

멀티모달 상황인지 서비스

정영준^{*}, 박성수, 안세열
KT 미래기술연구소 HCI연구담당

Multimodal Context-aware Service

Jeong, Yeong Joon^{*}, Seong-Soo Park, Se-Yeol Ahn

Human-Centric Interface Research Department, Advanced Technology Laboratory, KT.

E-mail : yjjung@kt.co.kr, soopark@kt.co.kr, syahn@kt.co.kr

요 약

다양한 서비스와의 서비스간 융합이 활발히 이루어지는 유비쿼터스 서비스 환경에서 사용자에게 보다 편리하게 맞춤 서비스를 제공하기 위해서는 다양한 입출력 수단을 통해 상황에 따라 적절한 서비스를 제공할 수 있는 기술이 필요하다. 멀티모달 상황인지 기술은 언제, 어디서나 사용자에게 최적의 서비스를 제공할 수 있도록 다양한 상황정보를 인지하여 적절한 입출력 수단을 사용할 수 있는 UI를 제공함으로써 사용자별 맞춤형 서비스가 가능하게 하는 기술이다.

본 고에서는 멀티모달 및 상황인지 기술에 대한 개요와 KT에서 개발하고 있는 멀티모달 상황인지 플랫폼 및 이를 기반으로 한 홈네트워크 서비스를 소개하고자 한다.

1. 서론

유비쿼터스 환경의 특징은 다양한 서비스의 융합과 이를 위한 단말기의 소형, 복잡화 증대, 이에 따라 사용자도 보다 편리하게 맞춤 서비스를 받고자 하는 점을 들 수 있다.

멀티모달 상황인지 기술은 언제, 어디서나 사용자에게 최적의 서비스를 제공할 수 있도록 사용자 프로파일, 단말 규격, 사용환경 등의 상황정보를 인지하여 적절한 입출력 수단을 사용할 수 있는 UI를 제공함으로써 사용자별 맞춤형 서비스가 가능하게 하는 기술이다.

본 고에서는 멀티모달 인터페이스 기술과 상황인지 기술에 대한 개요와 표준화/시장 동향에 대

해 소개하고, KT에서 개발하고 있는 멀티모달 상황인지 플랫폼 및 서비스를 소개하고자 한다.

멀티모달 인터페이스 기술은 기존의 독립적인 사용자 입출력 방법을 동시에 사용할 수 있도록 하여 상황에 따라 사용자가 선택할 수 있도록 하는 기술이며, 현재까지는 키보드/마우스와 음성을 결합하는 형태가 가장 많이 연구/표준화 되고 있다. 본고에서는 W3C (World Wide Web Consortium)에서 추진하고 있는 기술 표준에 대해 2장에서 소개한다.

상황인지 기술은 유비쿼터스 사회의 실현이라는 화두와 함께 새롭게 다양한 분야에서 연구/적용되고 있고, 3장에서는 그 기본 개념과 상황인지 서비스를 위한 미들웨어에 필요한 기술에 대해 주로

언급하였다.

4장에서는 두 가지 요소기술을 결합하여 멀티모달 상황인지 플랫폼 및 서비스를 연구하고 있는 KT의 개발 사례 및 방향에 대해 소개한다.

2. 본론

2.1. 멀티모달 인터페이스 기술 개요

기존의 문자, 음성, 영상 등의 유니모달(unimodal) 인터페이스 수단은 더욱 다양해지는 서비스와 점점 소형화, 복잡화 되고 있는 단말기에서 사용자가 충분히 편리하고 쉽게 쓰기에는 한계가 있다. 멀티모달 인터페이스는 사용자 주변의 다양한 입출력 수단을 활용하여 상황에 따라 가장 편리한 수단을 쓸 수 있게 한다. 또한, 여러 모달리티를 동시에 처리함으로써 사용자 의도를 더욱 간결하고 명확하게 전달할 수 있도록 한다. 각각의 모달리티가 갖는 부분 정보를 통합하여 불명확도를 줄이고, 결과도 다양한 형태로 제공함으로써 전달율을 높이는 것이다. 더 나아가 사용자 단말뿐만 아니라 주변의 기기도 이용할 수 있도록 해주며, 기기의 능력에 맞게 다양한 형태로 정보를 가공하여 제공할 수 있도록 하는 기술이다.

현재의 멀티모달 인터페이스 기술은 세계적으로도 시제품이 나오는 단계이며 표준화가 진행중인 기술로, 크게 3가지 표준이 주류를 이루고 있다.

