

가전제품의 VUI 가이드라인에 대한 연구

A Study on Voice User Interface for Domestic Appliance

채행식 Haeng Suk Chae*, 홍지영 Ji Young Hong*, 이주환 Ju Hwan Lee *, 전명훈 Myoung Hoon Jeon **, 김민선 Min Sun Kim **, 허우범 Ubeom Heo **, 안정희 Jeong Hee Ahn **, 한광희 Kwang Hee Han*

*연세대학교 인지공학연구소, **LG 전자 디자인 경영센터 Corporate 디자인실 UXD Gr.

요약 음성 사용자 인터페이스(Voice User Interface, 이하 VUI) 는 음성을 매개로 일어나는 인간과 기계 간 인터페이스를 뜻한다. 음성 인식율의 향상과 음성 재생 장치의 발달에 힘입어 최근 들어 휴대폰과 카 네비게이션 시스템에 주로 적용되고 있다. 최근 이러한 경향은 A/V 시스템 등 가전제품(Domestic Appliance) 에도 확대되고 있는데 본 연구에서는 사용자와 필수적이고 빈번한 상호작용이 일어나는 백색 가전을 대상으로 사용자를 만족시키는 음성 인터페이스의 주요 속성 중 음성 생성(Speech Generation)과 관련된 음성 표현을 중점 연구하였다. 연구방법으로 먼저 주부들이 느끼는 가전에서의 문제점과 VUI 로서 해결가능성에 대하여 F.G.I. 를 통하여 조사하여 주요 이슈를 도출하고, 대표적 백색 가전인 에어컨, 세탁기, 김치냉장고, 냉장고, 식기세척기, 오븐레인지 등 6 개 제품에 대하여 음성의 물리적 특성, 내용적 특성, 기능에 따른 배치에 대하여 조건에 따라 다양한 프로토타입을 제작한 후 실제 환경과 유사한 실험실 상황에서 사용자의 선호도, 적합도 및 수행을 측정하였다. 연구 결과 각 이슈에 따라 가전제품에 적합한 VUI 가이드라인 특성을 찾아내었다.

핵심어: Voice User Interface, HCI, Design, UI

1. 서론

가전제품은 시청각 미디어를 활용할 수 있는 A/V 가전과 비교적 사용이 용이한 세탁기, 냉장고, 식기세척기, 에어컨, 오븐레인지와 같은 백색가전으로 나뉘어 질 수 있다. A/V 제품은 기능이 다양하고 사용하기가 복잡하여 사용자 인터페이스에 관한 연구가 많이 이루어졌다. 최근에 백색가전도 많은 기능이 탑재됨에 따라 사용자 인터페이스의 중요성이 커지고 있고 사용자들의 요구를 반영한 인터페이스의 개선은 기술의 발달과 더불어 중요하게 되었다. 백색가전의 사용자 인터페이스를 향상시킬 수 있는 것 중 음성 인터페이스는 중요한 역할을 할 것이다. 본 연구는 백색가전에서의 음성 인터페이스의 향상을 위한 연구를 수행하였다. 특히 음성 인터페이스에서 백색가전의 가장 큰 특징 중 조작음 형성과 경고, 알림에 대한 부분의 연구를 진행하였다. 가전제품의 조작과정에서 제품의 사용성, 학습용이성 등이 저하되어 제품에 탑재된 유용한 기능을 배제하는 상황으로 갈 수 있다. 이러한 조작과정에서 음성 인터페이스를 통해 조작의 흐름을 가시화한다면 사용편리성의 향상이 제품의 가치의 상승효과를 가져올 수 있다[3][4].

본 연구는 LG전자 가전제품 VUI 가이드라인 개발 연구(과제번호 2006-8-1403)에 의해 연구비 지원을 받아 수행되었음

음성 인터페이스는 음성인식, 음성합성 등과 관련된 기술에 관련된 부분이 중점적으로 연구가 되었는데 본 연구에서는 백색가전의 특성을 반영한 음성표현 부분에 관한 실증적인 연구를 진행하였고 연구결과는 실제 제품에 적용 가능하도록 하였다. 이 후 백색가전에서 사용자 중심의 음성인식 인터페이스 연구의 바탕이 되는 연구를 수행하였다.

음성 인터페이스는 음성정보를 이용하여 상호작용하는 휴먼-컴퓨터 인터페이스 기술이다[4]. 음성기술의 영역은 다양해지고 TTS(text to speech), STT(speech to text) 등의 기술 역시 점차 정확율이 개선되어 전문화 기기뿐 아니라 TV의 EPG(Electronic Programmable Guide) 등 일상적인 가전제품에까지 사용이 확대되고 있는 추세이다. 에어컨, 세탁기 등 일상적인 가전제품은 사용자가 매일같이 사용하는 제품이고 주부 뿐 아니라 가정 내의 여러 구성원들이 사용하게 된다. 또한 점차 강력해지고 복잡해지는 가전제품의 기능들을 효율적으로 사용하기 위해 단순한 신호음과 간략한 액정 디스플레이뿐 아니라 정교한 컬러 디스플레이와 사용성과 심미성을 만족시키는 AUI(Auditory User Interface)를 적용한 제품이 출현하고 있다. 음성 인터페이스는 GUI(Graphic User Interface) 나 AUI 보다 구체적 의미를 직접 전달할 수 있기 때문에 효율적이고 만족스럽게 사용할 수 있는 가능성을 제공한다[5].

음성 인터페이스는 대체로 음성 인식 인터페이스와 동의어로 인식되는 경향이 있지만 기계로부터 표현되는 음성 자체

에 대한 고려 또한 사용성 및 만족도를 높이기 위한 매우 중요한 특성이다[1]. 본 연구에서는 음성 사용자 인터페이스에서 가전제품의 사용 만족도를 높이기 위한 요인을 조사하고 각 요인에 따른 프로토타입을 제작하여 적절한 조건을 탐색하였다.

기존의 음성 인터페이스 가이드라인에 관련된 연구 중 명령어(command vocabulary) 설계, 명령구문(command syntax)의 선택, 피드백 등의 연구는 있었다[2]. 하지만 음성 인터페이스와 관련하여 특정 제품군을 대상으로 실제 사용자들이 경험에 볼 수 있도록 프로토타입을 제작하고 실험 결과와 사용자 의견을 가진 제품에 적용 가능한 연구의 의미가 있다 할 수 있다[6].

