

음성기반 멀티모달 사용자 인터페이스의 사용성 평가 방법론*

홍기형 (성신여대)

<차 례>

- | | |
|------------------------|------------------------------------|
| 1. 서론 | 3.2. PROMISE |
| 2. 사용성과 사용자 중심 설계 | 4. 음성기반 멀티모달 인터페이스 사용성
평가 가이드라인 |
| 2.1. 사용성 | 4.1. 멀티모달 인터페이스의 사용성 |
| 2.2. 사용자 중심 설계 및 중요성 | 4.2. 사용자 중심 설계 방법론 |
| 3. 멀티모달 인터페이스 범용 평가 척도 | 5. 결론 |
| 3.1. PARADISE | |

<Abstract>

Usability Test Guidelines for Speech-Oriented Multimodal User Interface

Ki-Hyung Hong

Basic components for multimodal interface, such as speech recognition, speech synthesis, gesture recognition, and multimodal fusion, have their own technological limitations. For example, the accuracy of speech recognition decreases for large vocabulary and in noisy environments. In spite of those technological limitations, there are lots of applications in which speech-oriented multimodal user interfaces are very helpful to users. However, in order to expand application areas for speech-oriented multimodal interfaces, we have to develop the interfaces focused on usability. In this paper, we introduce usability and user-centered design methodology in general. There has been much work for evaluating spoken dialogue systems. We give a summary for PARADISE (PARADigm for DIalogue System Evaluation) and PROMISE (PROcedure for Multimodal Interactive System Evaluation) that are the generalized evaluation frameworks for voice and multimodal user interfaces. Then, we present usability components for speech-oriented multimodal user interfaces and usability testing guidelines that can be used in a user-centered multimodal interface design process.

* Keywords : Speech-oriented multimodal interface, Usability, User-centered design.

* 이 논문은 2006년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음.

1. 서 론

정보화 기기의 소형화 및 이동성의 혁신적인 증가, 그리고 다양한 멀티미디어 정보서비스의 확산은 정보 접근 및 사용을 위한 음성 사용자 인터페이스의 중요성을 크게 증대시키고 있다. 음성 인식 및 합성 기술을 기반으로 하는 음성 사용자 인터페이스는 키보드와 같은 대형 인터페이스 하드웨어를 부착하기 어려운 휴대전화를 비롯한 개인용 멀티미디어 단말기에서 가장 활용하기 편한 인터페이스이다.

지금까지의 음성 기술의 발전에도 불구하고, 음성 인식 기술은 잡음에 취약하며, 사용자의 음성 특성에 따라 성능의 차이를 보이는 등 기술의 한계를 가지고 있다. 또한 사람이 가장 자연스럽게 사용하는 정보 교환의 수단이 음성 대화인 것은 분명하나, 실제로는 음성과 함께 손동작, 몸동작을 함께 사용하는 멀티모달 인터페이스가 보다 자연스럽다[1][2][3].

음성기반 멀티모달 인터페이스는 음성과 함께 터치스크린, 키패드 등의 모달리티를 사용자가 동시에 사용할 수 있는 사용자 인터페이스이다. 이러한 멀티모달 인터페이스는 각 모달리티가 가진 기술적인 한계를 상호 보완할 수 있고, 사용자가 자신이 선호하는 모달리티를 선택하여 사용할 수 있는 장점이 있다.

현재, 음성기반 멀티모달 인터페이스를 구현하기 위한 기반 기술인 음성 합성, 음성 인식, 터치스크린 인식, 멀티모달 퓨전(fusion) 등은 나름대로 기술적인 한계를 가지고 있으나 이러한 한계를 감안한 응용 분야에서 상업화가 가능한 기술이다. 그러나 음성기반 멀티모달 인터페이스의 확산이 제대로 이루어지고 있지 못하고 있는 실정이다. 이에 대한 가장 큰 이유는 음성 인터페이스와 이를 기반으로 하는 멀티모달 인터페이스에 대한 사용자의 만족도가 크지 못하고 편리함을 주지 못하기 때문이다. 음성기반 멀티모달 인터페이스에 대한 사용자의 만족도를 높이고 사용자가 쉽게 적응하고 편리하게 사용할 수 있는 인터페이스임을 느낄 수 있도록 하기 위해서는 사용성(혹은 가용성이라고도 함, usability)에 초점을 맞추어 충분한 사용성 평가를 시행하여 개발하여야 한다[4].

사용성은 시스템 평가와는 다르다. 시스템 평가가 일반적으로 기능의 정확성과 성능에 초점을 맞추는 것이라면, 사용성은 사용자가 어떻게 받아들이는 지에 초점을 맞추는 것이다. 사용성에 초점을 맞춘 개발 방법론을 사용자 중심 설계라고 한다.

사용성이나 사용자 중심 설계는 기존의 다양한 제품 인터페이스 개발에서 적용되어 사용되고 있으나, 음성 인터페이스의 개발에 있어서는 최근에 많은 주목을 받고 있다. 물론, 음성 인터페이스 개발에 있어서 사용성과 관련된 평가가 전혀 없었던 것은 아니며, PARADISE[5] 등과 같은 음성 대화 시스템을 위한 평가 프레임워크에서 사용성 관련 척도가 시스템 평가 척도와 함께 연구되었다. 하지만 사

용자 중심 설계와 같은 설계 개발 과정에서 적용할 수 있는 방법론으로까지는 발전하지 못하였으며 특히, 음성기반 멀티모달 인터페이스에 대한 사용성 평가는 매우 초기 단계에 있다. 또한 PARADISE를 멀티모달 인터페이스용 평가 프레임워크로 확장한 PROMISE[6]가 있다. PROMISE는 멀티모달 대화 시스템 연구를 위한 독일 SmartKom 프로젝트[7] 결과물의 일부이다.

본 논문에서는 2장에서 사용성과 사용성 평가에 대하여 소개하고, 3장에서는 PROMISE를 중심으로 멀티모달 인터페이스의 사용성 평가에 대하여 설명한다. 4장에서는 음성 기반 멀티모달 인터페이스 개발을 위한 사용자 중심 설계에 대하여 논하고 5장에서 향후 연구 과제를 기술하고 결론을 맺는다.

2. 사용성과 사용자 중심 설계

2.1. 사용성

J. Nielsen[8]은 일반적인 사용성(usability)을 다음과 같은 특성으로 정의하였다. 여기서 일반적이라 함은 다양한 제품, 기기, 인터넷 홈페이지 등 사용자 인터페이스가 적용되는 모든 시스템에서 공통으로 적용할 수 있다는 뜻이다.

- 1) 학습의 용이성(learnability): 사용자가 배우기 쉬워야 한다.
- 2) 효율성(efficiency): 사용자가 인터페이스를 사용하여 높은 생산성을 달성할 수 있어야 한다.
- 3) 기억의 용이성(memorability): 인터페이스의 사용 방법을 기억하기 쉬워서 한 동안 사용하지 않고 있다가 다시 사용하더라도 처음부터 다시 배우는 수고가 없어야 한다.
- 4) 오류에 대한 대처(errors): 낮은 오류율을 가져야 하며, 오류가 있더라도 사용자가 이를 쉽게 수정하거나 오류 발생 전의 상태로 복귀할 수 있어야 한다.
- 5) 만족도 (satisfaction): 사용하기에 즐거워야 하고, 사용을 두려워하거나 싫어하지 않아야 한다.