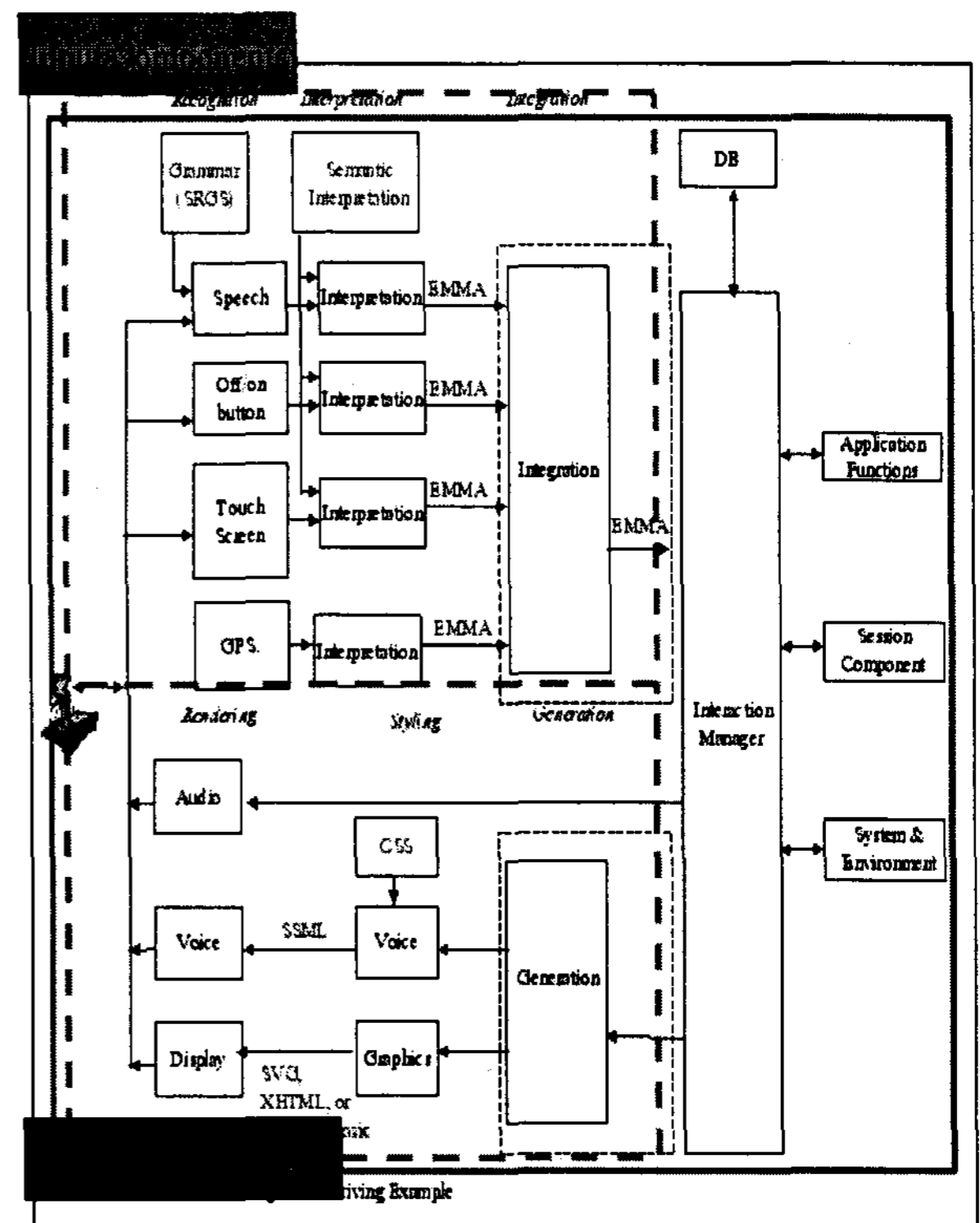
첫째로 W3C의 멀티모달 인터랙션(Multi-Modal Interaction) 워킹 그룹에서 웹기반의 다양한 모달리티를 수용할 수 있는 royalty free인 표준이 있고, 둘째는 IBM이 주도하고 있는 X+V 기술로 HTML과 음성(VoiceXML)을 동시에 처리,, 휴대폰에서의 멀티모달 인터페이스로 많이 연구되고 있으며, 마지막으로 마이크로소프트에서 추진중인 SALT(Speech Application Language Tags)로 .NET 기반으로 기존의 XHTML에 음성처리 기술을 접목하는 표준이 있다.

아직까지는 기존 인터페이스에 음성기술을 접목하는 것이 가장 많이 연구되고 있으며, 제스처나

얼굴 등의 생체인식에서도 연구가 되고 있다.

2.1.1 Multimodal Interaction Framework [1]

W3C에서의 멀티모달 인터페이스는 문자, 음성, 잉크 인터페이스 및 보편적인 그래픽 인터페이스를 모두 포함하고 있으며, 앞으로 새로이 발전하는 인터페이스 기술들을 추가할 수 있도록 다양한 모달리티 인터페이스간의 상호관계 (Interaction)에 따라 다음과 같은 Multimodal Interaction Framework을 정의 하고 있다.



