3 차원 오브젝트 직접조작을 위한 데이터 글러브 기반의 멀티모달 인터페이스 설계

임미정[1], 박 범[2]
아주대학교 미디어학과[1]
아주대학교 산업공학과 [2]
{sanctus50[1],ppark[2]}@ajou.ac.kr

Design of dataglove based multimodal interface for 3D object manipulation in virtual environment

Mi Jung Lim[1], Peom Park[2]
Ajou University[1 2]

요 약

멀티모달 인터페이스는 인간의 제스처, 시선, 손의 움직임, 행동의 패턴, 음성, 물리적인 위치 등 인간의 자연스러운 행동들에 대한 정보를 해석하고 부호화하는 인지기반 기술이다. 본 논문에서는 제스처와 음성, 터치를 이용한 3D 오브젝트 기반의 멀티모달 인터페이스를 설계, 구현한다. 서비스 도메인은 스마트 홈이며 사용자는 3D 오브젝트 직접조작을 통해 원격으로 가정의 오브젝트들을 모니터링하고 제어할 수 있다. 멀티모달 인터랙션 입출력 과정에서는 여러 개의 모달리티를 병렬적으로 인지하고 처리해야 하기 때문에 입출력 과정에서 각 모달리티의 조합과 부호화 방법, 입출력 형식 등이 문제시된다. 본 연구에서는 모달리티들의 특징과 인간의 인지구조 분석을 바탕으로 제스처, 음성, 터치 모달리티 간의 입력조합방식을 제시하고 멀티모달을 이용한 효율적인 3D Object 인터랙션 프로토타입을 설계한다.

Keyword : Multimodal interface, Gesture based interface, 3D Object manipulation

1. 서론

멀티모달 인터페이스는 인간의 제스처, 시선, 손의 움직임, 행동의 패턴, 음성, 물리적인 위치 등 인간의 자연스러운 행동들에 대한 정보를 해석하고 부호화하는 인지기반 기술이다. 멀티모달 인터페이스는 인간과 컴퓨터 인터랙션 과정에서 동시 여러 모달리티의 입출력을 허용하며 다수의 모달리티의 조합과 입력 신호 통합해석 등을 통해 상호 의사교환을 한다. 모달리티(Modality)란 인터랙션 과정에서 사용되는 커뮤니케이션 채널을 의미한다. 현재 시스템에서 휴먼-컴퓨터 인터랙션은 한 번에 한 가지 사용자 입력을 허용하기 때문에 하나의 오브젝트가 활성화되면 다른 오브젝트들은 모두 비활성화되어 두 개 이상의 오브젝트의 동시적 입력이 불가능하다. 그러나 멀티모달 인터페이스에서는 청각이나 촉각 등의 모달리티 입력은 동시 입력이 가능하여 여러 프로세스의 동시적 제어가 가능하다. 본 연구에서는 멀티모달 병렬입력을 효율적으로 설계하기 위해 W3C(월드와이드웹컨소시엄), ETSI(유럽표준화기구)에서 발표한 멀티모달 시스템 프레임워크, 요구사항, 관련 기술에 대한 자료를 분석하고 여러 해외논문과 저널에서 발표된 모달리티들 특징들을 수집하였다. 그리고 수집된 자료를 바탕으로 Speech 와 Touch, Gesture 를 이용한 멀티모달 인터랙션 과정에서 필요한 모달리티 입력방법과 동기화 방법, 인터랙션 메커니즘 등에 대해 제안하였다.

2. 연구배경

멀티모달 인터페이스는 MIT Media Lab.과 Microsoft, HP, Intel 등에서 연구 진행 중이며 음성

인식, 표정인식, Haptic Interface, Body Tracking, DataGlove 기반 제스처 인식 기술 등을 중심으로 연구가 진행되고 있다.

기존 인터페이스는 한 번에 한 가지 입력만을 허용하며 일단 하나의 오브젝트가 선택되거나 활성화되면 다른 오브젝트들은 모두 비활성화되어 입력이 불가능하기 때문에 사용자의 작업 효율성에 한계가 있었다. 그러나 멀티모달 인터페이스에서는 청각과 촉각 등의 모달리티 입력들을 동시에 받아들일 수 있으며 모달리티 정보들을 분석하여 여러 개의 입출력 형태로 통합하여 이를 어플리케이션 프로세스에 병렬적으로 적용 가능하다. 멀티모달 인터페이스는 기존 직렬 입력 방식과 달리 동시에 여러 다차원적인 입력 채널을 허용하기 때문에 기존 인터페이스와 구현상에서 본질적으로 차이가 난다.

