다. 미들웨어를 홈엔의 여러 서비스 중 가전기기 제어에 연동하여 멀티모달 제어가 가능하도록 웹응용 프로그램을 제작하였다.

<그림 5> PDA초기화면(좌측)와 제어화면(우측)

<그림 5>에서 초기화면으로부터 음성 또는 펜입력 모드로 메뉴 탐색을 할 수 있다. 가전기기 제어는 펜과 음성을 선택적으로 사용하거나 펜 + 음성과 같은 복합 멀티모달 제어도 가능하다. 제어 목록은 난방, 가스밸브, 조명, 커튼, 콘센트, 도어락 등이다.

메뉴이동시 해당 모달리티의 이벤트가 전달되면 SCXML은 다른쪽 모달리티 콤포넌트에 메시지를 보내어 동기화를 수행한다. 가전기기 제어는 SCXML이 각 모달리티 콤포넌트로부터 제어 파라메터를 넘겨받아 통합하여 홈엔 서버로 넘겨주도록 작성되었다.

Ⅴ. 결론

본 논문에서는 미들웨어를 홈엔에 적용하였다. 멀티모달 실행 프레임워크는 미들웨어의 근간으로서 다른 모달리티 컴포넌트와 통신할 수 있는 제어권을 갖고 SCXML을 해석하여 처리한다. 모달리티 컴포넌트는 사용자의 시스템간 상호작용할 수 있도록 마크업 언어로 동작하는 클라이언트로 규정하고 있다. XHTML과 VoiceXML은 이러한 상호작용을 위한 시나리오를 작성하는 데 이용된다. 대화 매니저는 SCXML 시나리오를 해석하여 복수의 모달리티 컴포넌트로부터 독립 또는 복합 입력을 처리한다.

본 미들웨어를 홈엔에 적용시켜 본 결과 미들웨어 기반에서 멀티모달 응용 프로그램이 효율적인 사용자 인터페이스로 운용될 수 있음을 보여주었다.

참고문헌

[1] Speech Application Language Tags(SALT) 1.0 Specifications http://www.saltforum.org/saltforum/downloads/SALT1.0.pdf, July, 2002

[2] Voice Extensible Markup Language(VoiceXML) Version 2.0, http://www.w3.org/TR/voicexml20, 2004

[3] James A. Larson, Raman, Dave Raggett, "W3C Multimodal Interaction Framework" World Wide Web Consortium (W3C), May 2003; http://www.w3.org/TR/mmi-framework.

[4] RJ Auburn, Jim Barnett, Michael Bodell, T.V. Raman, "State Chart XML (SCXML): State Machine Notation Abstraction 1.0" World Wide Web Consortium (W3C), July 2005; http://www.w3.org/TR/WD-scxml-20050705/.

[5] Wu Chou, Deborah A. Dahl, Gerry McCobb, Dave Raggett, "EMMA: Extensible MultiModal Annotation markup language" World Wide Web Consortium (W3C), September 2005; http://www.w3.org/TR/emma/

[6] Keith Waters, Rafah Hosn, Dave Raggett, Sailesh Sathish, and Matt Womer, editors. "Delivery Context Interfaces (DCI) Accessing Static and Dynamic Properties" World Wide Web Consortium, 2004; http://www.w3.org/TR/2005/WD-DPF-20051111/

[7] 홍기형, "음성기반 멀티모달 인터페이스 및 표준", 말소리 제 51호, 2004.