음성 인터페이스, 단일 모달리티 인터페이스 시스템에 대한 사용성도 아직까지는 위의 일반적인 사용성에 더하여 어떠한 특성이 명확하게 정의하여야 하는 지 연구 중에 있다.

2.2. 사용자 중심 설계 및 중요성

특정 시스템을 개발하기 위한 기술의 발전이 상용화에 문제가 없을 정도로 발전한 경우, 시스템의 설계는 그 시스템이 가져야 하는 기능의 검증이나 성능보다

는 사용자의 사용성에 초점을 맞추어야 한다. 이와 같이 사용성에 초점을 맞추어 설계하는 방법론을 사용자 중심 설계라고 하고, 여기에 중심은 사용성이다.

음성 인식, 합성, 키패드, 터치스크린 인식 기술, 멀티모달 퓨전 등과 같은 음성 및 음성 기반 멀티모달 인터페이스를 위한 기반 기술은 각기 나름대로 기술적인 한계를 가지고 있으나, 각 기술 요소가 가지는 현재의 기술적 한계를 감안하여 적용할 수 있는 응용 영역이 존재하는 단계에 도달했다고 판단된다. 현재, 많은 음성 기반 사용자 인터페이스가 상용 시스템에 탑재되어 있으나 그 사용은 크게 확산되고 있지 못한 상황이다. 상용 시스템에 채용되어 있다는 것은 음성 인식이나 합성 기술과 같은 기반 기술이 앞에서 언급한 대로 상용화 수준에 도달하였으나 아직까지 사용성 평가에 중심을 둔 사용자 중심의 시스템 설계가 제대로 이루어지지 못하고 있다고 할 수 있다.

일반적인 사용자 중심 설계는 다음과 같은 3가지 원칙이 있다 [9].

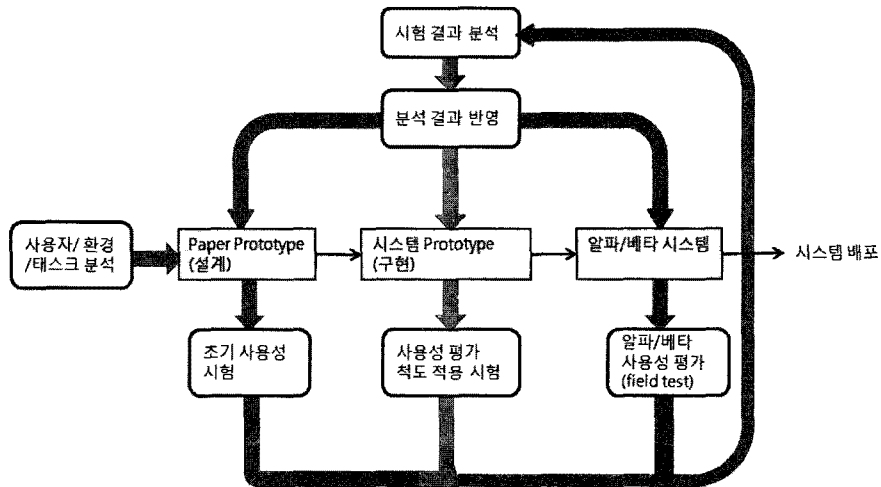
- 1) 사용자와 사용자가 수행하는 작업(task)에 먼저 초점을 둔다. 사용자에 대한 이해, 사용자가 수행할 작업에 대한 이해, 그리고 사용자의 작업 수행 환경에 대한 이해를 먼저 한다.
- 2) 사용성 평가는 실 사용자를 대상으로 실험 및 관찰을 통하여 시행한다.
- 3) 사용성 평가를 시스템 개발 생명 주기에 걸쳐 반복 시행하고 그 결과를 반영한 반복 설계를 한다.

<그림 1>은 사용자 중심 설계를 통한 일반적인 제품의 개발 과정을 보여준다. 사용자 중심 설계의 첫 단계는 아래에서 설명하는 바와 같이 사용자/환경/태스크 분석이다.

- 사용자분석 : 개발하고자 하는 시스템의 대상 사용자(target user group)에 대한 분석을 수행한다. 대상 사용자 그룹의 연령 대, 성별, 첨단 기술 제품에 대한 적응력 등의 특성을 바탕으로 대상 사용자 그룹을 인지한다.
- 사용 환경 분석: 목표 시스템의 주사용 환경의 특성을 알아낸다. 사무실, 집안, 이동 중 등과 같이 주로 사용하게 될 환경에 대한 분석을 한다.
- 태스크 분석: 시스템을 이용하여 사용자가 수행하여야 하는 태스크를 상세히 분석한다. 여기에는 태스크 수행과정에 따라 사용자로부터의 입력 정보, 수행할 기능, 출력 정보를 태스크 별로 상세히 도출하여야 한다.

이상과 같은 사용자/환경/태스크 분석 결과를 바탕으로 프로토타입(paper prototype)을 설계한다. 설계의 결과는 초기 사용성 시험을 통하여 수정 보완을 반복한다. 초기 사용성 시험에서는 대상 사용자 그룹을 대표할 수 있는 사용자를 선정하고 이 사용자를 대상으로 한다. 대표 사용자와의 인터뷰를 통하여 설계 결과를 설명하고 설계에 대한 느낌과 태스크 수행과정에서 인터페이스의 편리성 및 개선할 부분에 대한 응답을 듣는다.

초기 사용성 시험과 수정 보완을 반복하여 대표 사용자의 만족도가 원하는 수



<그림 1> 사용자 중심 설계

준 이상이 될 경우, 프로토타입을 구현한다. 여기서 구현하는 프로토타입은 wizard-of-Oz (WOZ) 시험을 위한 프로토타입부터 구현할 수도 있다. WOZ 시험을 위한 프로토타입은 인터페이스 부분을 사용자 (피시험자) 시험에 편리한 형태의 기계에 먼저 구현하고, 실제 기능은 시험자가 대신 수행하는 형태의 초기 프로토타입이다. WOZ 시험에서는 피시험자를 대상 사용자 그룹에서 가능한 다수를 참여시키는 것이 바람직하다. WOZ 시험에서는 트랜잭션 성공률과 같은 정량적 사용성 평가 척도와 실제 상황에서 사용자의 발화 형태 등을 파악할 수 있으며, 이를 실제 시스템의 구현에 반영할 수 있다.

프로토타입에 의한 반복적인 사용성 시험을 거쳐 보완된 시스템은 알파 테스트를 위한 실제 시스템으로 구현된다. 여기서 부터는 실제 필드 테스트라고 하는 시험을 시행하게 된다. 알파 테스트를 거쳐 보완된 시스템은 실 사용자 그룹에게 배포하는 베타 버전이 되며, 일정 기간의 베타 테스트 기간을 거쳐 실 사용자를 대상으로 사용성 시험을 진행한 다음, 공식적인 배포에 나서게 된다.

