미들웨어 기반의 텔레매틱스용 멀티모달 인터페이스

박 성 찬, 안 세 열, 박 성 수, 구 명 완
KT 미래기술연구소

A Multimodal Interface for Telematics based on Multimodal middleware

Sung-Chan Park, Seyeol Ahn, Seong-Soo Park, Myoung-Wan Koo
Advanced Tech. Lab, KT
E-mail : {rapport, syahn, soopark, mwkoo}@kt.co.kr

Abstract

In this paper, we introduce a system in which car navigation scenario is plugged multimodal interface based on multimodal middleware. In map-based system, the combination of speech and pen input/output modalities can offer users better expressive power. To be able to achieve multimodal task in car environments, we have chosen SCXML(State Chart XML), a multimodal authoring language of W3C standard, to control modality components as XHTML, VoiceXML and GPS. In Network Manager, GPS signals from navigation software are converted to EMMA meta language, sent to MultiModal Interaction Runtime Framework(MMI). Not only does MMI handles GPS signals and a user's multimodal I/Os but also it combines them with information of device, user preference and reasoned RDF to give the user intelligent or personalized services.
The self-simulation test has shown that middleware accomplish a navigational multimodal task over multiple users in car environments.

I. 서론

최근 유무선 인터넷과 이동통신 기술의 발달로 위치 기반 서비스(LBS: Location Based Service)에 대한 사용자의 요구사항이 증가하고 있다. 차량 내비게이션으로 대표되는 위치기반 서비스는 사용자의 단순 위치만을 제공하는 초기 형태에서 차량 내에서 위치 정보를 이용한 다양한 컨텐츠 및 응용 서비스를 제공하는 복잡한 형태로 바뀌고 있다. 무선인터넷과 단말기 성능의 향상으로 사용자들은 동적 환경에서 이러한 서비스를 거의 무제한으로 이용할 수 있게 되었다.

그러나 차량 내의 내비게이션 기기를 제어하고자 할 때, 키보드나 마우스는 사용이 불편한 데다 운전 시 손과 눈을 쓰기 어려운 상황이 종종 발생한다[1]. 이 경우, 태스크의 성질에 따라 사용자가 이를 수행하는 데 필요한 모드의 변화와 다양화가 요구된다.

이와 같이 환경적 제약이 따르는 경우, 지원하는 입출력 모달리티에 대하여 사용자는 최적의 입출력 모드를 선택하여 활용하는 방안이 모색되고 있다. 멀티모달 인터페이스는 주어진 상황에 따라 선택적 또는 복합적으로 다수의 모달리티를 지원하여 사용자의 태스크 수행능력을 극대화시킨다.

실용적인 멀티모달 시스템은 마크업 태그를 이용하여 웹과 연동하는 방향으로 연구되어 왔다[2]. 상태 차트 XML(SCXML, [3])은 현재 W3C에서 제안하는 멀티모달 시나리오 저작용 언어로 XHTML, VoiceXML [4]와 같은 모달리티 콤포넌트를 제어하여 음성 또는 비음성 모달리티 입출력을 처리한다. SCXML은 제어와 사용자 인터페이스를 분리하여 모달리티의 추가 등 확장성이 뛰어난 장점이 있다.

본 논문에서는 SCXML 기반의 멀티모달 미들웨어(이하 미들웨어[5])를 구현하여 멀티모달 인터페이스를 텔레매틱스 서비스에 적용하였다.