본 연구는 가전제품에 적용 가능한 사용자 중심의 음성 인터페이스 가이드라인 개발을 그 목적으로 한다. 각 가전제품의 기능을 파악하고 이에 적합한 인지적이고 감성적인 인터페이스를 적용하고자 한다. 선호되는 음성의 물리적 특성에 대한 지침을 마련하고 신호음과 음성의 기능에 따른 할당 지침을 마련한다. 본 연구는 가전제품의 사용과정에서 소비자의 멘탈 모델을 고려하고 인간요인의 관점에서 주의와 의미를 고려하여 음성 인터페이스 연구를 진행한다. 인간의 정보 처리 과정에서의 지각, 이해, 판단의 과정을 적용한다[7][8].

구체적으로는 가전제품의 음성 인터페이스에서의 일반적인 이슈에 대한 언급과 기능별 분석을 진행하고 사전 포커스그룹인터뷰(FGI)를 통해 실험설계를 구체화 하였다. 또한 실험에 필요한 사운드 녹음을 포함한 프로토타입 작업을 진행하였다. 이러한 실험결과를 바탕으로 가전제품의 음성 인터페이스 가이드라인을 확정하고 최종적으로 최적의 음성 사용자 인터페이스 프로토타입 제작을 하였다. 본 가이드라인과 최적의 프로토타입은 실제 제품의 제작에 반영되었다.

2. 연구방법

2.1 기능분석

본 연구는 여섯 가전제품에 음성을 적용하는데 있어 제품과 기능별로 적합한 음성의 특성 발견 및 가전 제품에서 음성 사용에 대한 가이드라인을 위하여 시행되었다. 이에 제품 및 기능별로 적합한 음성에 대한 조사 및 실험을 시행하기에 앞서, 각 제품별로 어떠한 기능들이 제공되고 있는지에 대한 조사가 요구되었다.

냉장고, 세탁기, 김치냉장고, 식기세척기, 오븐, 에어컨 등 여섯 가전제품은 각기 상이한 기능들을 수행하고 있으나, 각 제품들의 기능들을 공통된 범주에 따라 구분하는 것이 기능별, 제품별로 적합한 음성의 속성에 대한 효율적인 guideline을 작성케 할 것이다.

더욱이 가전 제품들은 일상생활에서 주로 주부들에 의하여 사용되며 주방, 거실, 그리고 다용도실이라는 비교적 근접한 공간에서 사용된다는 공통점을 지니고 있으므로, 이상의 기능들은 각각의 속성에 따라 몇 개의 공통된 범주로 나누어질 수 있다. 이에 아래에서는 각 제품의 기능들을 일곱 범주로 나누고, 각 범주들이 어떤 속성을 지니고 있으며, 제품별 기능들이 어떠한 범주에 속하는지에 대하여 연구하였다.

2.2 사전조사

본 실험에 들어가기 이전에 4명의 주부를 대상으로 사전

연구인 포커스 그룹 인터뷰 (Focus-Group Interview, 이하 F.G.I)를 통해 가전제품 사용자 인터페이스에 불편했던 점과 개선방향, 그리고 음성 인터페이스가 적용되면 편리한 부분에 대하여 의견을 얻었다. 대표적 가전제품으로 오븐, 세탁기, 식기세척기, 냉장고, 김치냉장고의 6개 제품을 대상으로 하였다.

사전 F.G.I 결과 음성의 물리적 특성으로 성별, 톤, 억양, 언어적 특성으로 경어체, 서술어미의 유무, 기능적 할당 측면에서 신호음과 음성의 전반적 할당 방법, 조작시의 음성표현과 설정 완료시의 음성 표현, 경고/알림 기능에서의 음성표현 방법의 중요성이라는 주요 관심 요인을 얻었다.



그림 1. F.G.I 진행장면

본 실험에서는 사전 F.G.I 에서 조사된 요인에 따라 6개 가전제품의 형태 및 기능을 모사하여 터치스크린으로 동작하는 음성 사용자 인터페이스 (Voice User Interface) 프로토타입을 제작하였다.

2.3 프로토타입 제작

본 실험 전 사전 F.G.I 분석결과와 위 분석된 제품별 기능 분석을 바탕으로 실험에 필요한 프로토타입을 제작하였다.

본 실험에 사용된 프로토타입은 실제 제품의 인터페이스를 반영하고 기기 조작에 따른 음성 피드백을 제공하여 줄 수 있도록 제작되었다. 이는 일종의 고품상도 프로토타입(High-Fidelity Prototype)으로서, 그림이나 외형 등과같이 실제로는 조작할 수 없는 저해상도 프로토타입(Low-Fidelity Prototype)과는 달리 본 연구에서 다루고자 하는 음성 피드백의 효과를 단순히 들어보게만 하는 것이 아니라 실제 제품을 이용할 때와 마찬가지로 조작을 하면서 들을 수 있도록 하였다.

본 실험에 사용된 프로토타입은 Macromedia Flash Professional 8을 이용하여 5개 각 제품의 외형을 갖추고, 조작할 수 있는 인터페이스를 제공하여 본 실험 과제 수행시 조작할 때마다 음성 피드백이 제공되도록 하였다. 또한 프로토타입은 제작 원칙인 사용성(usability), 신뢰성(reliability), 유지가능성(maintainability), 제품간 일관성(consistency)을 반영하였다. 특히 본 연구의 프로토타입은 제품 UI가 아닌 음성 피드백이므로 최대한 조작이 간편하고 사용하기 어렵지 않도록 설계되었다.

2.4 실험절차 및 내용

34명의 주부를 대상으로 각 요인에 대해 5개 세션으로 나누어 선호도, 적합도 및 수행 오류, 반응시간을 분석하였다. 1세션에서는 음성의 물리적 특성에 대하여 남성과 여성, 높은 톤과 낮은 톤, 보통 억양과 강한 억양에 대하여 실험하였고, 2세션에서는 언어의 내용으로 경어체와 관련된 어투에 관해 실험하였고, 3세션에서는 음성표현의 간결성, 4세션에서

는 경고 및 알림시의 사용자의 수행, 5세션에는 프로토타입의 전반적 비교를 실험하였다. 특히 경고 및 알림에서의 선호도와 음성 할당 세션에서는 주부들이 TV의 소음에 노출된 상황에서 효과적으로 주의를 전환하는지의 여부와 안내메시지를 듣고 정확하게 반응할 수 있는지의 여부를 실험을 통해 조사하였다.