3. 멀티모달 인터페이스시스템의 범용 평가 척도

음성 인터페이스 시스템의 사용성 평가는 정성적인 평가가 일반적이나 정량적 평가 척도 역시 존재한다. 트랜잭션 성공률(transaction success rate), 작업 완료 시간(task completion time), 사용자 인터랙션 회수(number of interactions), 부적절 발화 비율(inappropriate utterance ratio), 그리고 개념 정확도(concept accuracy) 등이 있다.

상업화가 완료된 음성 인터페이스 시스템에 대한 사용성 평가는 통계자료에 기초하여, 트랜잭션 성공률로 평가되는 경우가 많다. 그러나 높은 트랜잭션 성공률을 보이는 시스템이라고 하여, 반드시 사용자 만족도가 높다고는 할 수 없다. 또한, 위에서 언급한 각각의 정량적 평가 방법들은 특정한 태스크나 특정한 사용 환경을 가정한 특정한 사용 시나리오에 국한하여 사용할 수 있으나, 다양한 형태의 음성 인터페이스 시스템 또는 멀티모달 인터페이스 시스템에 대하여 일반화하기가 어렵다.

본 장에서 소개하는 PARADISE는 이러한 개별 평가 척도를 음성 인터페이스 시스템에서 일반화한 프레임워크이다. 또한 PROMISE는 PARADISE의 평가 프레임워크를 멀티모달 인터페이스로 확장한 것이다.

3.1. PARADISE

PARADISE (PARAdigm for DIalogue System Evaluation)[5]는 음성 대화 에이전트의 평가 프레임워크로 음성 인터페이스 시스템의 평가 방법을 하나의 함수로 정의 하였다. 이 함수는 <그림 2>와 같은 평가 모델을 기반으로 생성한 것이다.

사용자의 만족도를 최대화하기 위해서는 태스크 성공률을 최대화하고, 대화 비용을 최소화하여야 한다. 태스크 성공률을 계산하기 위하여, attribute-value matrix (AVM)로 사용자와 음성 인터페이스 시스템 사이의 정보 교환을 표현하고, 실제 음성 대화 시나리오에서 AVM에 정의된 정보 아이템들의 성공 회수를 혼동 행렬 (confusion matrix)로 조사한다. 이를 바탕으로 다음과 같은 Kappa 계수로 태스크 성공률을 표현하였다.

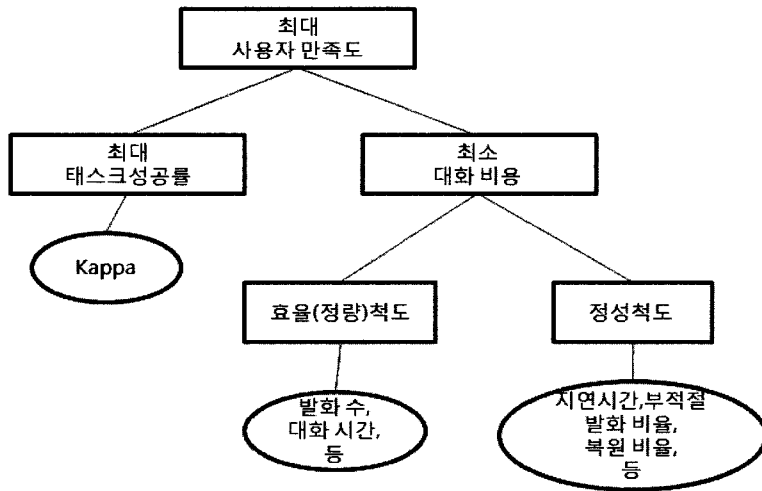
$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (1)$$

여기서 $P(A)$ 는 실제 사용자 대화에서 원하는 정보를 획득한 비율, $P(E)$ 는 대화 시나리오에서 원하는 정보를 획득할 기대치로, κ 가 0이면, 음성인터페이스 시스템의 태스크 성공률이 0%, 1이면 100%가 됨을 뜻한다.

대화 비용은 개별 척도를 고유의 함수로 계산한다. 예를 들면, 발화 회수를 c_1 으로, 태스크 완료 시간을 c_2 로 계산한다. 이런 방식으로 n 개의 척도를 사용한다면, $\{c_1, c_2, \dots, c_n\}$ 의 n 개의 척도 계산 값이 나온다.

사용자 만족도(성능)는 태스크 성공률 (Kappa 계수)과 대화 비용 (n 개의 척도 계산 값)을 기반으로 다음과 같이 정의 하였다.

$$\text{성능} = \alpha N(\kappa) - \sum_{i=1}^n \omega_i N(c_i) \quad (2)$$



<그림 2> PARADISE의 평가 모델

여기서 α 는 κ 에 대한 가산치(weight)이며, ω_i 는 c_i 에 대한 가산치이다. $N(\)$ 은 Z 스코어 함수로 $N(x) = \frac{x - \bar{x}}{\sigma_x}$ 이고 σ_x 는 x 의 표준편차이다.

3.2. PROMISE

PROMISE (PROcedure for Multimodal Interactive System Evaluation)는 SmartKom 프로젝트[7]의 부산물이다. SmartKom은 집, 공공장소, 이동 중에 사용하는 멀티모달 인터페이스 개발을 위하여 1999년부터 2003년까지 독일 정부 지원에 의해 수행된 대규모 프로젝트이며, 주요 결과물로는 상황 기반 인식, 멀티모달 인터페이스 언어, 음성 기반 웹 서비스 등이 있다.

PROMISE는 음성인터페이스 시스템 평가를 일반화한 PARADISE 프레임워크를 멀티모달 인터페이스로 확장한 것이다. 앞 절에서 소개한 성능 평가 함수는 거의 동일하나, 개별 평가 척도를 멀티모달 인터페이스를 위하여 확장하였다. 확장한 평가 척도는 정성 평가 척도와 정량 평가 척도로 나누어 볼 수 있다. <표 1>에서 보인 바와 같이 SmartKom에서는 음성, 제스처, 표정 및 GUI를 동시에 활용할 수 있는 멀티모달 인터페이스를 개발하였다. 표에서의 각 평가 척도가 PARADISE의 성능평가 함수에서 c_i 에 해당한다. 사용성 평가를 PROMISE와 같이 하나의 함수로 집적하여 수행하였다. 그러나 PROMISE 및 PARADISE의 사용성 평가 프레임워크는 개별 성능 평가 척도를 계산하여야 하는 단점이 있어 응답시간과 같이 시스템이 어느 정도 구축이 완료되어야 적용할 수 있다.