그러나 아직 멀티모달리티를 이용한 병렬입출력 방법에 대해서는 아직 많은 연구가 이루어지지 않은 상태이다. MIT Media Lab., IBM, HP, Microsoft, Intel 과 같은 연구소와 기업체에서는 주로 신호처리와 인공지능, 사용자 서비스 시나리오 등에 대한 연구가 진행되고 있는데 은닉 마코브 모델(Hidden Markov Model), 칼만 필터(Kalman Filter) 등을 이용한 패턴분석과 Bayesian Network, Neural Network, Machine Learning 을 이용한 인공지능, RF 센서나 센서 네트워크 기반의 휴먼 센싱 인터페이스 등을 중심으로 연구가 진행되고 있다. W3C, ETSI, ISO 와 같은 표준화 기관에서는 멀티 모달리티 사용에 관한 가이드라인과 Multimodal Interface Framework, Multimodal Markup Language 등과 관련한 멀티모달 인터페이스 표준화 방안에 대해 모색중이며 계속해서 새로운 표준안들이 제안되고 업데이트 되고 있는 상황이다.

그러나 멀티모달을 이용한 사용자 입출력방식에 관련한 형식화 설계방안이나 출력 메타포에 관해서는 아직 많은 부분에서 연구 보고되지 않은 실정이다. 현재까지 개발된 멀티모달 사용자 인터페이스는 주로 특정 어플리케이션의 일부 기능을 지원하기 위해 사용된 입출력 형식일 뿐 보편화될 수 있고 범용성 있는 입출력 표준형식은 제시하지 못하였다. 이러한 이유는 아직 멀티모달 인터페이스가 대중들에게 보편화되지 않은데다가 새로운 기술들이 계속적으로 등장하고 있어 현재까지 사용자 입출력방식을 표준화하기에 어려움이 있기 때문인 것으로 사료된다.

3. 설계 프로세스

그림 1 은 멀티모달 인터페이스 설계 프로세스를 나타낸 것이다. 크게 출력부분과 입력부분설계로 구분되며 출력부분은 다시 정보 설계, 인터랙션 설계, 프리젠테이션 설계, 입력 설계로 나뉜다.

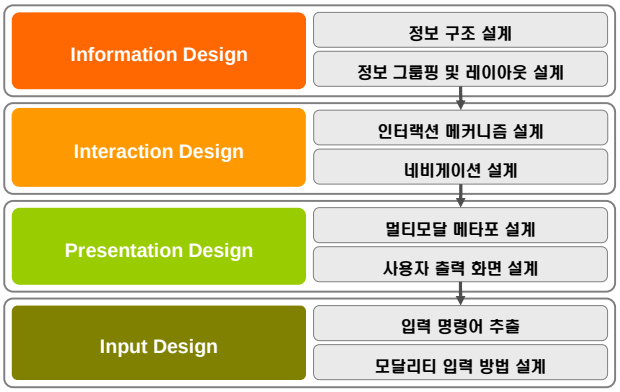

그림 1. 멀티모달 인터페이스 설계 프로세스

4. 출력 설계

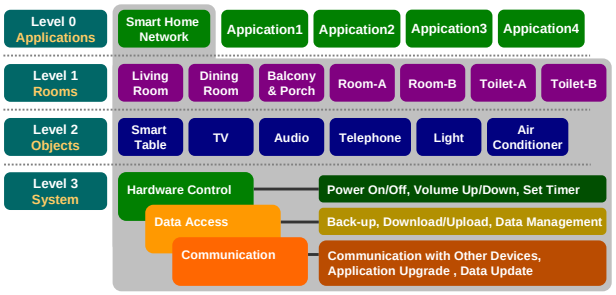

그림 2. 정보 설계

각 오브젝트를 크게 하드웨어 제어 가능한 오브젝트, 하드웨어 제어와 데이터 접근이 가능한 오브젝트, 하드웨어 제어와 데이터 접근, 네트워크 통신 3 가지가 모두 가능한 오브젝트로 나누었다. 그림 3 은 멀티모달 인터랙션 메커니즘을 적용한 출력화면을 나타낸 것이다. 메커니즘은 크게 Dual-Prompt 와 Input Grammar, Modality Switch, Direct Menu Access 로 구분될 수 있다.