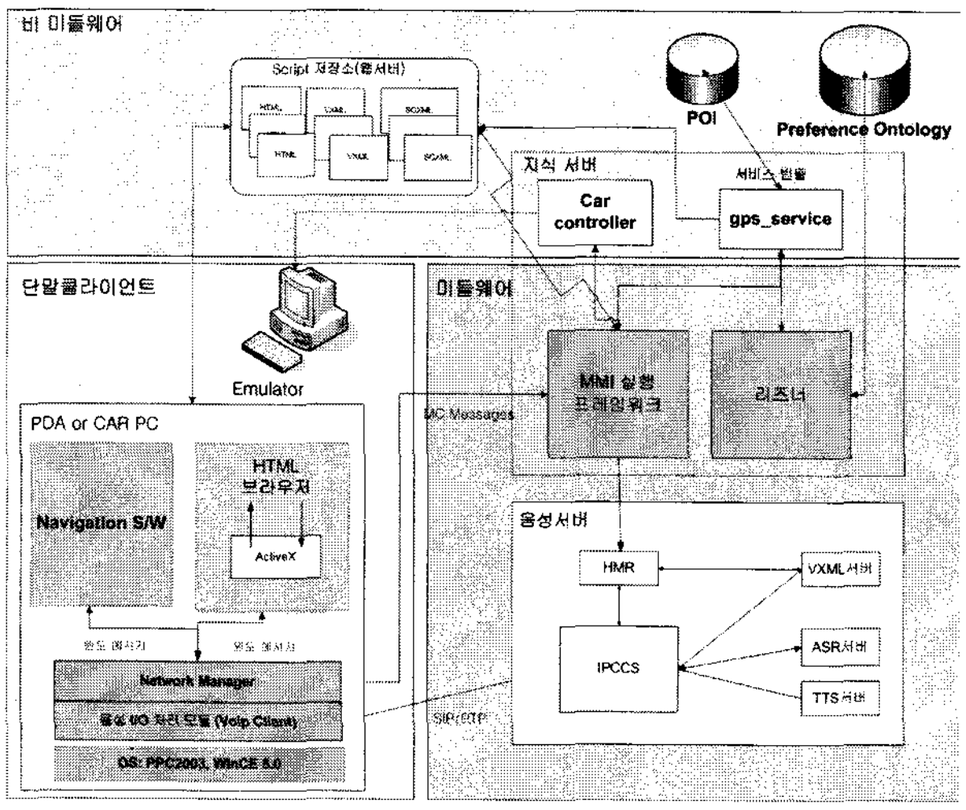
클라이언트 단말에서 발생된 HTML, VoiceXML 입력이나 GPS 신호는 네트워크 매니저에서 EMMA[6] 언어로 변환된다. 특히, GPS 신호는 주기적으로 미들웨어에 전달되어 Multi-Modal Interaction Runtime Framework(이하 MMI)의 로컬 변수로 저장된다. MMI는 SCXML로 하여금 전달된 EMMA를 이용하여 순차 또는 복합 멀티모달 입력을 처리하게 한다. 차량 관련 응용 서비스의 경우, 지식서버는 사용자의 멀티모달 입력, GPS신호, 각종 메타 정보를 모아 질의문 형식으로 서비스 번들에 보낸다.

리즈너는 선호(Preference) 온톨로지와 POI(Point Of Interaction) DB를 참조하여 RDF로 결과를 얻어낸다[7]. 지식서버의 서블릿은 RDF를 이용하여 동적으로 VoiceXML, HTML 스크립트를 생성하여 웹서버에 저장하면 단말 클라이언트는 스크립트를 수행한다.

본 논문의 구성은 다음과 같다. 2장에서는 텔레매틱스용 미들웨어 구조를 기술한다. 3장에서는 SCXML로 동작하는 위치기반 멀티모달 인터페이스를 디자인한다. 4장에서는 실험을 통해 서비스에 대한 미들웨어의 성능을 진단하고 5장에서 결론을 맺는다.

II. 텔레매틱스용 미들웨어

2.1 개요



<그림 1> 텔레매틱스용 미들웨어 구조

<그림 1>에 도시된 바와 같이 전체 구조는 비 미들웨어, 미들웨어, 단말 클라이언트로 나누어진다. 미들웨어는 지식서버의 MMI와 리즈너 그리고 음성서버로 구분된다. 음성 서버는 음성 인식과 합성을 처리하기 위한 IPCCS(IP call control server), HMR(Huvois Message Router), VoiceXML 서버, TTS 서버로 이루어져 있다.

비 미들웨어는 지식서버의 일부인 GPS 서비스 번들과 차량 콘트롤러를 포함하는 동시에 웹서버, 온톨로지와 POI 등의 DB의 집합이다.

단말 클라이언트는 PDA, 차량용 PC, 모의 주행을 위한 에뮬레이터로 구성된다. PDA와 차량용 PC에는 내비게이션 소프트웨어, HTML 브라우저, 음성 I/O 처리 모듈이 탑재되어 VoiceXML, HTML등 마크업 기반의 모달리티 콤포넌트를 처리하며 기능과 역할이 서로 유사하다.