(그림 1) W3C의 멀티모달 인터랙션 프레임워크

가. 입력 요소 (Input Components)

사용자의 입력을 받고, 이를 인지하는 인식 모듈(Recognition), 인식된 결과의 의미(semantic) 정보를 판단하는 해석모듈 (Interpretation)을 거쳐서 사용자의 명령 및 모달리티의 정보를 포함하는 데이터(EMMA:Extensible Multi-Modal Annotation)를 생성한다. 여러 입력 요소의 결과는 통합모듈(Integration)을 거쳐서 하나의 데이터로 통합된다.

나. 출력 요소 (Output Components)

정보를 어떤 모달리티를 활용하여 출력할 것 인지를 결정하는 생성모듈, 스타일 정보들을 기술하는 스타일 모듈, 모달리티의 특성에 맞게 음성, 그래픽 등을 출력하는 렌더링 모듈로 구성된다.

다. 인터랙션 매니저 (Interaction Manager)

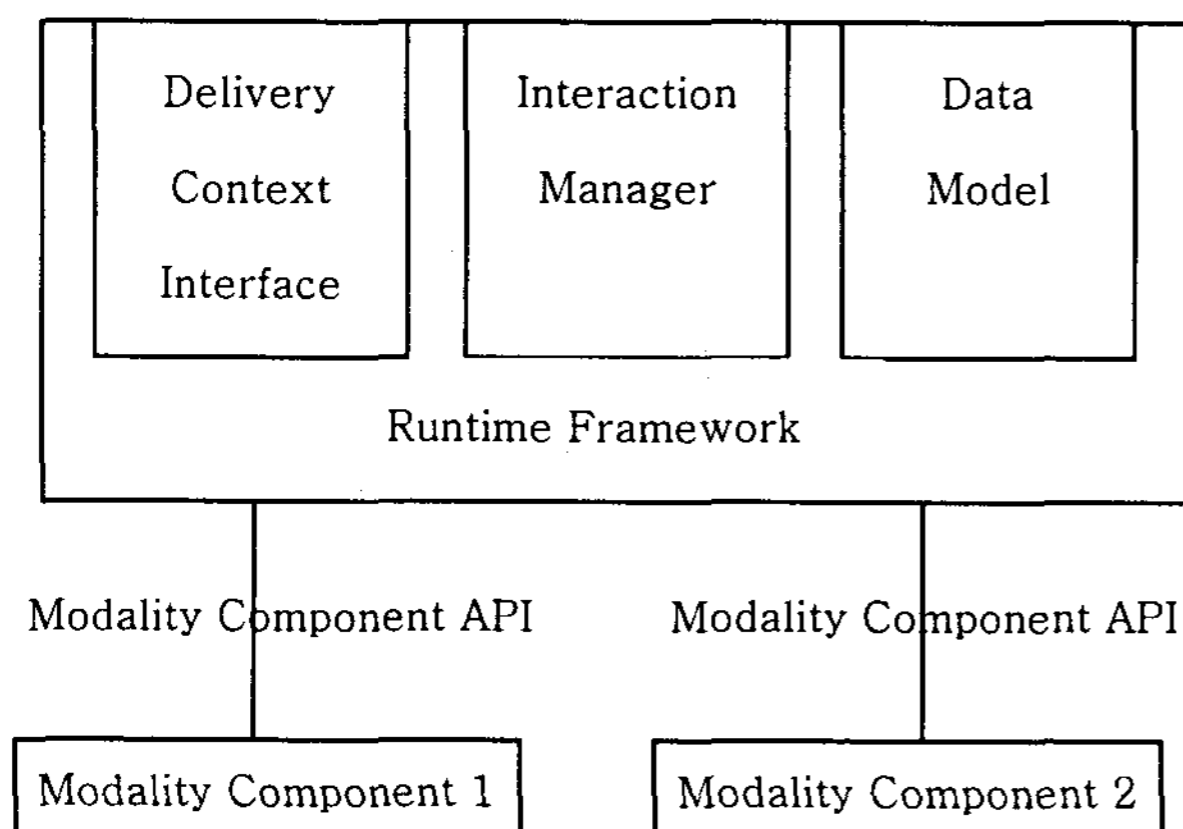
입력 요소로부터의 EMMA 데이터를 이용하여 실제 응용 서비스를 실행 한 후, 그 결과를 출력 요소에 제공한다. 다양한 모달리티의 인터랙션을 통합하여 모달리티 간의 동기화를 이룬다. 상태 천이 기계(state machine)로 구성되며, 현재 상태 및 이벤트 정보를 바탕으로 응용 서비스 실행 및 사용자와의 상호작용(Interaction)을 관리한다.