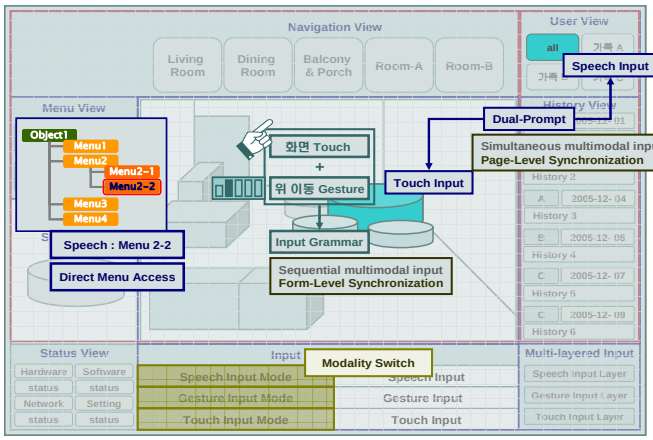

그림 3. 인터랙션 메커니즘

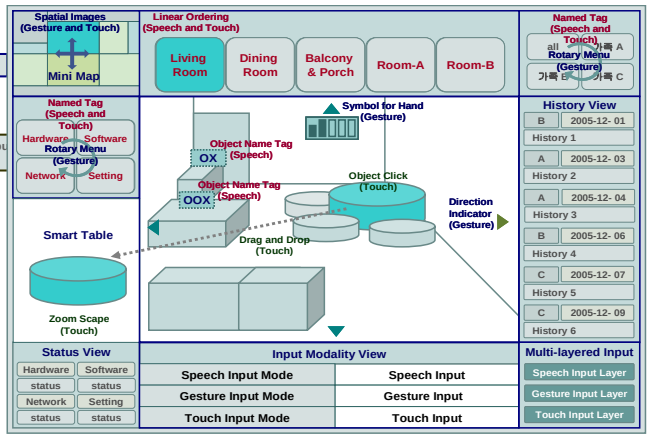

그림 4. 화면 출력 설계

- **Dual Prompt:** Simultaneous multimodal input 을 위한 Dual-prompt 이다. 두 가지 모달리티를 두 개의 프로세스에 동시 입력 가능하다. Dual-Prompt 는 Session, Page, Form, Field, Event 의 독립적인 프로세스 간에 이루어진다[8].
- **Input Grammar:** 모달리티 입력 순서와 결합 방식 등에 관한 입력 형식이다.
- **Input Modality Switch:** 입력 모달리티 전환
- **Direct Menu Access:** 음성과 제스처를 이용하여 2 차, 3 차 Depth 메뉴의 직접 접근이 가능하다.

5. 입력설계

5-1 [OAV] 멀티모달 입력 문법

멀티모달리티를 이용한 입력문법 [OAV]형식을 제안하고 Speech, Touch, Gesture 입력의 통합방법 및 동기화 방법에 대해 설명한다. 각각의 모달리티는 청각, 시각, 촉각 모달리티에 해당한다.

[OAV] 형식은 그림 6 에서 나타낸 것과 같이 멀티모달 언어입력 순서를 바탕으로 정해진 이름이다. O 는 오브젝트(Object)를 뜻하며, A 는 속성(Adverb), V 는 행위(Verb)를 의미한다. 그림 5 테이블에서 보여주듯이 영문형식을 기준으로 멀티모달 언어의 문장구조를 크게 '동사(V)', '동사+목적어(V+O)', '동사+목적어+부사(V+O+A)'로 구분하였다. 그리고 각각의 형식의 입력 순서는 그림 6 와 같이 지정하여 모달리티 입력방식을 형식화 하였다.

Type	Verb		Object		Adverb
V	Yes/No				
	Exit				
	Redo/Undo				
V+O	Save	+	the file		
	Execute	+	the application		
	Delete	+	the text		
V+O+A	Move	+	the object	+	here to there
	Scale	+	the image	+	x 2 or x3
	Rotate	+	the map	+	clockwise

그림 5. 멀티모달 입력 문법

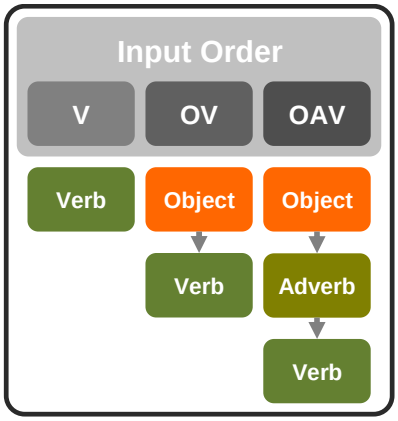

그림 6. 모달리티 입력 순서

많은 종류의 명령어들이 객체(object)에 행위(action)를 적용시킨다. 예를 들어 글자의 크기를 변경하고자 하는 경우 사용자가 크기를 변경하고자 하는 글자를 선택한 후에 글자의 크기를 변경할 것이다. 포토샵과 같은 그래픽 툴에서는 행위에 대한 명령을 선택하고 객체를 선택하기도 한다. 예를 들면 이미지를 이동시키고자 할 경우

"Move" 툴을 선택하고 이동하고자 하는 객체를 이동한다. 이 경우에는 세밀한 조작을 필요로 하거나 정교한 수정을 해야 하는 경우 이미지 속성 값이 여러 번 적용될 수 있기 때문에 속성 값은 맨 나중에 지정하는 것이 더 효율적이다.