2.2 단말 클라이언트

멀티모달 입력은 PDA와 차량용 PC의 브라우저를 통하여 음성/비음성 이벤트로 처리된다. 음성 입력의 경우, 분산 음성 인식(DSR) 방법을 사용한다. 마이크로부터 입력된 음성 신호는 음성 I/O 처리 모듈에서 특징 데이터를 얻은 다음 결과를 음성 서버로 전송한다. 음성 서버의 ASR 서버는 결과 텍스트에 각종 메타데이터를 더하여 EMMA 형식으로 프레임워크로 전달한다.

비음성인 입력의 경우, HTML 브라우저와 내비게이션 S/W가 처리한다. 네트워크 매니저는 HTML 브라우저의 펜 입력 이벤트를, 내비게이션 S/W의 GPS 신호를 EMMA 포맷으로 변환하여 각각 MMI로 전송한다.

멀티모달 출력은 지식서버와 단말 클라이언트가 담당한다. 지식서버의 GPS 서비스 번들은 RDF에 스타일시트를 추가하여 동적으로 HTML, VoiceXML 스크립트를 생성한 다음 스크립트 저장소에 저장한다. 단말 클라이언트의 HTML 브라우저, 내비게이션 S/W, 음성 I/O 처리 모듈은 HTML과 VoiceXML 스크립트를 구동한다. 음성의 경우는 TTS 서버에서 음성콘텐츠를 인코딩하여 단말로 최종 합성음을 전송한다. 비음성의 경우는 HTML 브라우저를 통해 단말에 출력된다.

차량 모의실험용 에뮬레이터는 내비게이션 로그 파일을 이용하여 모의 주행한다. 에뮬레이터를 이용하면 차를 실제로 운행하지 않고도 멀티모달 입출력 테스트가 가능하다.

2.3 비 미들웨어

비 미들웨어는 텔레매틱스 서비스를 구동하기 위한 GPS 서비스 번들과 차량 제어 장치(Car Controller), 기타 DB로 구성된다. 미들웨어가 특정 서비스에 종

속되지 않는 반면 비 미들웨어와 단말 클라이언트는 텔레매틱스 서비스에 종속되어 필요에 따라 미들웨어와 분리가 가능하도록 설계되었다.

GPS 서비스 번들은 사용자 위치 정보와 커맨드로 POI DB에 접근, 선호(Preference) 온톨로지를 참조하여 최종 검색 결과를 얻는다. 차량 제어 장치는 MMI로부터 주기적으로 GPS 신호를 수신하여 에뮬레이터에 차량 상태 정보를 제공한다.

2.4 미들웨어

미들웨어는 지식서버의 MMI와 리즈너, 음성서버로 이루어져 있다. 단말 클라이언트에서 EMMA 메시지를 받아 차량 제어장치로 GPS 신호를 송출하거나 GPS서비스 번들에 사용자 질의문을 보낸다. MMI는 SCXML을 읽어 멀티모달 시나리오를 진행한다. SCXML은 HTML 브라우저와, VoiceXML 서버로 동기 또는 구동 신호를 보내어 모달리티 콤포넌트를 제어하거나 외부 번들과 통신한다.