라. 세션요소(Session Component), 시스템 및 환경요소 (System & Environment Component)

세션 요소에서는 다양한 디바이스가 사용되더라도 세션이 끊어지지 않도록(seamless) 멀티모달 응용서비스와의 지속적인 연결 및 다양한 단말기 출력을 위한 싱크 기능이 있다. 시스템 환경 요소는 디바이스 선정 및 사용자 선호에 따라 자동으로 변화하도록 시스템 환경을 표현해 준다.

2.1.2 W3C의 멀티모달 구조 [2]

(그림 2)와 같이 W3C의 멀티모달 구조는 크게 런타임 프레임워크 (Runtime Framework) 와 모달리티 컴포넌트(Modality Components)로 구성되며 주요 구성요소는 다음과 같다.



(그림 2) W3C 멀티모달 구조

가. Runtime Framework (RF)

- Host language, Component software 초기화
- Lifecycle event 생성, Event loop
- Coordination between asynchronous message and synchronous message.

나. Dynamic Property Framework (또는 DCI : Delivery Context Interface)

- 사용 단말의 성능, 설정값, 사용자 선호도 그리고 환경 상태정보 등의 특성에 대한 계층적 구조에 동적으로 접근하도록 구성

다. Interaction Manager (IM)

- coordinates data and manages execution flow from various I/O modalities.
- Representation language : SCXML (StateChart XML) – Harel State Chart
- 역할(Message subscribe/publish) : modality component에서 발생하는 메시지 처리

라. Data Model

- Modality component 간의 shared data model
- 예를 들어, GPS application에서 현재 위치, destination 및 traffic information

마. Modality Components (MCs)

- responsible for controlling various input and output modalities on the device.
- 각 modality에 매치되는 namespace를 가짐

2.2. 상황인지 기술

유비쿼터스 환경 또는 서비스 구현에 있어 또 하나의 핵심 요소는 상황인지 기술이다.

상황(Context)이란 특정 시점에 특정 Entity의 상황(또는 상태)을 설명해 줄 수 있는 모든 정보의 집합으로 Entity는 사람, 공간, 서비스와 관련된 모든 객체를 의미하며, 특정한 사람, 장소, 시간, 이벤트에 따른 다양한 상황에 맞는 적합한 서비스를 제공하기 위하여 그와 관련된 Device로부터 상황정보를 얻고 활용하는 기술이 상황인지(Context-

Awareness) 기술이다.

초기에는 주로 상황정보와 센싱 기술이 조합된 특정 플랫폼에서의 개별적인 프로토타입 서비스가 선보였으나, 2000년 이후 현재까지는 응용서비스 개발에 필요한 공통 기능을 미들웨어 형태로 제공하는 것이 연구되고 있다.

먼저 상황인지 서비스를 위한 요소기술을 살펴보면 다음과 같은 것이 있다. [3]

□ 상황정보 센싱 기술 (Context Sensing)

상황정보는 사용자 단말기의 UI를 통해 직접 수집하거나 센서 또는 센서네트워크 통신을 통해 수집할 수 있다. 초소형 센서노드를 통한 객체의 식별이나 GPS를 활용한 위치 정보 등을 얻을 수 있는 다양한 방법이 있으며, 주기적인 폴링 또는 대량의 정보를 미리 처리하여 특정 조건을 만족할 때만 보내는 형태도 가능하다.

□ 상황정보 모델링 (Context Modeling) 및 융합(Context Fusion) 기술

수집된 정보를 시스템에서 처리하고 다른 시스템과의 교환을 통해 공유하기 위해서는 표준화된 형태로 정보를 표현하고 모델링하고 관리할 필요가 있다. 이를 통해 상위계층의 상황정보로의 유도가 가능하게 된다.

□ 상황정보 추론기술 (Context Reasoning)

다양한 상황 (정적, 동적) 정보를 기반으로 룰 또는 인공지능 등을 이용하여 지능적으로 상황을 판단하고 상황간의 충돌을 해소하는 기술

이러한 핵심기술을 기반으로 상황인지 서비스를 제공하기 위해서는

- 상황정보를 표준화된 형태로 상호교환 (Context exchange)할 수 있는 기술
- 서비스에 적합한 시스템 구조 (System Infrastructure) 기술
- 상황인지 서비스를 묘사하고 다양한 서비스를 발견할 수 있는 기술

이 필요하다.

상기 요소기술들의 각각이 많은 연구가 필요한 분야이고, 여기에 보안 및 Privacy가 더욱 중요한 문제로 대두되고 있지만, 상황인지 기술은 현재 여러 분야에서 새롭게 적용이 되고 있으며, 다양한 상황인지 서비스를 개발하고 서비스간 융합을 위해서는 상황인지 미들웨어가 매우 중요하다. [4]

Middleware 가 필요한 이유로는

□ Context-aware agent를 쉽게 개발할 수 있게 한다.