입력 순서는 가장 잦은 수정 빈도를 나타내는 명령을 가장 나중에 입력하고 가장 적은 빈도의 명령은 제일 처음에 입력하도록 하는 것이 바람직하다. 본 연구에서는 오브젝트 객체를 먼저 선택하고 후에 행위명령을 입력하는 형식으로 멀티모달 언어를 설계하였다. 만약 어플리케이션이 포토샵과 같은 그래픽 툴과 같이 잦은 '속성' 변경과 '동작' 명령을 수행하는 경우에는 입력 순서를 조정해야 할 필요가 있을 것이다. [VOA] 또는 [VAO]

5-2 Touch, Gesture, Speech 멀티모달 입력

음성은 Name Tag 기반의 인터페이스로 작동하여, Touch 는 메뉴클릭기반으로, Gesture 는 이동커서 및 제스처 명령기반으로 작동할 수 있다. 만약 사용자가 Undo 명령을 입력하고자 할 경우 입력해야 할 명령어 형식은 [V]에 해당할 것이다. 각 모달리티를 사용하여 '되돌리기(Undo)'명령을 입력하는 방법은 Touch 를 사용하여 '되돌리기' 메뉴를 클릭하거나 Speech 를 사용하여 '되돌리기' 명령을 말하거나 Gesture 는 '되돌리기' 명령으로 부호화된 Gesture 를 입력할 수 있을 것이다. 또한 사용자가 만약 세 번째 [OAV]형식에 해당하는 '그림을 2 배로 확대시켜라'는 명령을 수행할 경우 사용자는 Touch 나 Speech 로 해당 오브젝트(이미지)를 선택한 후 Speech 나 Gesture, Touch 등으로 '2 배'를 입력하고 마지막에 마찬가지로 Speech 나 Gesture, Touch 등을 사용하여 '확대하라' 명령을 수행할 수 있을 것이다.

5-3 [OAV] 문법과 관련한 Gesture, Speech, Touch 모달리티 입력 비교

- Speech 와 Touch 는 오브젝트를 선택하기 용이한 반면 Gesture 는 그렇지 못하다.
- Speech 와 Touch 를 사용하여 오브젝트를 가리키는 것은 '오브젝트를 선택하다'의 동사의미를 내포한다. Gesture 의 경우 [Roll & Pitch 정보]+[Gesture]를 이용하여 오브젝트 방향과 관련된 '속성(Adverb)'정보와 '행위(Verb)'정보를 동시에 나타낼 수 있다.
- 속성정보와 행위정보를 적용할 오브젝트가 한 개밖에 없는 경우에는 오브젝트를 선택하지 않아도 인터랙션 매니저에 의해 자동적으로 해당 오브젝트에 명령이 적용되도록 한다.
- Speech 의 경우 음성인식기의 단어 인식 범위에 따라 오브젝트와 속성, 행동을 하나의 명령어처럼 입력할 수 있다. Touch 와 Gesture 는 동시 입력이 불가능하며 연속적으로 사용한다 하더라도 두 모달리티 모두 손을 사용하여 입력되기 때문에 사용자가 입력 방식에 혼란을 겪을 수 있다.

5-4 Touch, Gesture, Speech - Object Selection

Speech, Touch 와 Gesture 멀티모달 입력을 이용하여 Object 를 선택하는 방법들을 살펴본다. 사용자가 만약 Speech, Touch 를 사용하여 오브젝트를 선택한다고 할 경우 Speech 이나 Touch 둘 중에 어느 하나를 사용하더라도 동일하게 효과적인 수행이 가능하다. 예를 들어 사물에 태그(다른 태그들과 사용에 있어 구별되어야 함)가 제시되면 태그는 음성인식 어휘를 할당하여 오브젝트 선택이 가능하며, 터치나 펜 입력을 사용하여서도 동일한 오브젝트 선택이 가능하다. 오브젝트에 태그(Tag)를 달거나 손으로 가리키는 구체화된 행동을 통해서 사물의 고유한 identification 을 충분히 나타낼 수 있고 애매하거나 잘못된 참조를 피할 수 있다[6]. 그러나 제스처 명령의 경우 자유로운 오브젝트 선택이 어렵다. 이유는 제스처는 사용자가 행동하기 원하는 작업을 기준으로 부호화되는데 온갖 오브젝트들을 제스처로 부호화하기는 힘들기 때문이다. 제스처를 사용하여 오브젝트를 선택하기 위해서는 새로운 방식의 메커니즘이 설계되어야 한다. 예로 Roll & Pitch 를 이용한 포인터 네비게이션 방법이 있다. 이 외에 Passive input modality 들과 결합한 Blended input mode 도 가능한데[5] Passive modality 로는 동공의 위치나 머리의 방향 움직임 값을 사용하여 오브젝트의 위치를 추적하

고 해당 오브젝트에 포인터가 위치할 때 제스처를 입력하는 방법도 가능할 것이다[2]. 제스처 대신에 음성과 동공의 움직임을 이용하여 마찬가지의 오브젝트 선택이 가능한데 이 경우에는 시선을 오브젝트에 고정시키고 음성으로 '이것', '저것' 등과 같은 지시적 표현을 이용해 오브젝트 선택이 가능하다[1][3].