<표 1> PROMISE에서의 평가 척도

정성평가척도		정량평가척도	
의미 평가	가능한 동시 입력 수, 오해 유도 입출력 수, 입출력의 의미 정확 성 등	바지-인 (barge-in)	사용자와 시스템의 오버랩 회수
도움 평가	실 인터페이스 상황에서 제공하 는 도움 회수 등	취소	바지-인으로 인한 계획된 시 스템의 인터럽트 수
인식 모달리티	음성, 표정, 제스처 등	off-talk	시스템이 유도하지 않은 사 용자 발화 수
트랜잭션 성공률	성공한 태스크 수	경과시간	표정인식입력 시간, 제스처 입력 시간, 음성 입력시간, 음성인식 시간, 제스처 인식 시간, 평균 시스템응답시간, 평균 사용자 응답시간, 태스크 완 료 시간
오류 메시지	오류 프롬프트 수		
대화 복잡도	태스크 복잡도, 입력 복잡도, 대화관리 복잡도 (사용자로부터)		
인터페이스 방법	동시에 가능한 모달리티 수		
출력 동기성	멀티모달 출력의 동시성	시간경과	출력오류율, 입력오류율
대화 주도권	사용자/시스템 주도, 혼합 주도 등	거절	사용자의 반복 작업을 요구 하는 입력오류 회수
시스템 협동성	잘못된 입력에 대한 시스템의 대 처 능력	사용자/ 시스템 평가	대화전환 회수, 발화 단어 수, 제스처 수, 사용자 입력 오류 회복 비율

사용자 중심 설계를 통한 개발 전 과정에 걸쳐 사용성 제고를 위한 사용성 시험과 평가는 개발 과정 각 단계 별로 적절한 사용성 시험 및 자료의 수집이 필요하다. 다음 장에서는 음성 기반 멀티모달 사용자 인터페이스의 사용성에 대하여 알아보고, 사용성 평가가 중심이 되는 멀티모달 인터페이스의 사용자 중심 설계 각 단계에서 실시해야 하는 자료 수집과 분석 및 사용성 평가에 대하여 알아본다.

4. 음성기반 멀티모달 인터페이스 사용성 평가 가이드라인

4.1. 멀티모달 인터페이스의 사용성

음성 기반 멀티모달 인터페이스 시스템의 사용성에 영향을 미치는 요소는 2장에서 소개한 일반적인 사용성 요소를 다음과 같이 보다 자세히 정의할 수 있다.

- 1) 모달리티 입력 인식 정확성: 사용자 입력에 대한 인식 정확도는 사용자로

하여금 시스템에 대한 신뢰도를 높이는 것이다. 음성 인식 정확도, 터치 제스처, 표정 인식 등 개개의 입력 모달리티의 인식 정확도뿐만 아니라 음성과 제스처가 동시에 입력되는 경우와 같이 모달리티 퓨전의 정확도가 매우 중요한 요소이다. 즉, 개별 인식 정확도, 모달리티 퓨전 정확도가 사용성에 큰 영향을 미친다.

- 2) 사용자 입력의 자연성: 음성인터페이스 단일 모달리티의 경우에는 발화의 자연성, 즉 인식 문법의 설계와 관련 있다. 그러나 멀티모달 인터페이스의 경우에는 터치 제스처 등과 같은 다른 모달리티와의 동시 사용을 고려한 멀티모달 음성 인식 문법이 사용자가 입력하기 쉽게 설계되어야 한다.
- 3) 양질의 멀티모달 음성 출력: 사용자에게 응답하는, 또는 질의하는 정보는 사용자가 이해하기 쉽고, 편하게 받아들일 수 있는 형태로 가공되어야 한다. 음성 합성의 사용자 만족도, 스크린 상의 정보 레이아웃(GUI)과 같은 개별 모달리티의 출력 만족도와 다수 출력 모달리티를 결합한 출력의 만족도가 함께 측정되어야 한다.
- 4) 출력 정보의 적절성: 사용자가 태스크를 수행하는 세부 단계에서 시스템이 제공하는 출력 정보는 내용의 정확성이 담보되어야 하고, 해당 단계에 적절한 양의 정보를 포함하여야 하며 과도한 정보를 출력함으로써 사용자의 인지 부담을 주지 말아야 한다.
- 5) 피드백 적절성: 현재 시스템이 사용자의 입력정보를 잘 이해하고 있는지, 무엇을 하는 중인지 등을 사용자가 알 수 있도록 하여야 한다.
- 6) 멀티모달 대화 주도권의 적절성: 태스크의 성격, 사용자의 개별 특성과 초보 사용자인지, 고급 사용자인지에 따라, 가장 자연스러운 대화의 주도권을 제어할 수 있어야 한다. 고급 사용자는 자신이 주도권을 갖기를 원하고, 초급 사용자는 시스템이 주도하는 것이 바람직하다.
- 7) 멀티모달 대화 구성의 자연성: 태스크 수행과정의 각 단계에서 필요한 멀티모달 대화는 단계 별 필요한 정보의 배분과 대화의 구성을 사용자가 가장 자연스럽게 받아들일 수 있도록 구성하여야 한다.
- 8) 태스크 별 정보 및 서비스 지원의 충분성: 사용자는 자신이 수행하는 태스크에서 필요한 정보보다 더 많은 정보를 기대하는 경우가 많다. 이를 충분히 지원할 수 있어야 한다.
- 9) 충분한 추론 능력: 태스크에 대하여 획득한 사실로부터 추론할 수 있는 정보는 추가적인 사용자 입력을 줄일 수 있어서 충분한 추론 능력을 가져야 한다.
- 10) 충분한 사용자 안내: 시의 적절한 인터페이스 사용에 대한 도움말을 제공하여야 하며, 제공하는 도움말은 사용자에게 유용하여야 한다.
- 11) 시스템 오류 대처 능력: 시스템이 보유하고 있는 다수의 모달리티 인식 모

들 중에 하나에 오류 (또는 고장)가 있다고 할지라도 사용자와의 의사소통에 문제가 없어야 한다. 시스템 오류에 대한 대처뿐만 아니라 복원 능력이 고려될 수 있다.

- 12) 사용자 오류 대처 능력: 사용자 입력의 오류에 대하여 복원할 수 있는 기능을 갖추어야 한다.
- 13) 사용자별 대처 능력: 사용자 그룹 별로 적절한 대화 시나리오의 제공은 개개의 사용자를 만족시키는 데 매우 유용하다. 시스템 초급/고급사용자, 사용자 별 모달리티 선호도, 수행 태스크의 전문성 정도에 따라 차별화된 멀티모달 시나리오를 제공할 수 있어야 한다.
- 14) 사용환경별 대처 능력: 동일한 태스크의 수행 환경이 하나가 아닌 경우에는 사용환경별로 모달리티의 선택, 대화 시나리오의 차별화가 필요하다.

이러한 각 사용성 요소에 대한 평가 방법 및 평가 척도의 개발은 특정 분야의 멀티모달 인터페이스 개발 연구에서 이루어진 것이 일부 있으나, 일반화된 척도에 대한 연구는 시작단계라 할 수 있다.

4.2. 사용자 중심 설계 방법론

2장에서 소개한 일반적인 제품의 사용자 중심 설계의 각 과정에 대하여 음성 기반 멀티모달 인터페이스의 사용자 중심 설계에 초점을 맞추어 정리하고자 한다. 또한, 휴대전화에서의 멀티모달 음성인식 다이얼링을 위한 설계를 설계방법론에 따라 사례 연구를 수행하였다.