- 서로 다른 센서로부터 context 를 얻는 방법이나 context 추론 및 학습 메커니즘을 middleware가 제공한다.

□ Seamless한 서비스를 가능하게 한다.

- agent들 간의 통신에 middleware가 context 를 교환할 수 있게 해줌으로써 서로 다른 agent 간의 semantic gap 이 없게 해준다.

등이 있으며, 미들웨어가 갖추어야 할 기본 기능들은 다음과 같다.

□ Functional Requirements

- 서로 다른 센서로부터 context를 수집하고 이를 다른 agent에 전달
- 수집된 context로부터 상위단계의 상황을 추론
- agent가 고유의 추론 및 학습 메커니즘을 사용 가능
- 특정 context에서 특정 behavior를 지정
- 서로 다른 agent 간에 언어와 의미 교환

□ Architectural Requirements

- Hardware, OS, Programming Language에 독립적이어야 함 (Extensibility)
- 분산된 context-aware agent 들 사이의 상호작용을 기반으로 복잡한 시스템을 만들 수 있어야 한다. (Interoperability)

다음은 상황인지 서비스 인프라 관련 대표적인 연구 사례를 소개한다.

□ **Scalet Context-aware Infrastructure [5]**

- 주관기관 : Illinois Institute of Technology
- 연구분야 : 이질적 플랫폼간 상황정보 교환
- 내용 : 센서 노드와 휴대용 PC간의 상황 정보는 Scalet 인프라의 SOAP over HTTP를 통해 전송되고, 가용한 웹 서비스는 WSDL을 이용하여 정의되고 알려진다

□ **Service Globe [6]**

- 주관기관 : TU 뮌헨 (2002~)
- 연구분야 : 상황인식 웹서비스 플랫폼
- 내용 : 다양하고 이질적인 단말 능력과 고객의 위치 등과 같은 상황정보를 고려하여 더 나은 상황인지 웹서비스 제공

□ **Gaia (Context Infrastructure) [7]**

- 주관기관 : University of Illinois
- 연구분야 : smart space 라는 유비쿼터스 컴퓨팅 환경을 위한 infrastructure
- 내용 : 다양한 형태의 task를 수행하기 위해 여러 종류의 agent로 구성

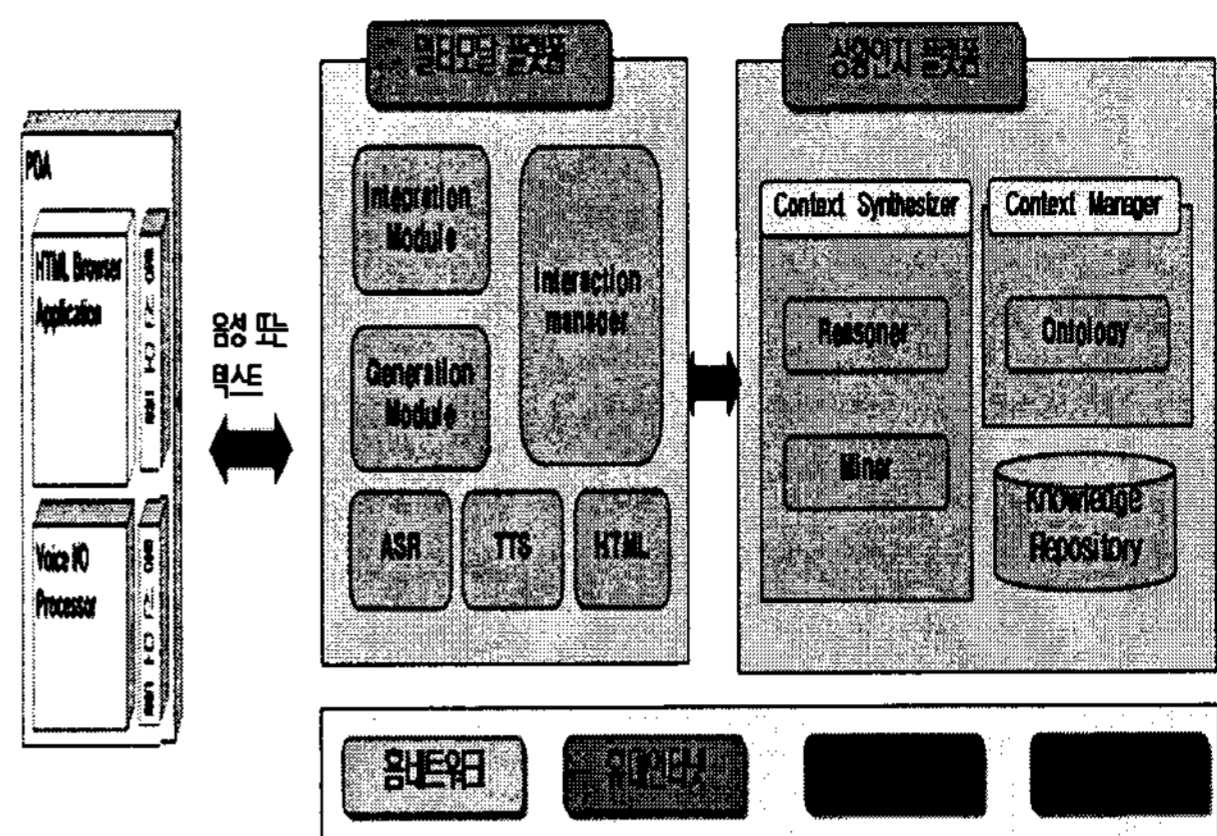
□ **SOCAM (Service Oriented Context-Aware Middleware) [8]**