III. 멀티모달 인터페이스

3.1 텔레매틱스 시나리오

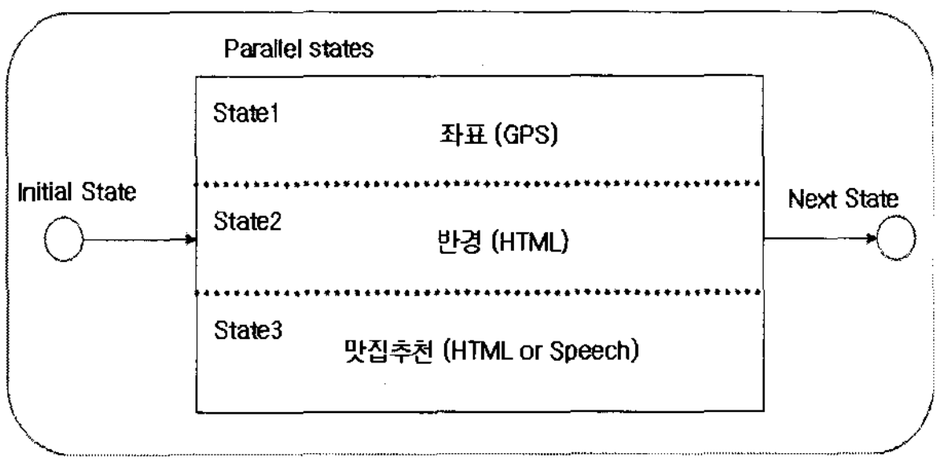
본 논문에서 채택한 테스트 시나리오는 위치기반 맛집추천 서비스이다. 맛집추천 서비스란 사용자가 주행 중에 지도상의 위치 또는 반경을 선택하고 이어서 "맛집추천" 메뉴를 클릭하거나 음성으로 "맛집추천"을 발성하여 커맨드를 완성한다. 서비스 번들은 리즈너로 하여금 POI DB를 참조하여 반경 내의 식당목록을 찾는다. 그리고 사용자의 선호도를 참조하여 결과 리스트를 보여주고 그 중 선택된 식당의 경로로 바꾸어 주는 서비스이다. 리스트를 찾기 위해 서비스 번들에서 필요한 정보는 좌표, 반경 정보, 키워드("맛집추천")이다. 좌표는 주행 중에 5초 간격으로 자동으로 수신하지만 사용자가 직접 지도 위의 특정 좌표를 지정할 수 있다. 이 경우 사용자 이벤트의 우선순위를 높였다.

3.2. 멀티모달 입력

MMI는 GPS, HTML, VoiceXML의 3개 모달리티 입력을 처리하고 시나리오에 따라 순차 또는 복합(보조결합,[5]) 모달리티를 지원한다. 순차 모달리티 입력은 음성과 비 음성 두 가지 모드 중 하나를 선택한다. 따라서 한 번에 하나의 모달리티 입력만을 허용한다. 복합 모달리티 입력은 여러 모드를 조합하여 동시에 여러 모달리티의 입력이 허용된다. 미들웨어는 GPS, HTML, VoiceXML의 3개 모달리티를 지원한다.

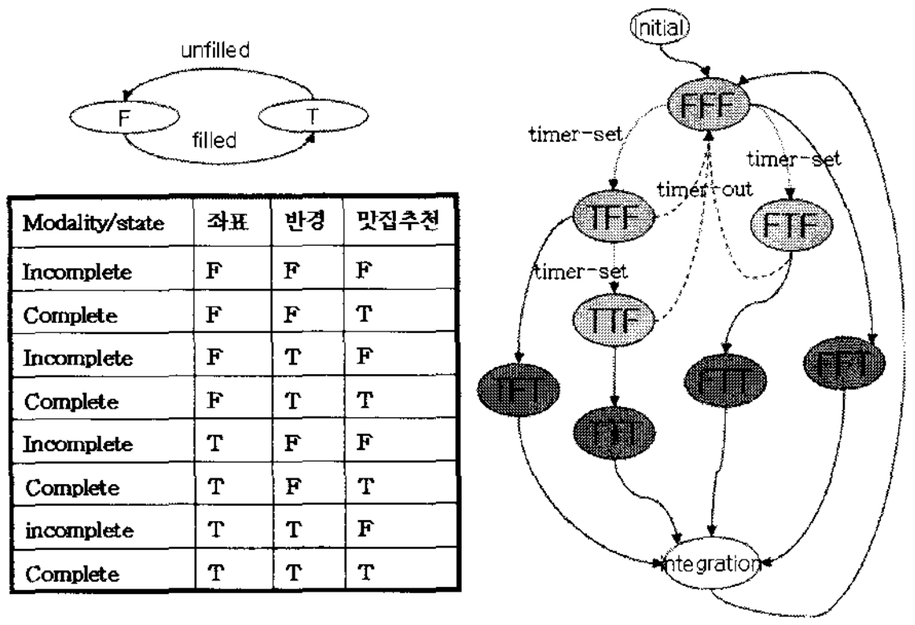
3.3 복합 멀티모달 처리

복합 모달리티 입력은 <그림 2>와 같이 활성화된 병렬(Parallel) 상태에 의하여 이루어진다.