실험은 Usability Testing Room(이하 UT room)에서 실시하였으며, UT room에는 실험 참가자 한 명에, 실험을 진행할 진행자(moderator) 한 명, 실험 진행 보조자 한 명이 배석하였다. 또한 이들의 대화내용을 기록하는 스크립터, 진행을 체크하는 관찰자(Observer)가 관찰실(Observation Room)에서 간접적으로 실험에 참여하였다. 이 관찰실(Observation Room)에서는 UT room에 설치된 카메라와 마이크를 통해 실험 진행상황이 실시간으로 전송되었다.

2.4.1 실험절차

실험은 사전조사 및 연습시행 단계, 본 실험 단계, 사후 인터뷰 및 설문조사 단계의 순서로 실시되었으며, 사전조사 및 연습시행 단계는 참여자로 하여금 실험에 사용될 터치스크린의 조작법에 익숙해지도록 하기 위해 실시되었다. 여기서 사용되는 터치스크린과 연습 시행용 프로토타입은 실제 실험 진행에 이용되는 것과 동일하며, 실험에 쓰이는 모든 프로토타입의 모든 조건을 한 번씩 조작해보므로써 사전 학습효과를 최소화하였다. 이 연습시행은 연습시행 진행자에 의해 진행되었으며, 참여자가 궁금해 하는 사항에 관한 답변도 병행하였다. 연습시행을 마치면 기본 인적 사항을 조사하는 사전 설문지를 작성한 다음, 진행자의 안내를 통해 UT room에 있는 실험 진행자(moderator)에게 인계되었다.

2.4.2 실험내용

실험 참여자와 진행자는 터치스크린이 놓여진 탁자를 앞에 두고 나란히 앉아 실험을 진행하였다. 실험이 진행되는 동안에는 관찰에 이용되는 카메라와는 별도로 캠코더를 통해 터치스크린의 내용을 촬영하였다.

실험은 진행자가 개별 기기의 프로토타입을 보여주고, 이에 맞는 과제를 제시하면, 참가자는 사용된 목소리에 유의하면서 이 과제에 적혀있는 대로 프로토타입을 조작해보았다. 각 조건에 대한 선호도와 적합도를 7점 척도로 체크 하였다. 이때 참가자에게는 매 조건을 실시한 후 다음 조건에 대한 선호도와 적합도를 체크하면서 이전에 체크한 사항을 수정할 수 있도록 하였다.

3. 실험결과

3.1 음성의 물리적 특성

음성을 통해 사용자에게 정보를 전달하는 인터페이스 설계에서 음성 그 자체의 물리적 특성이 중요한 변수가 될 것이다. 즉 음성의 물리적 특성에 따라 사용자에게 전달되는 정보의 인지적 효율성과 감성적 만족도가 다를 수 있기 때문에 가전제품에 어떤 물리적 특성의 음성을 음성 사용자 인터페이스(VUI)로 적용해야 하는지에 대한 가이드가 중요하다. 이를 위해 본 연구에서는 크게 세 가지의 음성 특성을 음성의 성별(남성 vs. 여성), 음성의 음높이(tone: common tone vs. high tone), 그리고 음성의 억양(intonation: flat vs. dynamic)으로 나눠 비교 분석하였다.

3.1.1 음성의 성별

연구에 사용된 여성과 남성의 목소리에 대한 사용자의 선호도와 적합도는 그래프에 나타난 바와 같이 여성 음성이 더욱 긍정적인 평가를 받았다. 특히 남성의 목소리는 평가의 수치가 낮을 뿐 아니라 부정적인 평가를 받았는데, 사용자들은 일반적인 전자제품에 적용된 음성이 여성이기 때문에 남성 목소리가 낯설고, 어색하게 느껴진다고 반응하였다. 이런 측면에서 남성 목소리를 가전제품의 음성 인터페이스로의 적용하는 것은 사용자에게 정서적으로 부담이 될 수 있으므로 여성 목소리를 적용하는 것이 더 적합한 것을 알 수 있다.

3.1.2 음성의 음높이

일반적인 여성의 음높이와 상대적으로 높은 음성을 지닌 여성의 목소리에 대한 사용자의 선호도와 적합도는 보통 수준의 음높이를 지닌 여성의 음성이 상대적으로 더 높은 평가를 받았다. 높은 톤의 목소리에 비해 보통의 목소리에 대한 긍정적인 평가는 안정감 있고 신뢰감을 주는 편안한 목소리로 인식되어 들떠있거나 경박한 듯한 높은 목소리보다 선호되는 것으로 나타났다. 그러나 보통 톤의 경우는 약간 기계적이고 딱딱한 느낌을 지니고, 반대로 높은 톤은 발랄하고 상냥하게 느껴져 음성 인터페이스의 적용 목적에 따라 음높이를 다르게 선택하는 것도 필요할 수 있다. 결론적으로는 일반적인 경우에는 무난하고 신뢰를 주는 안정된 목소리가 적절하다.

3.1.3 음성의 억양

목소리가 가질 수 있는 또 다른 변수로는 발생하는 방법에 따른 억양의 차이가 있다. 억양은 다양한 구분이 가능하지만 본 연구에서는 가전제품에 적용될 수 있는 일반적인 억양과 이보다는 약간 역동적인(dynamic) 억양의 음성 인터페이스를 비교 평가하였다. 그 결과, 과장된 억양의 표현이 아니라면 큰 차이가 없는 것을 알 수 있었다. 즉 보통의 억양과 보다 활동적인 억양에서 사용자들은 큰 차이를 느끼지 못하였다. 그러나 음성의 억양에서도 음높이와 마찬가지로 음성 인터페이스 적용 분야와 목적에 따라 편안하고 안정된 일반적 억양을 사용하거나 명량하고 자신감 있게 느껴지는 역동적인 억양을 사용할 필요가 있을 것이다. 결론적으로는 일반적인 억양으로 안정되고 편안함을 표현하되 너무 사무적이지 않도록 유의하는 것이 중요하다.