- 주관기관 : National Univ. of Singapore
- 연구분야 : 상황인식 모바일 서비스를 위한 미들웨어
- 내용 : 미들웨어내에서 상황정보 모델링을 위해 OWL 을 사용

2.3. 멀티모달 상황인지 플랫폼

이 장에서는 현재 KT에서 연구/개발되고 있는 멀티모달 상황인지 플랫폼을 소개하고자 한다. (그림 3)과 같이 기본적으로 2장 및 3장에서 언급한 멀티모달 인터페이스 기술과 상황인지 기술을 결합한 것으로, 유비쿼터스 사회에 맞는 다양한 서비스를 쉽게 개발하고 제공할 수 있는 미들웨어를

목적으로 한다.



(그림 3) 멀티모달 상황인지 플랫폼 구성

2.3.1 멀티모달 단말기

멀티모달 단말기는 스타일러스 펜과 음성을 통한 멀티모달 입력을 받아들이는 역할을 수행한다. 즉, 사용자와 그래픽 및 펜 인터페이스를 수행하는 HTML 브라우저가 내장되어 있다. 그리고, 음성 입출력을 처리하기 위한 Voice I/O 프로세서가 있어, 입력된 음성을 DSR(Distributed Speech Recognizer) 표준에 따라 특징을 추출한 뒤, 무선 랜을 통해 DSR 서버로 전달하며, 서버에서는 입력된 특징 정보에 따라 인식을 수행한다[9].

2.3.2 멀티모달 상황인지 플랫폼

(그림 3)에서 멀티모달 플랫폼은 기본적으로 W3C에서 제정되고 있는 멀티모달 프레임워크의 기본 구조를 따른다.

애플리케이션 제작을 위한 Container Language는 SCXML을 사용하며, 음성응용 서비스를 위한 언어는 VoiceXML을 사용한다.

멀티모달 표현언어인 EMMA를 사용하며, 입력 결과가 N-best 형태로 표현된다.

Runtime Framework은 외부 서버와의 연동을 위해 RDF, OWL 데이터를 처리한다.

상황인지 플랫폼은 여러 상황정보를 처리/관리하는 마이너와 상황에 따라 룰 또는 추론에 의해

지능적으로 결과를 판단하는 리즈너를 핵심요소로 갖추고 있으며, 서비스 도메인에 따른 온톨로지를 구성하여 인터넷상의 다양한 지식환경과 연동할 수 있도록 구성하였다.

멀티모달 인터페이스로부터의 사용자의 서비스 환경 정보뿐만 아니라 인터넷상의 지식 정보를 수집/가공하여 다양한 상황에 따라 사용자 맞춤형 지능 서비스를 제공할 수 있도록 하였다.

다음은 주요 구성 요소에 대한 기능 개요이다.

□ 컨텍스트 매니저 (Context Manager)

- Ontology 및 메타데이터 관리
- OWL(RDF) 문서의 parsing과 synthesizing

□ Context Synthesizer

- Reasoner : Context model의 class 및 individual간 관계성을 이용하여 다양한 상황을 추론(context model 및 rule기반)하고 현재 상황에 맞는 지식 서비스 결정
- Miner : 입력된 정보를 언어 분석하고, 수집된 문서를 분류/군집/요약

□ Knowledge Repository

- Home Network Ontology (Context Model) : 사용자, 각종 Device 정보, 시간 정보 등을 저장 관리하고 각 기능별로 필요한 관계성을 OWL로 기술한 Knowledge
- Rule Set : Reasoner의 Rule-based Engine에서 필요로 하는 Rule들을 저장한 DB

2.4. 홈네트워크 기반의 멀티모달 상황인지 서비스

(그림 4)는 멀티모달 상황인지 기반의 홈네트워크 서비스의 구성도이다.