5-5 Touch, Gesture, Speech - Menu Navigation

메뉴 선택을 하기 위한 멀티모달 입력을 형식화하는 것도 앞에서 설명한 Object Selection 과 비슷하다. Speech 와 Touch 는 메뉴입력에 효율적으로 사용이 가능하지만 Gesture 의 경우 새로운 입력 메커니즘을 필요로 한다. 예로 사용자가 Speech 와 Touch 를 이용하여 풀다운 메뉴를 선택하고자 한다고 가정하자. Speech 로 풀다운 메뉴를 선택하는 경우 시각적으로 메뉴표시를 보지 못할 것이기 때문에 사용자가 선택하려는 메뉴 옵션에 대한 사전지식이 있어야 한다. 만약 사용자가 메뉴에 대한 사전지식이 있는 경우 여러 개의 Menu Depth 를 거치지 않고서도 한 번에 효율적으로 자신이 실행하고자 하는 메뉴를 선택할 수 있다. 음성은 계층메뉴구조 또는 Direct Menu Access 와 관련해서 발생하는 행동 제어 문제들을 해소시킬 수 있다. 극단적인 예를 들면 스크린과 사용자가 멀리 떨어져 있거나 손을 제대로 사용하지 못하는 경우 Speech 는 Touch 에 비해 탁월한 성능을 보일 것이다[9].

하지만 초보 사용자에게 음성은 펜과 거의 동일한 수행 효과를 보이거나 그보다 못한 수행효과를 보일 수 있다. 이유는 초보자는 Menu Depth 정보를 알지 못하기에 모든 메뉴를 일일이 다 확인해야 하고 Speech 의 경우 Touch 입력에 비해 인식정확도가 떨어지기 때문이다. 초보자는 대부분의 경우 Direct Menu Access 가 불가능하기 때문에 올바르게 수행한다 할 경우에도 Speech 는 Touch 과 거의 동일한 수행효과를 보인다. 또 Speech 의 경우 에러발생 시에 이전 상태로 되돌리는데 Touch 입력과 비교해서 에러발생확률이 더 많기 때문에 수행속도가 더 떨어진다. Touch 는 사용자로 하여금 메뉴 옵션 경로를 더욱 더 분명하게 알도록 하기 때문에 초보 사용자들의 계층구조 메뉴 네비게이션 입력 매개체로 선호되어 진다.

Gesture 를 이용하여 메뉴 네비게이션을 하기 위해서는 입력을 위한 새로운 메커니즘이 필요하다. 본 연구에서는 '이동커서'를 이용한 메뉴 네비게이션을 제안하였다. 알기 쉬운 예로 일반적인 윈도우 창에서 '탭(Tab)'키를 누르면 선택 가능한 버튼들이 돌아가면서 활성화되는 것을 볼 수 있을 것이다. 이것을 Gesture 입력에 적용하여 Gesture 입력이 한 번 적용할 때마다 Tab 키와 같이 해당 메뉴들이 활성화되고 이 후 사용자가 2 차로 선택 Gesture 를 입력할 수 있도록 하였다. 인터랙션 레벨은 Session> Page> Form> Field> Event 로 구분될 수 있으며 사용자는 Gesture 를 이용하여 인터랙션 수평레벨 간, 수직레벨 간 이동이 가능할 것이다. 그림 7 은 인터랙션 레벨에 따라 Gesture 메뉴 네비게이션 과정을 도식화한 것이다. Gesture1 은 수평적 이동을 나타내며 Gesture2 & Gesture3 은 수직적 이동을 나타낸다.

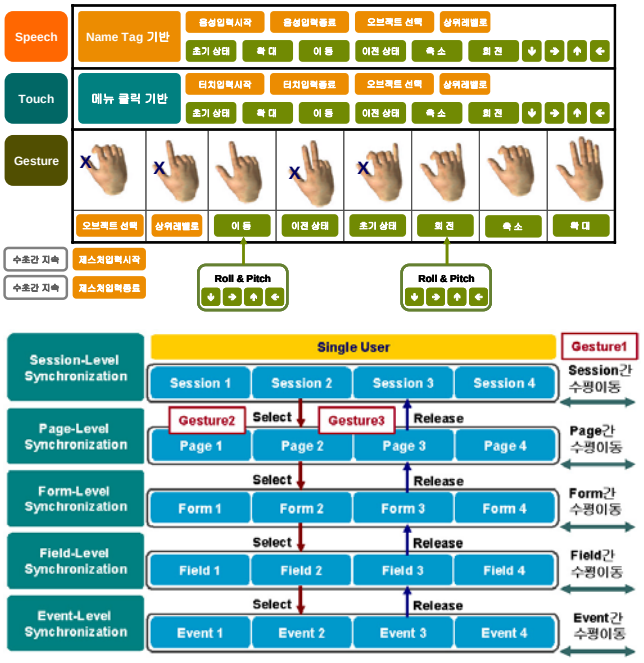

그림 7. Speech, Touch, Gesture 를 이용한 메뉴 네비게이션

5-6 Touch, Gesture, Speech - Object Manipulation

Touch 와 Gesture, Speech 를 이용하여 가상공간에서 오브젝트를 직접 조작하는 방식에 대해 살펴본다. 먼저 Touch 를 사용하여 사물을 직접 조작하는 경