3.1.4 음성의 물리적 특성 종합

위와 같이 세가지 음성의 물리적 특성들에 대한 비교 분석의 결과를 종합하여 보면, 남성보다는 여성의 안정적인 톤과 억양의 목소리가 가전제품의 음성 사용자 인터페이스로 사용되는 것이 적합하다. 그러나 친근하고 상큼한 느낌을 살려 딱딱하고 사무적이지 않도록 적용해야 한다

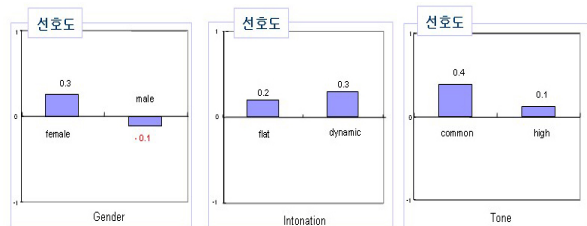


그림 2. 음성의 물리적 특성에 관한 선호도

3.2 어투(합쇼체/해요체)

가전제품에서 음성을 사용할 때 고려해야 할 한국어의 특징 중에 하나는 높임법을 사용한다는 것이다. 가장 빈번하게 사용되는 상대높임법의 격식체를 살펴보면 해라체, 하계체, 하오체, 합쇼체가 있고 비격식체에는 해요체와 해체가 있다. 이중 일상대화나 방송언어로 가장 많이 사용되는 것은 격식체인 합쇼체와 비격식체인 해요체이다. 따라서 공식적이고 격식 있는 느낌을 주는 합쇼체 (예. “문을 닫아 주십시오”)와 친근하고 부드러운 감성을 전달하는 해요체 (예. “문을 닫아 주세요”) 중 어떤 어투가 더 선호되고 적절하게 평가되는가를 조사하여 음성 인터페이스에 적용하는 것이 필요하다.

3.2.1 일반조작

본 세션에서는 가전제품에서 버튼의 일반 조작 음으로 제공되는 음성이 합쇼체로 제공될 때와 해요체로 제공될 때를 나누고 사용자가 어떤 경우를 더 선호하고 어떤 느낌을 갖는지를 알아보기 위해 동일한 멘트를 합쇼체와 해요체 두 가지로 들려주고 선호도와 적합도를 평정 받았다.

실험 결과 일반 조작 음에서 합쇼체를 사용하는 것은 사용자들에게 정중하고 격식을 유지하는 느낌을 주며 신뢰를 주고 사용자들이 매우 받는 것 같은 긍정적인 느낌을 주었다. 또한 가전제품이 아닌 일반적인 음성안내의 경우에 주로 합쇼체가 사용되어왔기 때문에 사용자들은 합쇼체에 더 익숙하다는 느낌을 받았다. 반면 명령조로 말하는 것 같아 부정적인 느낌을 주어 비 선호된다고 보고한 사용자도 있었다.

해요체의 경우에는 친절하고, 편안하며, 부드러운 느낌을 주는 반면에 일상에서의 안내음성과는 다른 어투를 사용해 어색한 느낌을 주며, 정중하게 예의를 갖춰 말하는 느낌이 들지 않고 오히려 낮춤말 같다는 부정적인 느낌들이 보고되었다.

전체적인 실험결과를 보면, 합쇼체가 해요체보다 더 정중하고 신뢰를 주며 익숙하기 때문에 일반 조작 시 제공되는 음성의 어투로 더 선호되고 적절하다고 평가되었다. 따라서 일반 조작음으로 음성을 제공할 때에는 사용자들이 선호하고 적절하다고 평가한 합쇼체를 사용하는 것이 바람직하다.

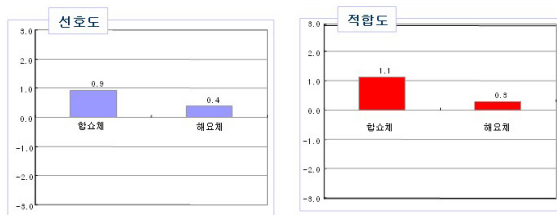


그림 3. 일반조작시 경어체 선호도 및 적합도

3.2.2 경고/알림

본 세션은 가전제품의 경고 또는 알림 멘트를 음성으로 알려줄 때 합쇼체가 더 적절하게 느껴지고 선호되는가, 아니면 해요체가 더 적절하게 느껴지고 선호되는가를 알아보기 위해, 사용자들이 직접 조작해본 직후에 선호도와 적합도를 평정하여 보고하는 방법으로 실시되었다. 일반 조작 음과는 달리 경고 또는 알림 멘트는 다른 곳에 주의를 두고 있는 상황에서 주의를 환기하고, 기기의 상태 또는 취해야 할 행동에 대한 메시지를 전달한다는 특수성이 있다. 동일한 경고와

알림 멘트를 합쇼체와 해요체로 들려주고 선호도와 적합도를 평정 받은 결과, 전체적으로 합쇼체가 해요체 보다 선호되고 적절한 것으로 드러났다. 경고나 알림 멘트의 경우 멘트를 주는 행위자체가 경각심, 놀람, 짜증과 같은 부정적인 정서를 유발할 수 있다. 따라서 어투를 선택할 때에도 이를 고려하여 설계해야 한다.

합쇼체를 선호하고 적절하다고 평가하는 대다수의 사용자들은 간결한 느낌을 주며 귀에 더 잘 들어온다고 보고하였고, 반대로 마치 명령을 받는 느낌이 들어 합쇼체를 비선호하는 사용자들도 있었다. 또한 배려를 받는 느낌이고, 편안한 느낌이 들어 해요체를 선호하고 적절하다고 생각하는 사용자들이 있는 반면, 대부분의 사용자들은 해요체가 너무 친숙한 느낌을 주어 듣기 “기분스럽고”, “방정맞고”, “가볍다”는 생각이 들어 비선호되고 적절하지 못하다고 생각하였다. 경고 멘트 자체에 대해서는 음성과 함께 소리가 제시되어 주의를 더 환기시키는 기능이 필요하다는 의견과 제시되는 음성의 속도가 더 빠르고 톤이 높아서 더 긴박한 느낌을 주어야 한다는 의견이 제시되었다. 반대로 알림 멘트에 대해서는 당장 행동을 취하지 않아도 되는 덜 중요한 멘트이므로 더 부드럽게 말하는 것이 좋다는 요구사항이 있었다.