멀티모달 상황인지 플랫폼은 홈 외부(IDC 등)의 홈서버에 위치하며, 홈내의 홈게이트웨이와 인터넷으로 연결되어 있는 구조이다.

사용자는 멀티모달 인터페이스가 탑재된 PDA의 무선랜을 통해 홈서버에 접속하여 상황인지 기반의 홈네트워크 서비스를 이용할 수 있다. 기본적으로 런타임 프레임워크와 상황인지 모듈이 탑재

된 멀티모달 상황인지 서버, 음성 서버, 서비스 스크립트가 있는 시나리오 서버로 구성된다.

다음은 상기 플랫폼을 기반으로 구축한 멀티모달 상황인지 서비스의 예이다.

□ 멀티모달 지도 검색 서비스

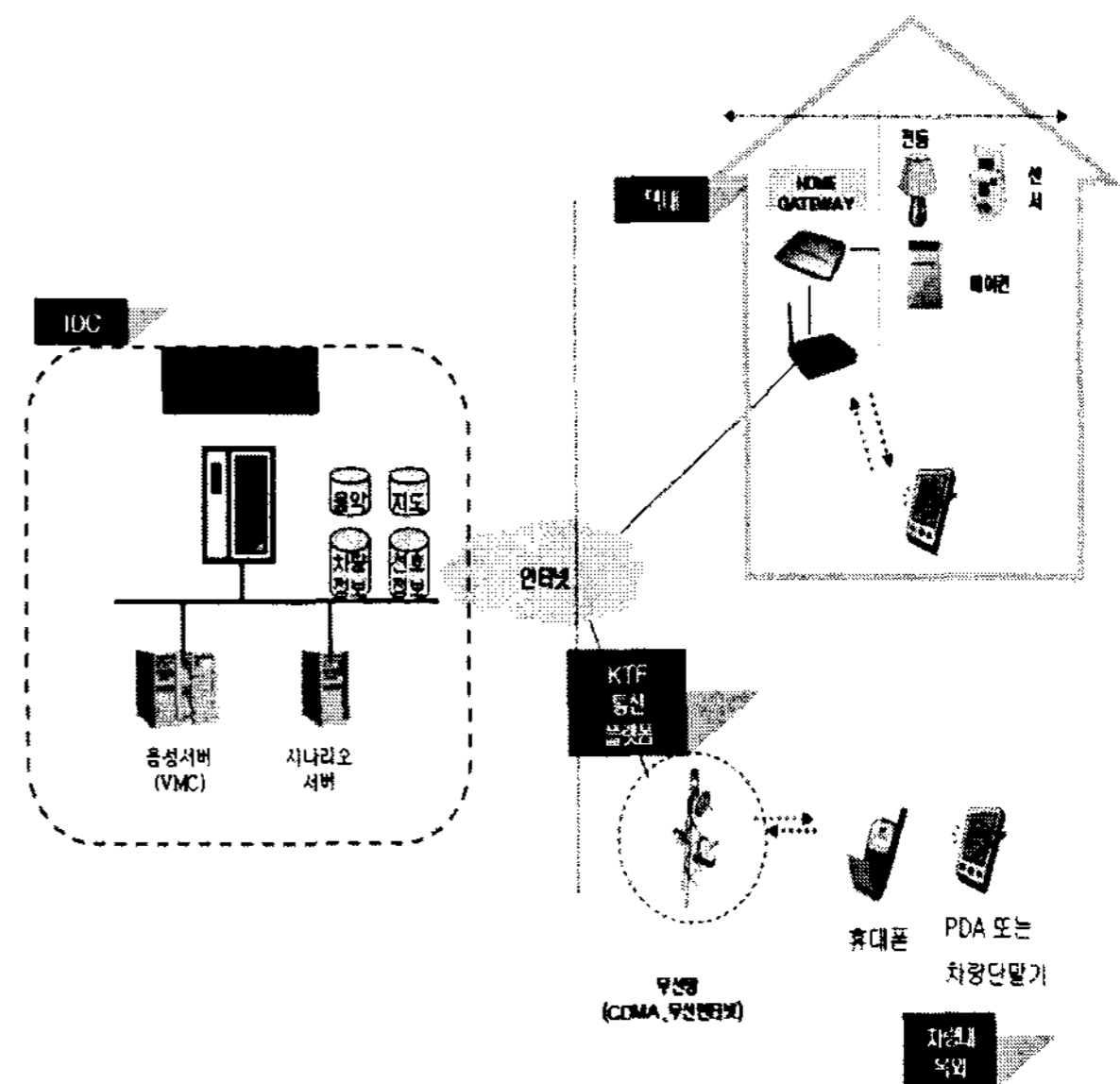
- 음성과 포인팅, 클릭 등의 여러 모달리티를 동시 적용, 지도기반의 주변 정보 검색 서비스