우 윈도우 폼 안에서 이동하고자 하는 방향으로 손가락을 움직여 오브젝트를 이동시킬 수 있을 것이다. 또는 오브젝트의 각 모서리를 잡고 변경하고자 하는 크기만큼 오브젝트를 늘리거나 축소시키거나 회전시킬 수 있을 것이다. 마찬가지로 Gesture 를 이용해서도 오브젝트의 효과적인 직접 조작이 가능한데 Roll 과 Pitch, Static Gesture, Dynamic Gesture 등을 이용한 다차원적인 오브젝트 조작이 가능하다. Gesture 의 경우 3D 가상 환경에서 오브젝트 직접조작에 탁월한 수행능력을 보인다. Speech 명령을 사용하여서 오브젝트를 디폴트 값의 크기와 위치를 갖는 오브젝트를 생성할 수도 있을 것이다. Speech 입력으로는 구체적인 공간상에 위치나 크기를 정하기 어렵기 때문에 직접 조작에는 적합하지 않다[6].

6. 입력 모달리티 통합

멀티모달 입력 동기화 형식은 크게 다음과 같이 구분될 수 있다. [8][그림 9]

Unimodal input 단일 모달리티 입력.

Simultaneous multimodal input 여러 어플리케이션 작업에서 두 가지 이상의 입력을 동시 다른 어플리케이션에 전달할 수 있어 병렬작업이 가능하다.

Composite multimodal input Simultaneous multimodal input 과 유사하지만 여러 모달리티 정보가 결합하여 한 가지의 어플리케이션 명령을 수행하는데 사용된다는 차이점이 있다.

Sequential multimodal input 순차적으로 모달리티를 입력하는 것으로 두 개 이상의 모달리티들이 합쳐져 하나의 명령어를 이룬다. 모달리티들 간의 결합이 가능한 time interval 을 명시해주어야 한다.

7. Dual-Prompt

그림 9 는 멀티모달 입력 동기화 레벨에 따라 인터랙션을 모형화한 것이다. 멀티모달 입력은 Simultaneous multimodal input 을 이용한 병렬작업이 가능하므로 이를 수용할 수 있는 Dual-Prompt 가 필요하다. Dual-Prompt 는 병렬작업을 진행하기 위한 어플리케이션의 준비상태라 볼 수 있다. 병렬작업 과정에서 동시 처리되는 프로세스는 서로 종속관계가 아닌 상호 대등관계, 독립적인 관계여야 한다. 왜냐하면 종속적인 프로세스의 경우 병렬로 신호를 입력할 경우 하부프로세스의 명령과 상위 프로세스 명령 간에 충돌이 일어날 수 있기 때문이다. Simultaneous multimodal input 을 이용하여 사용자는 결과적으로 2 개 이상의 Session, Page, Form, Field, Event 등에서 동시간 병렬작업이 가능하다. 하지만 Simultaneous multimodal input 의 경우 사용