그림 4. 경고.알림시 경어체 선호도 및 적합도

3.2.3 알림멘트에서의 해요체의 허용

경고 멘트와 알림 멘트에서 합쇼체와 해요체를 비교했을 때에는 앞에서 살펴본 바와 같이 일반적으로 해요체 보다 합쇼체를 선호하고 더 적절하다고 평가되었다. 그러나 알림 멘트의 경우, “~ 주세요”와 같은 요구의 표현은 사용자들에게 적절한 행동을 요구하는 듯한 느낌을 주기 때문에 해요체를 사용하는 것이 좋다. 따라서 일반적인 경우와는 달리 “린스를 보충해주세요”와 같은 식기세척기 알림 멘트의 경우에는 합쇼체 보다 해요체가 부드러운 느낌을 주고, 듣기에 좋으며, 딱딱하고 짜증이 나지 않게 말해주기 때문에 더 적절하다는 결과가 도출되었다.

따라서 일반적인 경고와 알림 멘트에서는 합쇼체를 사용하여 음성을 제공하되, “~해주세요”와 같은 요구의 표현인 경우에는 부드러운 느낌을 주는 해요체를 사용하는 것이 더 적절하다고 보여진다.

3.2.4 결과논의

전체적으로 이 세션의 실험내용을 정리해보면, 일반적인 경우에 알림과 경고 멘트에서 간결하고 귀에 더 잘 들어오는 합쇼체로 음성을 제시하되, 너무 명령조의 느낌이 들지 않도록 부드럽게 말하는 것이 가장 바람직하다고 할 수 있다. 또한 경고 멘트는 긴박한 느낌을 주기 위해 소리와 함께 제시되거나, 빠르고 높은 음성으로 제시하고 알림 멘트의 경우에는 당장 행동을 취하지 않아도 되기 때문에 부드럽게 말하는

것이 바람직하다.

3.3 음성표현의 간결성

3.3.1 음성표현의 길이

음성을 통해 가전제품의 기능과 상태 등을 전달하는 경우에 그 정보는 서술적 표현이 되기 쉬운데, 이런 경우 음성의 길이는 다양한 수준으로 생성될 수 있다. 또한 제품에 직접 적용하기 위해서는 음성이 긴 경우 사용자 조작에 의해 음성 정보가 중단되거나 잡음처럼 들릴 수 있는 문제점을 안고 있다. 이런 측면에서 사용자의 빈번한 조작으로 음성표현의 간결성이 요구되는 모드 전환이나 수치 설정 작동기능에서 음성 표현의 길이를 세 가지로 구분하여 비교 분석하였다.

“25도 입니다.”와 같은 서술어미를 지닌 음성표현과 이를 생략하고 특정 모드나 수치(“25도”)를 표현하는 경우, 그리고 이러한 빈번한 조작에 음성표현 대신 상승이나 하강 등을 표현하는 청각신호(sound signal)을 사용한 경우를 비교하였다.

사용자들은 빈번한 조작으로 상태를 전환시키는 기능에는 간결한 음성표현을 선호하고, 또한 적합하다고 평가하였다. 서술어미를 포함한 음성표현의 경우는 말이 너무 길다고 느끼거나, 공손하지만 짜증스럽다고 판단하였다. 이에 반해 핵심내용만 음성으로 표현한 경우는 음성 정보의 명확성을 전달할 수 있으면서도 간결하여 효과적인 것으로 나타났다. 그러나 이 또한 딱딱 끊어지는 말투로 인해 사무적인 느낌을 보인다. 음성표현 대신 사용된 청각신호의 경우는 기존 방식과 차별화 되지 않고, 청각신호 자체에 대한 평가 수준에 따라 부정적 영향을 크게 받는 것으로 나타났다. 즉, 음성표현에 비해 부적절한 청각신호의 적용이 기능과 제품에 대한 부정적 영향으로 이어질 수 있는 위험을 지니고 있다.

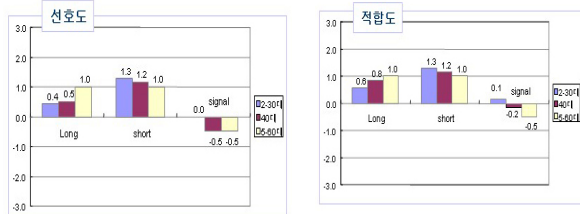


그림 5. 음성표현 길이 간결성의 연령별 결과

고 연령대 사용자들은 가전제품에서의 음성표현 자체의 기능적 참신성과 젊은 사용자들에 비해 빠쁘지 않은 생활 패턴으로 인해 보다 여유 있게 음성표현을 이용할 수 있음을 보여준다. 또한 고 연령대의 사용자들은 시력이 좋지 않아 조작 움직임과 같은 신체적 활동에 민첩하지 못한 부분을 음성표현으로 충족시킬 수 있는 이점을 지니며, 감성적으로도 보다 공손한 기계(가전제품)를 선호한다는 것을 알 수 있었다.

3.3.2 최종설정상태의 안내여부

음성표현의 간결성에 대한 다른 이슈는 가전제품 가운데 세탁기와 같이 특정한 상태로 몇 가지 설정 값들을 조작한 후에 작동하는 경우, 이들 설정 값들을 각 단계마다 하나하나 음성으로 표현하는 방법과 한번에 최종 정리하여 설정상태를 음성으로 표현하는 방법을 생각해 볼 수 있다. 이들 방

법에 대한 사용자 평가를 확인하기 위해 본 연구에서는 설정 값들을 조작할 때마다 해당 값들을 음성으로 표현해주는 조건과 이에 더해 최종 확인 시 설정을 정리해서 표현하는 조건, 또한 개별 설정 값에는 단순 청각신호만 제시하고 최종 확인 시에 설정 값을 한번 정리해서 표현하는 조건을 비교하였다.