□ 상황인지 기반 음악추천 서비스

- - 사용자 선호도 정보(장르, 프로파일 등), 무드, 날씨, 시간 등의 상황에 따른 추천
- - 다자(그룹)의 공통 선호, 비선호 장르, 공통 무드 및 상황 정보에 따른 추천

□ 멀티모달 홈네트워크 기기제어 서비스

- PDA상에서 터치스크린 및 음성인식을 통한 홈내 기기제어 서비스



(그림 4) 멀티모달 상황인지 서비스 구성도

향후에는 서비스 도메인을 확장하여 텔레매틱스와 홈네트워크 서비스를 연동하여 다양한 서비스 간을 Seamless하게 이용할 수 있는 구조로 확대할 예정이다. PDA뿐만 아니라 CDMA나 와이브로 단말기를 통해 이동 중에도 지속적인 서비스를 받을 수 있도록 하며, 모달리티도 센서(GPS 등)를 추가

하여 기존 모달리티와 동시처리 가능하도록 하고, 도메인 온톨로지의 구축 및 연동이 가능하도록 할 예정이다.

3. 결론

본 고에서는 멀티모달 인터페이스 기술과 상황 인지 기술에 대한 소개 및 기술 동향에 대해 기술 하였다. 그리고 두 기술이 통합된 멀티모달 상황 인지 플랫폼과 이를 기반으로 홈네트워크에서의 멀티모달 인터페이스를 이용한 지능형 상황인지 서비스의 구축 예를 소개하였다.

이 두 가지 기술은 유비쿼터스 환경의 다양하고 복잡/융합된 서비스의 개발 및 제공에 필수적인 요소이며, 이러한 멀티모달 인터페이스 및 상황인지 기술은 국제적으로 표준화가 활발히 진행되고 있고 향후 국내에서도 지속적인 활동과 기술개발이 필요한 분야이다.

KT에서 개발하고 있는 멀티모달 상황인지 플랫폼은 두 기술을 융합한 것으로 사용자에게 보다 편리하고 다양한 서비스를 상황에 따라 최적으로 제공할 수 있는 기반 기술이 될 것이다. 언제(Anytime), 어떤 기기(Any device)라도 이용할 수 있는 멀티모달 기술과 어떤 상황(Anywhere)에서도 누구에게나(Anyone) 맞춤형 서비스를 제공할 수 있는 멀티모달 상황인지 기술은 유비쿼터스 서비스 실현을 위한 한발 앞선 구체적인 기술이 될 것이다. 현재는 제한된 환경 (PDA 단말기, 무선랜, 홈네트워크 도메인)에서의 서비스에 국한되어 있지만 향후 텔레매틱스, 와이브로, 로봇릭스 등 여러 도메인 환경에서 다양한 무선망 및 단말기를 통해 서비스를 이용할 수 있는 플랫폼으로 진화해 갈 예정이다. 또한, 모달리티도 다양화하여 GPS 등의 센서기반 모달리티의 적용과 오감기반의 인터페이스로 더욱 인간 친화적인 인터페이스를 개발 할 예정이다.

[참고문헌]

- [1] W3C, MultiModal Interaction Framework, <http://www.w3.org/TR/mmi-framework/>, W3C Note, May, 2003.
- [2] W3C, Multimodal Architecture and Interface, <http://www.w3.org/TR/2006/WD-mmi-arch-20060414>, W3C Working Draft 14 April, 2006.
- [3] 김재호, 신경철, “상황인식 서비스 기술 연구 동향”, ITFIND, 주간기술동향 1178호, 2004.12.
- [4] Anand Ranganathan, Roy H. Campbell, “A Middleware for Context-Aware Agents in Ubiquitous Computing Environments,”
- [5] Abhay Daftari, Nehal Mehta, Shubhanan Bakre and Xian-He Sun, “On the Design Framework of Context Aware Embedded Systems,” Monterey Workshop on Software Engineering for Embedded Systems: From Requirements to Implementation, 2003.
- [6] The ServiceGlobe Project, <http://www.db.fmi.uni-passau.de/projects/sg/>
- [7] Gregory Biegel and Vinny Cahill, “A Framework for Developing Mobile, Context-aware Applications,” IEEE International Conference on Pervasive Computing and Communications (PerCom), 2004.
- [8] T. Gu, H.K. Pung and D.Q. Zhang, “A Middleware for Building Context-Aware Mobile Services,” In Proceedings of IEEE Vehicular Technology Conference(VTC), 2004
- [9] 박성준, 한기수, 구명완, “음성인식 WiFi 전화 시스템 개발,” 제22회 음성통신 및 신호처리 학술대회 논문집, pp. 89-92, 2005.