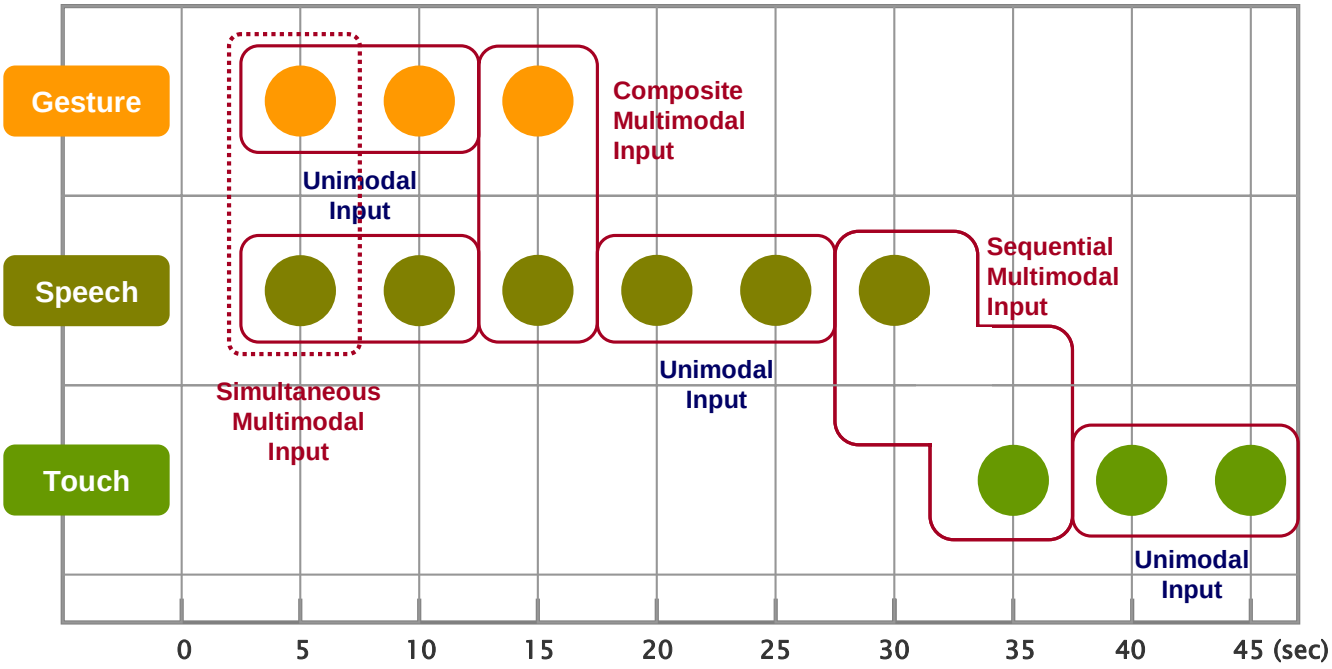

그림 8. 입력 모달리티 통합 및 동기화 방법

자의 주의 자원(attention resource)의 배분이 제대로 이루어지지 않을 수 있기 때문에 각각의 Speech, Gesture, Touch 입력을 함에 있어서 사용자가 양립적인 개념을 갖고 사용할 수 있도록 하는 것이 중요하다. Dual-Prompt & Dual-Task 간에는 개념적인 양립성을 유지하고 있어야 하며, 사용자가 손으로 Image 를 조작하면서 동시에 음성명령을 입력하는 등의 동시적 병렬작업이 가능해야 한다[4].

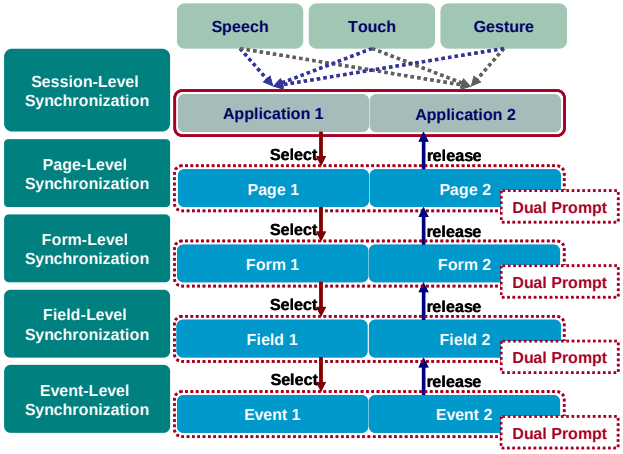

그림 9. 인터랙션 동기화 레벨과 Dual-Prompt

8. 통합 패턴과 동기화 레벨

그림 10 은 입력 모달리티 레이어와 동기화 레벨, 입력 동기화 방법 등을 4 개의 축을 이용하여 표현한 것이다. X 축은 시간을 나타내며, Z 축 양의 방향은 Multi-Layered Input 을, Z 축 음의 방향은 모달리티 동기화 레벨을, Y 축은 모달리티 입력 개수를 나타낸다. 그림 10 은 4 개의 축에서 보여지는 Composite multimodal input 과 Sequential multimodal input, Simultaneous multimodal input 을 표시한 그림이다.

9. 토론 및 향후 연구

본 연구은 멀티모달 인터페이스를 개발하기 위한 휴먼-컴퓨터 인터랙션 설계 방안 제시에 중점을 두었다. 현재까지 HCI 적인 관점에서 멀티모달 인터페이스를 체계적으로 분석한 자료가 많지 않고 출판된 책도 거의 없기 때문에 멀티모달 인터페이스에 대한 논문들과 표준화 자료를 중심으로 멀티모달 인터페이스에 대한 설계방안을 제안하였다. 본 논문에서 제안된 내용들을 실제 어플리케이션 상에서 어떻게 구현해야 할지에 대해서는 아직 더 연구가 이루어져야 할 것이다. 그리고 앞으로 이러한 멀티모달 인터페이스가 과연 사용자들에게 효율적인 인터랙션이 가능하도록 하게 할 수 있을지 없을지에 대해서 사용성 평가와 검증실험을 통한 평가가 이루어져야 한다. 향후 설계된 프로토