<그림 2> 복합 SCXML 입력 다이어그램

복합 멀티모달은 맛집추천 서비스의 경우 좌표, 반경, 키워드("맛집추천")의 3가지 입력이 조합되어 입력 정보가 완성된다. 좌표와 반경선택을 제외한 맛집추천은 펜 및 음성 입력이 가능하다. 입력 중에 키워드 정보는 복합 멀티모달 입력을 완성하고 서비스 번들을 깨우는 트리거 정보로 이용되므로 생략이 불가능하다. 반면 좌표와 반경선택은 생략 가능한데, 보통 좌표 + 반경 선택 순으로 진행된다. 좌표가 생략되면 현재 차량 위치가, 반경이 생략되면 디폴트 값이 사용된다.



Modality/state	좌표	반경	맛집추천
Incomplete	F	F	F
Complete	F	F	T
Incomplete	F	T	F
Complete	F	T	T
Incomplete	T	F	F
Complete	T	F	T
incomplete	T	T	F
Complete	T	T	T

<그림 3> "맛집추천" 서비스의 입력 과정 모델링

<그림 3>은 맛집추천 서비스에 필요한 3개 입력의 과정을 모델링한 것이다. 각 입력은 값이 채워진 든 상태(T)와 값이 채워지지 않은 빈 상태(F)의 둘 중 하나에 존재한다(왼쪽 테이블 위 다이어그램). 아래 테이블은 각 입력의 상태와 정보의 완성 여부를 나타낸 것이다. 커맨드 입력이 없는 상태에서 좌표 또는 반경이 입력되면 타이머가 작동(timer-set)되어 각 상태는 일시적으로 든 상태에 머문다. 이 때 키워드 입력이 발생하면 통합(integration)상태로 전이되며 MMI는 완성된 입력을 모아 서비스 번들로 넘긴다. 만약 키워드 입력이 없이 시간 초과(timer-out)되면 다시 빈 상태로 전이된다.

오른쪽 다이어그램은 이 과정을 묘사하고 있으며

실제로는 동등한 SCXML 스크립트에 의해 처리된다. 다이어그램은 IBM 멀티모달 변환 플러그인[8]을 활용하여 자동으로 SCXML로 변환된다.