그 결과, 사용자들은 설정 값의 개별 조작 음과 최종 설정 상태의 확인을 음성으로 표현하는 조건을 더 선호하고 적합하다고 평가하였다. 특히 여러 가지 설정 값을 조작할 때마다 음성표현을 통해 확인할 수 있는 점을 높이 평가하였다. 또한 필요에 의해 최종 설정상태를 표현하고자 하는 경우는 너무 길어지지 않도록 하는 것이 중요하다.

3.3.3 최종설정상태의 안내여부에 대한 사용자 연령별분석

최종설정과 관련한 연령별 분석 결과로 저연령층은 간결한 표현을 선호하고 고연령층은 의미가 확실치 않을 경우 시각적 표현을 보완할 수 있는 음성표현을 선호하는 것으로 나타났다.

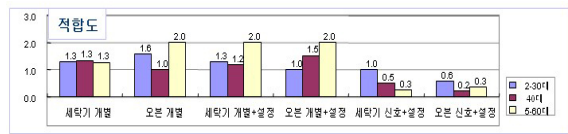
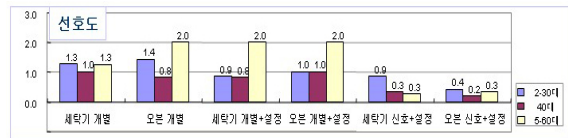


그림 6. 가전제품 기기별 음성표현 길이 간결성의 연령별 결과

3.4 경고/알림시의 음성메시지

3.4.1 개요 및 제작도구

본 실험과 같이 프로토타입을 이용하여 음성 피드백을 듣고 주관적인 선호도와 적합도를 평정하는 것은 실제 집에서의 작업 환경을 충분히 고려한 것이라고 보기 어렵다. 사람들은 가전 기기에만 집중을 하고 조작하는 경우는 거의 없다. 일반적인 환경이라면 집안에서 다른 가족 구성원과의 대화, TV 등의 소음, 다른 작업하기 등 무수한 방해(Distraction)가 일어나곤 한다.

따라서 본 실험의 4번 세션은 실제로 TV를 시청하면서 경고와 알림의 소리 피드백, 음성 피드백, 소리와 음성을 함께 제시하는 피드백 등을 듣고 가전 기기로 직접 이동해서 문제 상황을 조치하도록 설계되었다. 이를 위하여 Microsoft Visual Basic 6.0을 이용, 문제가 일어나는 상황 및 그에 대한 조치를 할 수 있는 프로그램을 제작하였다.

3.4.2 자극제시

프로그램은 프로토타입이 제시되는 모니터에 제시되며, 기기 명, 경고 혹은 알림의 내용이 제시되고, 이에 대한 선택을 할 수 있는 보기가 4개 제시된다. 과제는 총 12번 (6개 제품 * 경고 및 알림) 제시되었으며, 12번의 순서는 무선적으로 제시되도록 설계되었다.

3.4.3 실험수행

경고/알림 수행 비교 세션은 경고/알림 상황에 대해 신호음만 들려주는 경우, 음성만 들려주는 경우, 신호음과 음성을 연속하여 들려주는 경우에 대하여 사용자가 어떤 경우를 선호하고 효과적으로 상황에 대처할 수 있는지 알아보기 위해 설계되었다. 기존 제품들은 경고/알림 상황에 대하여 신호음만 제공하기 때문에 실제 어떤 상황인지 주부들이 제품 앞으로 다가가 디스플레이를 확인하기 전까지는 알 수 없다. 사전 F.G.I. 에서도 주부들은 중요하지 않거나 당장 대처할 필요가 없는 알림 상황에서도 어떤 상황인지 파악하기 위해 하던 일을 멈추고 해당 가전제품에 다가가야 하는 상황을 문제로 꼽았다. 따라서 음성 UI 를 이용하여 가장 효과를 거둘 수 있는 기능은 바로 경고/알림 메시지를 음성으로 알려주는 것이다. 그러나 음성의 심각한 단점은 주변의 소음이나 말소리에 의해 쉽게 차폐된다는 것이다. 신호와 소음(noise)의 구별이 불명확해짐에 따라 민감도가 떨어져 탈루(miss) 나 오경보(false alarm)가 증가하여 결국 중요한 신호음을 놓치는 불상사가 일어날 수 있게 된다. 따라서 이론적으로 최적의 해결책은 신호음과 음성을 같이 제시하여 주의 전환의 효과와 의미 전달의 효과를 함께 하는 것이다. 단, 이때 신호음이 음성을 차폐할 수 있으므로 신호음과 음성의 크기와 적절한 시간적 배열이 중요하다.

본 실험에서는 신호음이 먼저 나타나고 약 500 ms ~ 1000 ms 후에 음성이 제시된다. 긴 신호음의 경우 음성이 제시되는 동안 음량을 감소시켜 방해를 최소화 시켰다. 또한 주변 잡음에 의한 차폐 효과를 가상적으로 검증하기 위하여 실제 사용자가 제품과 5미터 이상 떨어진 상태에서 고개를 돌리고 TV를 시청하는 상태를 유지하게 한 뒤 갑작스럽게 경고/알림 메시지를 들려주었다. 참가자가 신호를 감지하면 바로 제품(프로토타입) 앞으로 다가가 디스플레이에 나타난 경고/알림 메시지를 확인하고 4가지의 대처 방법 중 하나의 적절한 대처 방법을 선택하는 방식으로 실험이 진행되었다.

대상 제품은 경고/알림이 가장 빈번하고 중요한 세탁기와 오븐레인지가 선택되었고 각각 경고와 알림 메시지 한 가지씩 포함되었다. 총 12회의 시행마다 각각의 메시지에 대한 선호도, 적합도, 반응 시간, 오류횟수가 측정되었다.