4.2.1. 요구 분석 단계

- 사용자 분석: 일반적인 제품 인터페이스 설계의 경우와 큰 차이는 없으나, 대상 사용자의 발화 음성의 특징을 알 수 있는 출신 지역이나 학력 수준, 그리고 첨단 기술에 대한 대상 사용자 그룹의 특성 (예를 들면, *early adapter*, *early majority*, *late majority*, *laggard*)을 인지하여야 한다.
- 환경 분석: 여기서는 멀티모달 인터페이스가 사용될 사용환경과 각 사용환경에 적절한 입출력 모달리티가 도출되어야 한다. 그리고 멀티모달 인터페이스가 적용될 하드웨어 및 소프트웨어 환경의 제약도 명확히 도출되어야 한다.
- 태스크 분석 : 멀티모달 인터페이스로 수행할 태스크를 상세히 분석하여야 한다. 태스크는 수행과정을 세분하여 하부 태스크를 정의하고, 하부 태스크 별로 사용자 입력정보, 출력 정보를 명세한다.

<표 2> 요구 분석 명세

사용자	대상 군	대상연령	성별	학력 수준	지역	특이사항
	A	20대 ~30대	남,여	고졸 이상	서울,경기	Keypad를 사용한 단순 작성이 편함
B	40대~ 60대	남,여	고졸 이상	서울,경기	Keypad를 사용한 단순 작성이 불편하고, 노안으로 작은 글씨를 보기 어려움	
휴대전화 모달리티	input modality		비고			
	keypad		12개의 번호 키			
	버튼		메뉴, 통화, 취소, 확인, 좌/우/위/아래 등			
	음성인식		고립단어 연속어			
	헤드셋 버튼		통화/종료, 볼륨 up/down			
	output modality		비고			
	화면		320x240			
사용 환경	번호	환경	사용 가능 input modality		사용 가능 ouput modality	
	1	사무실	keypad, 버튼, 헤드셋버튼		화면	
	2	가정	음성인식, keypad, 버튼		음성합성, 화면	
	3	이동중	음성인식, keypad, 버튼, 헤드셋 버튼		음성합성, 화면	
	4	운전 중	음성인식, 헤드셋 버튼		음성합성	
태스크	다이얼링		필요 정보			비고
			대상자 이름, 연락 장소(휴대전화, 집, 사무실)			

휴대전화에서 멀티모달 음성인식 다이얼링을 태스크로 가정하고, 요구 분석 단계에서 도출할 수 있는 정보를 명세하면 <표 2>와 같다. 사용자 군은 개발할 태스크의 사용 특성이 다를 것으로 예상되는 2개의 그룹(A, B)으로 나누었고, ‘사무실’ 환경은 소리를 내기 어려운 환경, ‘가정’은 소리를 낼 수 있는 환경을 대표하며, ‘이동 중’은 소리를 낼 수 있으며, 동시에 화면을 쳐다 볼 수 있는 장소 또는 이동 환경을, ‘운전 중’은 화면이나 키패드를 사용하기 어려운 장소나 이동환경을 대표한다. 태스크는 전화 걸기이며, 전화 걸기를 위하여 필요한 사용자로부터의 정보는 전화번호 또는 전화번호를 검색하기 위한 이름과 연락장소이다. 연락장소는 동일인의 전화번호가 휴대전화, 집 전화, 사무실 전화와 같이 여러 개가 존재할 수 있다.

4.2.2. 초기 사용성 시험

요구분석 단계의 결과는 인터페이스 설계서에 충분히 반영하여야 한다. 인터페이스 설계서에는 하부 태스크 별로 멀티모달 음성 대화 시나리오가 설계되어야 한다. 음성 대화 시나리오에는 사용할 입출력 모달리티와 예상되는 사용자의 입력 모달리티 조합에 충분히 대처할 수 있는 대화 설계가 이루어져야 한다.

<표 3>은 음성인식 멀티모달 다이얼링에 대한 사용자군 및 환경별로 사용자의 모달리티 친숙도에 따른 선호도를 보이고 있다. 사용자 군 B의 ‘이동 중’ 환경에서는 입력 모달리티로는 음성인식, 헤드셋 버튼, keypad, 버튼의 순으로 선호함을

<표 3> 사용자 및 환경 별 모달리티

사용자 군	환경	편한 input modality	편한 ouput modality
A	사무실	keypad, 버튼	화면
	가정	keypad, 버튼, 음성인식	화면, 음성합성
	이동중	keypad, 헤드셋 버튼, 음성인식	화면, 음성합성
	운전 중	음성인식, 헤드셋 버튼	음성합성
B	사무실	keypad, 버튼	화면
	가정	keypad, 버튼, 음성인식	화면, 음성합성
	이동중	음성인식, 헤드셋 버튼, keypad, 버튼	음성합성, 화면
	운전 중	음성인식, 헤드셋 버튼	음성합성

나타낸다.

<표 3>으로 부터 다음과 같은 사용자 인터페이스 시나리오가 마련되어야 한다.

- (1) Keypad+버튼, 화면만을 이용하는 시나리오. (사용자A-사무실/가정/이동중, 사용자B-사무실,가정)
- (2) 음성인식+헤드셋 버튼, 음성합성만을 이용하는 시나리오. (사용자A-운전중, 사용자B-운전중/이동중)
- (3) 음성인식+keypad+버튼, 음성합성+화면을 이용하는 시나리오. (사용자A-이동중, 사용자B-이동중)

실제로 시나리오 (1)은 모든 휴대전화에서 기본으로 제공하고 있는 시나리오 (전화번호 입력 후 통화 버튼, 또는 메뉴에서 전화번호 검색 후 통화 버튼 등)이며, 시나리오 (2)는 S사의 휴대전화 일부 모델에서 제공하는 시나리오이다.

<그림 3>의 시나리오에서 p는 시스템(휴대전화)을 u는 사용자의 발화를 의미하며, ②의 <2>은 keypad 2를 누르는 것도 가능함을 뜻 한다. <그림 3>의 시나리오는 keypad를 이용할 수도 있는 시나리오이나 이를 시나리오 (2)로 분류한 것은 사용자에게 각 단계에서 모두 음성 합성으로 프롬프트 하고 있으며, 실제로 화면을 볼 수 있는 상황에서는 음성 프롬프트가 사용자에게 매우 지루하게 느껴진다. 따라서 S사의 휴대전화에 구현된 <그림 3>의 시나리오는 시나리오 (2)의 해당 환경

- | |
|-----------------------------------------------------|
| ① p: "기능을 말하세요." |
| ② u: "이름으로 걸기" ; <2> |
| ③ p: "삐 소리 후 이름과 연결할 곳을 말하세요." |
| ④ p: Beep |
| ⑤ u: "홍길동 사무실" |
| ⑥ p: "홍길동 사무실 맞습니까?" |
| ⑦ u: "예" |
| ⑧ p: "홍길동 사무실로 연결합니다." (헤드셋을 사용하고 화면을 볼 수 없는 경우 유용) |
| ⑨ (Dialing) |

<그림 3> 음성인식+헤드셋 버튼, 음성합성을 이용한 시나리오 예

- | |
|--------------------------------------------|
| ① p: [음성인식 기능 메뉴 표시] |
| ② u: 이름으로 걸기 메뉴 선택 (메뉴 번호 또는 위/아래 버튼 후 확인) |
| ③ p: "삐 소리 후 이름과 연결할 곳을 말하세요." |
| ④ p: Beep |
| ⑤ u: "홍길동 사무실" |
| ⑥ p: "홍길동 사무실 맞습니까?" + [n-best 를 순서대로 표시] |
| ⑦ u: "예" <통화> |
| ⑧ (Dialing) |

<그림 4 > 음성기반 멀티모달 시나리오 설계안

과 사용자에게 맞춘 것이라 할 수 있다.