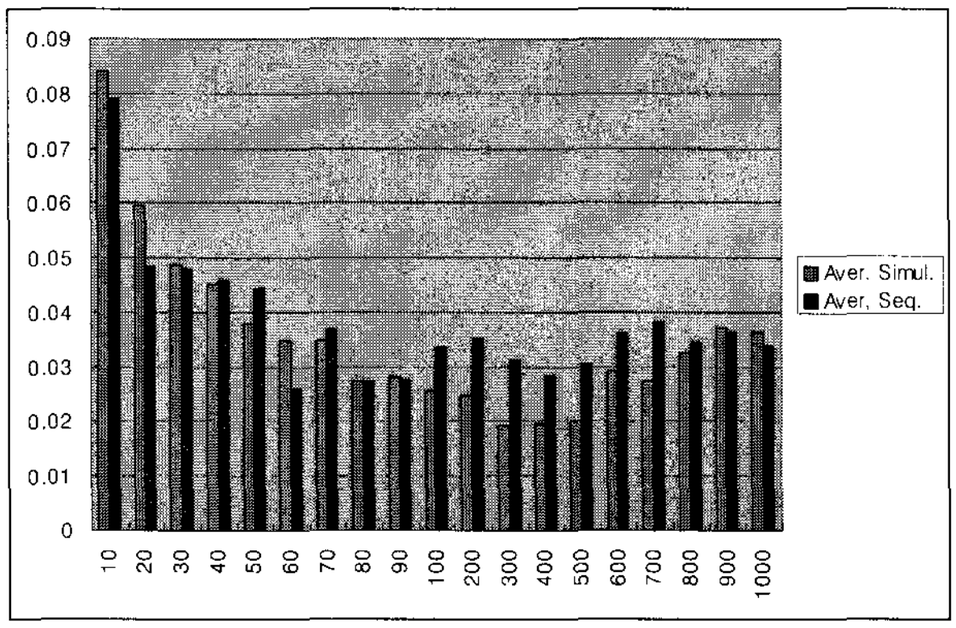
Ⅳ. 실험 및 분석

4.1. 모의실험

다수 사용자에 대한 미들웨어의 수행 능력을 알아보기 위해 모의실험을 수행하였다. 테스트는 맛집추천 시나리오가 기술된 SCXML의 동시(Simultaneous) 테스트와 순차(Sequential) 테스트로서 환경은 Intel Pentium-M 1.8Mhz, 1G Ram이다. 클라이언트는 모달리티 콤포넌트 없이 자체적으로 멀티모달 이벤트를 발생시켰다.

동시 테스트는 하나의 SCXML 스크립트를 읽어 미리 n개의 오브젝트 모델을 생성한다. 이어 *n*개에 대한 사용자 입력 (좌표 + 반경 + 맛집추천) 이벤트를 동시에 발생시켜 평균 처리시간을 체크하였다.

순차 테스트는 대상 SCXML 스크립트를 읽어 하나의 오브젝트 모델을 생성한다. 그리고 위와 동일한 이벤트를 발생시켜 인스턴스 개수만큼 순차적으로 반복하여 평균 처리 시간을 체크하였다.



<그림 4> 동시, 순차 테스트의 평균 처리 시간

<그림 4>의 동시, 순차 테스트 결과 초반에는 불규칙하게 진행되다 횟수가 지날수록 시간이 안정되었다. 약 800회까지 순차 테스트에서 조금 더 시간이 걸렸고 그 이후부터는 동시 테스트가 더 걸렸다.

횟수가 작을 때는 오브젝트 모델을 매번 새로 만드는 순차테스트가 불리하지만 횟수 800 이상에서 그래프는 역전되었다. 동시 테스트는 오브젝트 모델을 미리 만들어 둔다는 장점에도 불구하고 동시에 처리해야 하는 부담이 크게 작용한 것으로 여겨진다. 하지만 그 차이는 예상보다 미미했다. 실제 어플리케이션은 네트워크, I/O 프로세서에서도 영향을 받으므로 전체 소요 시간은 이보다 넉넉히 잡아야 할 것으로 보인다.

Ⅴ. 결론

본 논문에서는 미들웨어를 위치기반 텔레매틱스 시나리오에 적용하였다. MMI는 SCXML 스크립트에 따라 멀티모달 시나리오를 수행한다. SCXML은 모델링/변환 툴킷[8]에 의해 다이어그램으로부터 자동으로 쉽게 얻을 수 있다. SCXML은 사용자의 HTML, VoiceXML 마크업 뿐 아니라 시스템 혹은 사용자 GPS 신호를 처리한다. 모의실험 결과, 미들웨어는 다수 사용자 대상으로 텔레매틱스용 멀티모달 인터페이스를 운용할 수 있음을 보여주었다.

참고문헌

[1] James A. Larson, "Commonsense Guidelines for Developing Multimodal User Interfaces", *World Wide Web Consortium(W3C)*, April 2003.

[2] James A. Larson, Raman, Dave Raggett, "W3C Multimodal Interaction Framework" *World Wide Web Consortium(W3C)*, May 2003; http://www.w3c.org/TR/mmi-framework.

[3] 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/.

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

[5] 박성수, 박성찬, 안세열, 구명완, 김원우, "모바일 환경을 위한 멀티모달 미들웨어 설계 및 구현", *말소리* 제 60호, pp125-144, 2006.

[6] 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/

[7] Seyeol Ahn, Sung-Chan Park, Seong-Soo Park, Myoung-Wan Koo, Yeong-Joon Jeong and Myung-Sook Kim, "Multimodal Context-Aware middleware Interfaces for mobile applications" *제 23회 음성통신 및 신호처리 학술대회* (KSCSP2006 Vol23, No. 1), pp.209-212, 2006

[8] "IBM Modeling and integration Tools for State Chart XML 1.0", *IBM*, http://www.alphaworks.ibm.com/tech/scxml, 2007