3.4.3 실험결과

신호음만 제시하는 조건(Signal only), 음성만 제시하는 조건(Voice only), 신호와 음성 모두 제시하는 조건(Signal and Voice)의 세 조건으로 나누어 측정한 결과, 알림과 경고 신호에 대한 반응시간의 순서는 신호음만 제시하는 조건 > 음성만 제시하는 조건 > 신호와 음성 모두 제시하는 조건의 순서로 신호음만 제시하는 조건의 경우가 반응시간이 가장 길었다. 신호와 음성 모두 제시하는 조건과 음성만 제시하는 조건의 반응시간 차이는 미미하였는데, 이를 통해서 음성이 포함된 경우가 반응시간이 가장 빠르다는 결론을 내릴 수 있다. 실험 참가자 중에는 이 결과에 대한 이유로 음성을 통한 경고의 장점은 무슨 일이 일어났는지 알 수 있어서 시간을 절약할 수 있다는 점을 들었다. 결국 가전기기의 경고와 알림을 알려주는 경우에는 음성을 반드시 포함시켜야 함을 알 수 있다. 하지만 TV 소리에 묻혀 안 들리거나 급박함이 덜 느껴진다는 우려도 있었다. 하지만 이러한 점은 신호음을 추가함으로써, 신호음만으로 경고를 알아채고 그 다음 주의를 기울여 목소리를 듣는 방식의 개선이 가능하다. 이때 신호음은 목소리를 방해하지 않는 수준의 것이어야 한다.

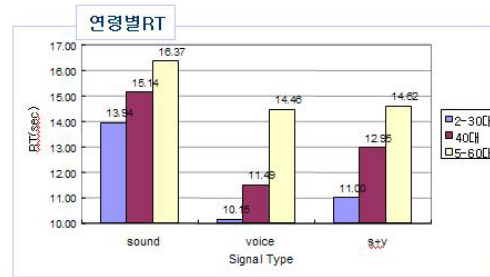


그림 7. 경고.알림 음성메시지 반응시간 연령별 분석

또한 이를 연령별로 분석해보면, 세 조건에서 각각 고 연령층의 반응시간이 길었다. 하지만, 연령을 고정시키고 각 조건별로 비교하였을 경우, 앞서의 결과처럼 신호음만 제시하는 조건 > 음성만 제시하는 조건 > 신호와 음성 모두 제시하는 조건의 순서는 유지되었다. 또한 음성을 포함시키는 조건에서는 저 연령층의 경우 음성의 효과가 확인하게 드러난다. 이러한 결과는 음성이 포함된 UI에 대한 선호도로 이어질 것이다.

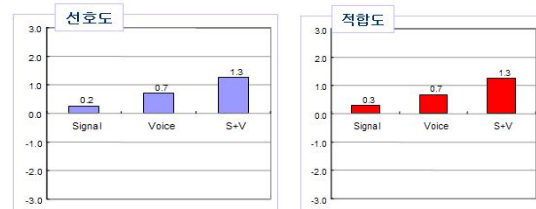


그림 8. 경고.알림 메시지 선호.적합 결과

반응시간이 기록되는 것과 더불어, 실험참가자는 각 소리에 대한 선호도와 기기에 대한 적합도를 평정하였다. 그 결과 경고의 경우 신호음만 제시하는 조건, 음성만 제시하는 조건, 신호와 음성 모두 제시하는 조건의 순으로 선호도와 적합도가 나타났고 알림의 경우 신호음만 제시하는 조건만이 선호도 및 적합도가 낮게 나타났다. 이러한 사용자들의 반응은 앞서의 이론적 고찰과 일치한다. 첫째, 신호와 음성 모두 제시하는 조건이 가장 선호도 및 적합도가 높았다는 사실은 사용자가 신호와 의미 모두를 필요로 한다는 사실을 의미하고 음성만 제시하는 조건이 다음 순위로 평정된 것은 사용자가 단순한 신호음보다는 의미를 더욱 중요시한다는 점을 시사한다.

이것은 일반적인 가전제품에서 대체로 경고 / 알림 메시지의 경우 신호의 의미를 알기 위해 매뉴얼을 뒤지거나 제품 옆에 붙은 경고메시지 리스트를 참고해야만 그 의미를 파악할 수 있는 상황이 매우 불편한 상황이었음을 간접적으로 암시한다. 또한 음성을 추가하는 것 자체만으로도 제품이 고급스러워 보이거나 첨단 느낌을 받는다는 의견도 무시할 수 없다. 신호음과 음성메시지가 결합된 VUI 인터페이스는 가장 큰 사용성의 개선을 가져오는 부분으로 파악될 수 있다.

한편 경고와 알림 메시지는 그 긴급한 정도에 있어서 차이를 보이는데 알림 메시지의 경우 대체로 긴급도가 높은 경우는 거의 없다. 오히려 사용자가 메시지를 듣고 중요하지 않은 메시지의 경우 반응을 보류하기를 원하는 경우도 많다. 알림 메시지의 경우 음성 메시지만 들려주는 조건과 신호음

을 포함시킨 경우에 거의 차이가 보이지 않고 단지 신호음만 있는 경우를 비 선호하는 경향을 보이는데, 이 결과는 긴급도가 높지 않은 알림 메시지의 경우 반드시 의미를 파악할 수 있는 음성메시지를 사용하여 사용자가 제품 앞으로 움직이기 전에 자신의 반응을 결정할 수 있게 해야 한다는 의미를 담고 있다고 볼 수 있다.

연령별 분석 결과 고 연령층은 신호음만 제시하는 조건과 신호와 음성 모두 제시하는 조건에 비해 음성만 제시하는 조건을 비 선호하였고, 저 연령층은 신호음만 제시하는 조건 조건에 비해 음성만 제시하는 조건과 신호와 음성 모두 제시하는 조건을 선호하였다. 이러한 결과는 고 연령층의 경우 경고 / 알림으로 음성만 주어질 때 주의 전환 능력이 약하여 사용성을 충분히 반영하지 못한다고 판단한 것으로 추정되고, 저 연령층의 경우 오직 Voice 가 포함되는 경우에 대하여 선호한 것으로 보인다.