시나리오 (3)은 다음 <그림 4>과 같이 설계할 수 있다. <그림 3>과 <그림 4>의 차이는 <그림 3>의 경우는 화면을 볼 수 없는 상황이므로 음성인식 다이얼링에서 '이름으로 걸기', '전화번호로 걸기'와 같이 음성 인식 기능 메뉴를 음성으로 선택 하도록 하지만, <그림 4>의 경우는 화면에 음성인식 기능 메뉴를 표시하여 사용자가 버튼으로 선택할 수 있도록 한 것이다. 또한 <그림 4>의 ⑥에서 사용자 발화의 인식 결과를 n-best로 보여 주고 선택할 수 있도록 한 것과 ⑦에서 사용자의 대답 이 발화 ("예")와 <통화> 버튼을 모두 이용할 수 있도록 한 것이 차이가 있다.

이러한 시나리오 설계를 한 다음에, 이루어져야 하는 초기 사용성 시험은 개발자 자신이나 대상 사용자 그룹을 대표할 수 있는 대표 사용자를 이용하여 실시한다. 하부 태스크에 대한 다수의 멀티모달 음성 대화 시나리오에 대하여, 대화 완성을 위한 사용자의 모달리티별 사용 회수, 자연성 및 편리성 등을 평가하여 최선의 시나리오를 선정하기 위한 사용성 평가를 수행하여야 한다. 이 단계에서의 사용성 평가는 시스템이 구현되지 않은 상태이므로 설계서를 이용한 피실험자와의 인터뷰를 통하여 시행한다. 인터뷰에서의 check list에는 다음과 같은 것이 반드시 포함되어야 한다.

- 시나리오 각 단계 (프롬프트)에서 출력 모달리티별로 사용자에게 제공되는 출력의 설계가 자연스러운가? 화면 출력의 경우, 시인성과 함께 표시한 문장에 대한 사용자의 인지 정확성 및 자연성을 시험하여야 하고, 음성 합성의 경우에는 합성 문장의 설계가 사용자가 시스템이 요구하는 것이 무엇인지를 쉽게 인지할 수 있도록 하였는가를 각 프롬프트에 대하여 시험한다.
- 시나리오의 각 단계에서 사용자가 할 수 있는 입력이 입력 모달리티별로 어렵지 않게 설계 되었는가? 입력하여야 하는 정보의 발화 자연성, 입력 복잡성 등을 각 사용자 입력 단계에서 시험한다.
- 설계한 시나리오에서 태스크 성공에 이르기까지 필요한 모달리티별 사용

회수와 사용자 입력 오류에 대한 대처 방안의 지원에 대해서 그 용이성 및 만족도를 물어본다.

<표 4> 모달리티별 사용성 요소의 시나리오 의존 관계

	Keypad+버튼	음성인식	화면 출력	음성합성
시나리오 설계	특정기능 사용 버튼의 선택	인식 분법 편리성	레이아웃 및 폰트, 색상 등	프롬프트 문장의 인지용이성
개발 환경 의존	크기, 버튼 연동 기능	개별 발화 인식을	화면 크기 등	음색만족도

<표 4>는 다이얼링에서 사용 가능한 모달리티별로 사용성에 영향을 미치는 요소들이 시나리오 설계에 의존적인 것과 개발환경에 의존적인 것으로 분류해 보았다. 여기서 개발환경에 의존적이라 함은 별도의 HW 또는 SW 개발 비용이 있을 경우 개선될 수 있는 요소로, 예를 들면, 개별 단어의 인식율은 추가적인 메모리와 비용이 비싼 인식기를 사용하여야 한다. 이러한 개발환경 의존적인 요소는 일반적으로 인터페이스 설계에서 생산 비용 등의 측면에서 고정되는 경우가 많아서 이러한 요소들에 의하여 제한되는 사용성에 대해서는 시나리오 설계로 개선될 수 없다. 따라서 시나리오 설계를 통하여 개발환경에 의존적이지 않은 사용성 요소를 개선하여야 하는 경우가 일반적이다. 4.1절에서 제시한 사용성 요소 중에서 ‘모달리티 입력 인식 정확성’이 개발환경 의존적인 요소이다.

4.2.3. WOZ 사용성 시험

초기 사용성 시험을 반복 수행하여 완성된 음성 기반 멀티모달 인터페이스는 먼저 WOZ 시험을 위한 형태로 구현한다. WOZ 시험을 위한 시스템은 실제 태스크를 시험자가 할 수 있도록 사용자와의 실제 인터페이스 부분만을 구현한다. 시험의 편의를 위하여 실제 개발 대상 하드웨어가 아닌 PC 등을 이용할 수도 있다. WOZ 시험에서는 대상 사용자 그룹에서 가능한 많은 수의 실제 사용자를 대상으로 한다. WOZ 사용성 시험에서는 대부분의 정성적 사용성 평가 척도를 적용하여 시험을 실시할 수 있다.

WOZ 시험에서 가장 중요한 것은 사용자의 수행 태스크 별 멀티모달리티의 사용 패턴 조사(사용 패턴에는 사용자의 멀티모달 음성 발화문 수집과 [10]에서와 같이 멀티모달 입력에서 모달리티 사이의 관계를 포함한다.)와 사용자가 일으킬 수 있는 오류 입력의 종류이다. 사용 패턴은 사용자 입력의 자연성, 멀티모달 대화 주도권의 적절성, 사용자별 대처 능력, 멀티모달 대화 구성의 자연성을 높이기 위한 기초 자료이다. 또한 사용자 오류 입력의 사전 조사는 사용자 오류 대처 능력을 제고하기 위하여 필요하다. WOZ 사용성 시험에서는 또한 시험자(wizard)가