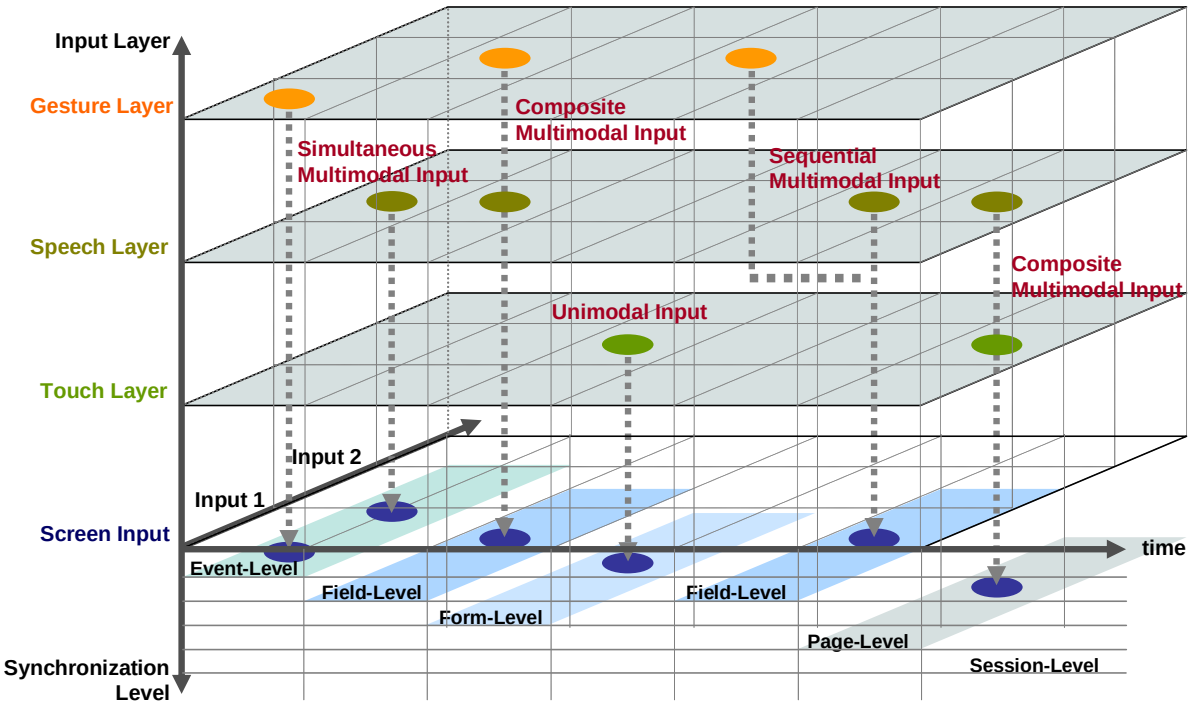

그림 10. 멀티모달 통합패턴과 동기화 레벨

타입을 보충하여 완전한 시뮬레이션 형태의 멀티모달 인터페이스를 제작해 볼 계획이다. 인간중심의 멀티모달 인터페이스를 구현하기 위해 계속적인 인지실험과 사용성 평가를 바탕으로 한 인터페이스 수정, 보완이 이루어져야 할 것이다.

10. 참고 문헌

1. Qiaohui Zhang, Atsumi Imamiya, Kentaro Go, Xiaoyang Mao, "Gaze and Speech Multimodal Interface," *Int. Conf. on Distributed Computing Systems Workshops(ICDCSW'04)*, pp.208-214
2. D.B. Koons, C.J. Sparrell, & K.R. Thorisson, "Integrating simultaneous input from speech, gaze, and hand gestures. M.Maybury(Ed.)," *Intelligent Multimodal Interfaces*, pp.257~276, Menlo Park, CA: MIT, 1993.
3. S.L. Oviatt, Mutual disambiguation of recognition errors in a multimodal Architecture, in *Proc. CHI' 99 Human Factors in Computing Systems Conf.*, Pittsburgh, PA, 576-583.
4. Chih-Yuan Ho, Mark I. Nikolic, and Nadine B. Sarter, "Multimodal information presentation in support of timesharing and effective interruption management," *Digital Avionics Systems, 2001. DASC. The 20th Conference Vol. 1*, 14-18 Oct. 2001 Page:5D2/1 - 5D2/8 vol.1
5. S. L. Oviatt, "User-Centered Modeling and Evaluation of Multimodal Interfaces," in *Proc. of the IEEE*, vol. 91, NO. 9, 2003, pp. 1457-1468
6. Jennifer L. Leopold and allen L. Ambler, "Keyboardless Visual Programming Using Voice, Handwriting, and Gesture", in *Proc. of the 1997 IEEE Symposium on Visual Languages(VL '97)*, pp. 28-35
7. http://www.w3.org/TR/2005/WD-mmi-arch-20050422
8. http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/
9. Oviatt, S.L. User-centered modeling for spoken language and multimodal interfaces, *IEEE Multimedia*, winter 1996, vol. 3, no. 4, 26-35