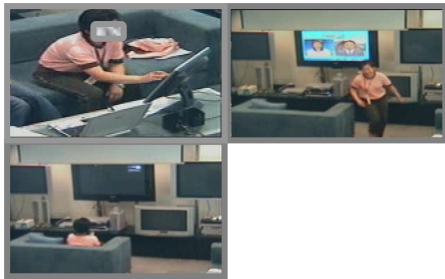


그림 9. 경고,알림 메시지 실험과정

3.5 프로토타입의 전반적 비교

본 연구의 사전 F.G.I 결과에 의거 Signal only, Full voice, Optimal 1(S/V: Signal + 설정 값 정리), Optimal 2(S+V: 매 조작 때마다 Signal + Voice)의 네 조건의 프로토타입이 제작되었고 각각의 전체적 VUI 에 대하여 선호도와 적합도를 평정 받았다. Signal only 조건은 시작과 조작 중, 작동 시작 시 신호음만 나오는 조건이고, Full Voice 조건은 모든 조작음을 음성으로 알려주는 조건, Optimal 1은 전원과 특수기능은 음성으로 하되 모드전환과 값 설정 기능은 신호음으로 하고 최종 설정 시 설정된 사항을 정리해서 알려주는 조건, Optimal 2 는 Optimal 1과 비슷하지만, 모든 작동과정에서 신호음과 음성을 함께 들려주었다. 또한 다이얼을 돌려야 하는 경우 빠르게 돌릴 때는 Signal 만, 돌리다가 멈추었을 때는 Voice 가 제시됨으로써 가장 명확한 피드백이면서 효율적이 되도록 설계하였다

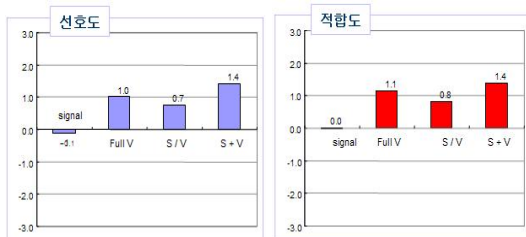


그림 10. 프로토타입 전반적 비교

평가 결과 참가자들은 Signal only 조건보다는 Voice 가

있는 다른 세 조건을 선호하였다. 이들 중 특히 S+V 조건이 가장 선호되는 경향을 보였다. Optimal 1과 Optimal 2 조건만을 비교한다면 Optimal 2를 비교적 더 선호하는 경향을 보였는데 이것은 명확한 피드백과 조작 시 음성이 함께 나오는 것에 대한 선호를 보여준다. 이러한 선호의 이유로는 여러 가지 있었지만, 특징적으로는 “실수가 적어서” 혹은 “정중하고 친절한 느낌을 받는다.” 등이 있었다. 또한 이는 “눈이 나쁘거나 어두운 환경에서 도움이 된다.”는 의견도 있었고, 반대로 소리만 나오는 경우는 “무슨 메시지인지 알 수 없다.” 혹은 “기존 제품과 구별이 안 된다.”는 등의 의견이 있었다. 하지만, 저 연령층에서는 “앞에서 보고 있는데 매번 읽어들 필요가 없다.”는 의견과 “주방에서 들리는 라디오 소리를 못들을 것 같다.” 혹은 “자주 들으면 식상할 것 같다.”는 의견도 있었으므로, 음성의 사용에 대한 특별한 주의를 필요로 함을 알 수 있다. 결국 Full Voice 조건은 Optimal 1과 Optimal 2의 선호도에 비해 중간 정도로 평가되었는데, 이를 정리해보면 조작시의 모든 경우에 Optimal 2와 같이 Signal과 음성을 연속하여 제시하고, 빠르게 값을 설정하는 경우 조작 음만을, 잠시 조작을 멈추었을 때 Voice로 알려주는 것이 가장 좋은 신호음과 음성의 배치방법이라는 사실을 알 수 있다.

4. 종합논의

목소리 측면에서는 보통 톤의 여성의 조용한 억양을 선호하였다. 인터뷰 결과 정중,공손, 격식 있는 느낌이기 때문임을 알 수 있었다. 경어체 측면에서는 역시 동일한 이유로 합쇼체를 선호하였고 기능에 따른 신호음과 음성 안내의 할당 측면에서는 명확한 피드백과 의미 전달이 모두 가능한 신호음과 음성안내가 결합된 인터페이스를 선호하였다. 특히 사용자들은 음성 인터페이스가 가장 필요한 기능으로 경고/알림을 손꼽았는데 사용자 반응 시간 조사 및 해당 세션의 인터뷰 결과 일상생활에 가까운 소음 조건에서 신호음은 사용자의 주의 전환을 유발하였고 음성 안내는 경고/알림의 의미를 빨리 전달하여 사용자가 신속히 의사를 결정하고 문제를 해결하는 데 도움을 준 것으로 나타났다. 전체적으로 조작음 및 경고음 모두에서 적절한 AUI 와 함께 간결한 음성 메시지를 정중하고 공손한 느낌으로 제시하는 것이 가전제품의 음성 표현에 적합함을 확인하였다. 본 연구결과를 바탕으로 음성 사용자 인터페이스 가이드라인을 바탕으로 최적의 프로토타입을 제작하였고 이는 실제 가전제품에 적용 가능하도록 하였다.

참고문헌

- [1] Damper, R. I, and Wood. S. D(1995) "Speech versus keying in communication and control applications",*Int.J.Human-Computer Studies.* 42, 289-305.
- [2] Jones,D.,Hapeshi,K.,& Frankish, C(1989) "Design guidelines for speech recognition interfaces", *Applied Ergonomics*, 20(1),47-52.
- [3] Mcleod, R. M. (1994) "Human factors issues in speech recognition", In Lovesey, E.J.(Ed),

Contemporary Ergonomics, Taylor and Francis, London.

- [4] Michael H., James P. Giangola., Jennifer Balogh
“ Voice user interface design”, Addison-Wesley,
Boston.
- [5] Nass C., & Gong L.(2000). Speech interfaces from
an evolutionary perspective. *Communications of
the ACM*, 43(9) 36-43
- [6] NCAM, A developer's Guide to creating talking
menus for set-top boxes and DVDs, National
center for accessible media , accessed 2006-10.
- [7]Neuhoff, J. G., Kramer, G., & Wayand, J.
(2002). Pitch and loudness interact in auditory
display: Can the data get lost in the map? *Journal
of Experimental Psychology*, 17-25.
- [8] Noyes, J. M., and Frankish, C. R(1994) “Errors
and error correction in automatic speech
recognition systems”, *Ergonomics*, 37(11), 1943-
1957.