<표 5> 음성과 터치 제스처 기반 멀티모달 인터페이스 사용성 평가를 위한 매트릭스

요소	척도	수행단위 및 방법	수행단계	피드백단계	정성/정량	
효율성	모달리티 입력 인식율	음성인식율	태스크의 각 음성인식 단계의 평균	알파/베타	알파/베타	정량
		제스처인식율	사용 제스처 별 인식율 평균	알파/베타	알파/베타	정량
		퓨전인식율	태스크의 각 퓨전 단계 평균	알파/베타	알파/베타	정량
	경과시간	평균 시스템 응답시간	사용자 입력에 대한 응답 시간 평균	알파/베타	알파/베타	정량
	태스크 성공률		태스크 단위별 성공률 평균	WOZ/알파/베타	알파/베타	정량
	태스크 완료 시간		태스크 별 완료 시간 평균	알파/베타	알파/베타	정량
	태스크 성공을 위한 인터랙션 회수 평균	태스크 별 회수 평균	초기/WOZ/알파/베타	WOZ/알파/베타	정량	
학습 용이성	도움말	시의성	도움말 제공 단계별 선호도	초기/WOZ/알파/베타	WOZ/알파/베타	정성
		충분성	도움말 별 제공 정보의 충분성 평균	초기/WOZ/알파/베타	WOZ/알파/베타	정성
		접근용이성	도움말 접근을 위한 인터랙션 회수	초기/WOZ/알파/베타	WOZ/알파/베타	정량
	피드백정보의 적절성		시나리오 각 단계에서 시스템 제공 정보에 대한 만족도 평균	초기/WOZ/알파/베타	WOZ/알파/베타	정성
	프롬프트 인지 용이성		프롬프트 별 내용의 인지 정도 평가	초기/WOZ/알파/베타	WOZ/알파/베타	정성
성과 시간	평균 사용자 응답시간	사용자가 입력하는데 걸린 시간 평균	WOZ/알파/베타	알파/베타	정량	
기억 용이성	동일 태스크의 학습성	동일 태스크 수행 시간 감소 비율	동일 태스크의 1회차 및 2회차 수행시간 비교	알파/베타	알파/베타	정량
	유사 태스크 학습성 (UI의 일관성)	유사 태스크 수행 시간 감소 비율	최초 태스크 A의 수행 시간과 태스크 B 수행 후 A 수행 시간 비교	알파/베타	알파/베타	정량
오류 대처	사용자 입력 오류 회수		태스크 각 입력 단계에서 오류 입력 회수	WOZ/알파/베타	알파/베타	정량
	사용자 입력 오류 회복 비율		오류 입력 별	알파/베타	알파/베타	정량
	입력 오류 회복에 필요한 인터랙션 회수 평균		회복 오류입력 별	알파/베타	알파/베타	정량
	off-talk 회수		태스크 각 단계에서 유도하지 않은 사용자 입력 회수	WOZ/알파/베타	알파/베타	정량
	무 응답에 대한 대처 만족도		무 응답/사용자 별 만족도	알파/베타	알파/베타	정성
만족도	입력 자연성	사용자 발화 (인식문법)자연성	인식 단계별 인식 문법의 자연성	초기/WOZ/알파/베타	WOZ/알파/베타	정성
		제스처 자연성	제스처 별 만족도	초기/WOZ/알파/베타	WOZ/알파/베타	정성
		발화 및 제스처 동시 입력 자연성	동시 입력 별 선호도	WOZ/알파/베타	WOZ/알파/베타	정성
		입력 모달리티 선호도	태스크/환경/사용자 별 선호도	초기/WOZ/알파/베타	WOZ/알파/베타	정성
	출력 자연성	합성 음성 만족도	합성 음성 선호도	알파/베타	알파/베타	정성
		GUI 시인성	화면 출력에 대한 선호도	초기/WOZ/알파/베타	WOZ/알파/베타	정성
	대화 자연성	동시 출력 모달리티 선호도	태스크/환경/사용자 별 선호도	초기/WOZ/알파/베타	WOZ/알파/베타	정성
			태스크 별 사용자 선호도 평균	초기/WOZ/알파/베타	WOZ/알파/베타	정성
사용자별 대처 능력	대화 주도권 적절성	태스크/사용자 군별 선호도 평균	WOZ/알파/베타	알파/베타	정성	
	사용자 군별 만족도	태스크/사용자 군별 만족도 평균	WOZ/알파/베타	알파/베타	정성	
환경 대처능력	사용 환경 별 (이동, 운전, 가정, 사무실) 만족도	태스크/ 환경별 만족도 평균	WOZ/알파/베타	알파/베타	정성	

사용자에게 응답하는 정보의 내용을 달리하면서 사용자의 반응을 조사하여 시스템의 피드백 적절성을 높일 수 있는 설계가 이루어 질 수 있도록 보완할 수 있다.

<표 5>는 음성인식/합성과 터치 제스처 그리고 디스플레이를 사용하는 멀티모달 인터페이스의 사용성 평가를 위해 제안하는 매트릭스이다. 여기서 요소는 2.1 절에서 설명한 J. Nielsen의 정의에 따른 사용성 요소이며, 척도는 각 요소를 평가하기 위한 세부 매트릭스를 의미한다.

수행 단위 및 방법은 각 척도의 수행 단위와 측정 방법을 수행 단계는 각 척도가 측정될 수 있는 설계 단계를 뜻한다. 피드백 단계는 각 척도의 결과를 피드백할 수 있는 설계 단계를 의미한다. 예를 들어 ‘사용자 입력 오류 회수’는 WOZ 사용성 시험 단계에서부터 적용 가능하고, 개발 환경에 의존적인 척도인 모달리티

인식률은 알파/베타 사용성 평가 단계에서 가능한 것으로 표시하였다. 이러한 척도는 인터페이스 시나리오의 설계에 좌우 되지 않고, 알파 단계에서 이러한 척도가 현저히 낮다면, 보다 나은 질의 인식기 합성기를 추가 비용으로 채용하여야 한다.

<표 5>와 같은 척도를 이용하여 PROMISE에서 개발한 평가 프레임워크에 적용하여 이전 시스템의 평가 결과에 비하여 시험 결과 반영 후의 평가가 점차 좋아지는 것을 확인하고, 평가 결과가 만족할 만한 수준에 도달하면 시제품 개발로 나아간다. 물론, <표 5>가 모든 멀티모달 인터페이스에서 적용 가능한 평가 매트릭스는 아니다. 예를 들어, 출력 모달리티로 햅틱(haptic)이 추가된다면, 출력 자연성 척도의 세부 척도로 햅틱 출력의 선호도와 같은 항목이 추가되어야 한다. 그러나 제안하는 매트릭스에서 요소와 대분류 척도는 다른 멀티모달 인터페이스에도 공히 적용할 수 있는 형태이며, 단지 세부 척도에서 사용 가능한 모달리티의 차이에 따라 추가하거나 수정될 수 있다.

4.2.4. 알파/베타 시스템 사용성 평가

WOZ 사용성 시험에서 보완된 인터페이스는 실제 대상 하드웨어 및 소프트웨어에 구현하여 시제품을 만든다. 알파 시제품은 WOZ 시험용 시스템에서 참여한 피시험자와 개발자 그룹을 대상으로 하는 사용성 평가 시험을 하는 것이며 베타 시제품은 실제 사용자 그룹에 무작위로 배포하여 시험한다.

알파/베타 시스템에 대한 사용성 평가는 모든 정량적인 평가 척도를 적용한 평가를 시행한다. 이 시점에서는 인식 시간 등과 같은 경과 시간 관련 평가 척도, 실제 사용상에서 발생하는 오류 패턴, 실제 사용 상황에서의 멀티모달 발화 회수, 오류율 등이 측정될 수 있다. 물론, 알파/베타 시스템의 경우에도 시험 후 반영 및 재출시를 반복하는 것이 바람직하다.

이러한 과정을 거쳐 개발자가 원하는 정도의 사용성 평가 결과가 나온 음성 기반 멀티모달 인터페이스 시스템을 정식으로 출시한다. 정식 출시 후에도 사용자의 불편 사항을 지속적으로 모니터 하고, 이를 반영하여 수정하는 노력을 계속한다. <표 5>의 평가 매트릭스에서 수행단계가 알파/베타인 척도들이 이 단계에서 사용성 평가를 수행할 수 있는 것이다. 즉, 모든 가능한 평가 척도가 사용되어야 한다.

5. 결 론

음성 인식, 음성 합성, 제스처 인식, 멀티모달 퓨전과 같은 멀티모달 인터페이

스를 위한 기반 기술은 나름대로 기술적인 제약을 현재 가지고 있다. 예를 들면, 음성 인식의 경우, 인식 대상 어휘의 수가 늘어날수록, 사용환경에 잡음이 많을수록 인식 성능이 저하된다. 이러한 기술적 제한을 감안하더라도 음성기반 멀티모달 인터페이스를 상용화할 수 있는 응용 분야가 존재한다. 그러나 음성기반 멀티모달 인터페이스의 응용 분야 확대를 위해서는 사용성 제고가 필수이다.

지금까지, 일반적인 제품 설계에 있어 적용되어온 사용성과 사용자 중심 설계 및 음성 인터페이스와 음성기반 멀티모달 인터페이스를 위한 범용 평가 척도로서 PARADISE[5]와 PROMISE[6]에 대하여 소개하였다. 범용 평가 척도는 기존의 특정 분야에서 개발된 특정한 인터페이스를 위해 개발된 다양한 개별 평가 척도를 다양한 분야의 다양한 형태의 인터페이스에 대하여 적용할 수 있도록 통합한 형태의 평가 방법을 말한다.

일반 제품을 위한 사용자 중심 설계 과정은 요구 분석의 단계에서부터 사용성에 초점을 맞추는 반복적인 사용성 평가가 중심이다. 이를 음성 기반 멀티모달 인터페이스 개발에 적용하여 음성기반 멀티모달 인터페이스를 위한 사용성과 사용자 중심 설계의 가이드라인을 제시하였다. 요구분석과 초기 사용성 시험에서 도출하여야 하는 사용성 관련 정보와 초기 설계를 휴대전화에서 멀티모달 음성인식 다이얼링을 사례로 예시하였고, 음성과 제스처를 활용하는 멀티모달 인터페이스 사용성 평가 매트릭스를 제안하였다.

실제로, 영어권이나 독일 등 유럽에서의 음성기반 멀티모달 인터페이스 개발은 사용자 중심 설계로 이루어지는 경우가 많으며, 인터페이스 개발의 비용 중에서 가장 많은 비율이 사용성 평가에 투자하고 있다. 물론 국내에서 개발되고 있는 멀티모달 인터페이스도 나름대로의 사용성 평가를 시행하고 있으나, 그 규모나 체계에서 멀티모달 인터페이스의 확산에 도움을 줄 수준에는 크게 못 미치고 있다.

사용성 제고를 염두에 둔 요구분석단계에서의 사용자 분석, 설계서를 기반으로 한 초기 사용성 평가를 위한 구체적인 방법론의 개발, WOZ 시험을 위한 멀티모달 인터페이스 구현 프레임워크의 개발이 다음 과제이다. 4장에서 제시한 멀티모달 사용성 요소 각각에 대한 정성/정량 사용성 평가 척도의 개발은 시작단계에 있다.

참고문헌 [11]에서는 설계 단계에 적용할 수 있는 사용성 예측 지표를 개발하였다. LCD 모니터의 메뉴에 대하여 검증하였다. 사용성 예측 지표는 설계한 시나리오만을 이용하여 정량적 수치로 시나리오의 사용성을 평가하고자 하는 것으로 실제로 사용성 예측 지표에서 높게 나오는 시나리오가 구현한 다음에 실 사용자의 사용성과 일치함을 검증하여야 한다. 향후, 음성기반 멀티모달 인터페이스 설계를 위한 사용성 예측 지표의 개발도 또 다른 연구 과제이다.

참 고 문 헌

- [1] 박성수, 안세열, 김원우, 구명완, “모바일 환경을 위한 멀티모달 미들웨어의 설계 및 구현”, *말소리*, 제60호, pp. 125-144, 2006.
- [2] L. M. Reeves, J. Lai, J. A. Larson, S. Oviatt, T. S. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.-C. Martin, M. McTear, T. Raman, K. M. Stanney, H. Su, Q. Wang, “Guidelines for multimodal user interface design”, *Communications of ACM*, Vol. 47, No. 1, pp. 57-59, 2004.
- [3] 홍기형, “음성기반 멀티모달 인터페이스 및 표준”, *말소리*, 제51호, pp. 117-135, 2004.
- [4] C. M. Barnum, *Usability Test and Research*, Longman Publishers, 2002.
- [5] M. Walker, D. Litman, C. Kamm, A. Abella, “PARADISE: A framework for evaluating spoken dialogue agents”, *Proc. 35th Annual Meeting of the Association of Computational Linguistics*, pp. 271-280, 1997.
- [6] N. Beringer, U. Kartal, K. Louka, F. Schiel, U. Turk, “PROMISE: A procedure for multimodal interactive system evaluation”, *Proc. Workshop on Multimodal Resources and Multimodal Systems Evaluation*, pp. 77-80, 2002.
- [7] W. Wahlster, “Dialogue systems go multimodal: The SmartKom experience”, *SmartKom: Foundations of Multimodal Dialogue Systems*, W. Wahlster (eds.), pp. 3-27, Springer, 2006.
- [8] J. Nielsen, *Usability Engineering*, Academic Press, 1993.
- [9] L. Dybkjaer, N. O. Bersen, “Overview of evaluation and usability”, *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*, W. Minker, D. Buhler, L. Dybkaer (eds.), pp. 221-246, Springer, 2005.
- [10] 김지영, 이경님, 홍기형, “TV 가이드 영역에서의 음성기반 멀티모달 사용유형 분석”, *말소리*, 제58호, pp. 101-117, 2006.
- [11] 윤영식, 윤완철, “UI 설계-평가 프로세스 지원을 위한 인터랙션 설계시안의 사용성 예측 지표 개발”, *대한인간공학회 2006 추계학술대회 논문집*, pp. 353-358, 2006.

접수일자: 2008년 8월 9일

게재결정: 2008년 9월 22일

▶ 홍기형 (Ki-Hyung Hong) : 교신저자
 주소: 136-742 서울 성북구 동선동3가 249-1
 소속: 성신여자대학교 미디어정보학부
 전화: 02) 920-7525
 E-mail: khhong@sungshin.ac.kr