

음성 및 비접촉 제스처 기반 차내 정보시스템  
인터랙션의 조작 종류에 따른 효과

The Effects of Operation Types on Interacting with  
Voice and Non-Touch Gestures in In-Vehicle Information Systems

주 저자: 김 현

서울과학기술대학교 기계시스템디자인공학과 교수

**Kim, Huhn**

Seoul National University of Science and Technology,  
Department of Mechanical System Design Engineering

공동 저자 : 박수현

서울과학기술대학교 NID융합기술대학원 석사과정

**Park, Soohyun**

Seoul National University of Science and Technology,  
Graduate School of NID Fusion Technology

\* 이 연구는 서울과학기술대학교 교내 학술연구비 지원으로 수행되었습니다.

1. 서론

2. 연구방법

- 2-1. 사전 실험
- 2-2. 실험 환경
- 2-3. 실험 태스크 및 시스템
- 2-4. 실험 계획

3. 실험결과

- 3-1. 그룹별 분석
- 3-2. '양 조절' 조작 분석

4. 결론

참고 문헌

(요약)

운전 중 차내 정보시스템 (IVIS: In-Vehicle Information System)을 조작하는 것은 운전자의 시선을 뺏고 주의력을 도로에서 벗어나게 하여 교통사고를 유발하는 주요한 요인 중의 하나이다. 이 문제를 완화하고자 손 조작이 필요 없는 음성인식에 의한 조작관련 연구들이 활발히 진행되었으나 음성인식은 차내 노이즈로 인해 인식 정확도가 떨어질 가능성이 크다. 이에 최근에는 비접촉 제스처에 의한 차내 정보시스템 조작에 대한 연구도 진행되고 있다. 따라서 본 연구는 효율적이고 안전한 차내 정보시스템 조작 인터랙션을 위한 비접촉 제스처와 음성인식의 최적 조합을 알아보기 위한 실험을 수행하였다. 이를 위해 차내 정보시스템 조작 인터랙션을 조작과 양(예. '볼륨조절 + 크기 3')의 개념으로 분리하여 (1) 모두 음성으로 조작, (2) 모두 비접촉 제스처로 조작, (3) 조작은 제스처, 양은 음성으로 조작, (4) 조작은 음성, 양은 제스처로 조작하는 네 그룹으로 구분하였다. 실험 결과, 음성으로만 조작한 그룹이 가장 조작시간이 짧고 오류도 적은 것으로 나타났다. 하지만 조작 명령은 음성으로 하고 양의 조작은 제스처로 하는 방식이 만족도가 가장 높았으며 시간이나 오류 측면에서도 효율적인 것으로 나타났다. 본 연구에서는 이러한 실험 결과를 기반으로 차내 정보시스템에 적합한 음성 및 비접촉 제스처 기반 혼합조작 방식을 제안하였다.

(주제어)

차내 정보시스템, 음성인식, 비접촉제스처, 멀티모달 인터랙션

(Abstract)

Operating in-vehicle information systems (IVISs) while driving is one of the main causes of car accidents. Thus, in order to lessen the risk, many studies have focused on the application of voice recognition, but its accuracy can be highly compromised due to surrounding noises. Thus, some researchers have paid attention to the operation of IVISs using non-touch gestures while driving. In this study, an experiment was conducted to investigate the efficiency and safety of operating IVISs using voices and/or non-touch gestures while driving. For the purpose, the interactions relevant to IVISs were divided into two kinds of operations: the operation itself and its degree (ex. 'adjust radio volume' to 'level 7'). Then, the following four combinations were derived: (1) All Voices, (2) All Gestures, (3) Gesture+Voice (gesture operation with the degree indicated by the voice), and (4) Voice+Gesture (voice operation with the degree indicated by the gesture). The experimental results showed that the 'All Voices' group required less operation time and resulted in less errors than the other groups. However, the 'Voice+Gesture' group, which could complement the limitation of voice recognition techniques, showed higher subjective satisfaction scores, as well as less operation time and less errors. From the experimental results, this study recommended a multi-modal interaction design based on both voice and non-touch gesture that is likely to be applied in the operation of IVISs.

(Keyword)

In-vehicle Information Systems, Voice recognition, Non-touch gestures, Touchless gestures, Multi-modal interaction

## 1. 서론

운전은 다양한 중요도를 갖는 여러 가지 운전 및 비운전 과제들을 포함한다. 도로상의 안전에서 가장 중요한 것은 운전자의 전방에 보이는 영역인 일차 시각주의영역(Primary Visual Attention Lobe, PVAL)으로부터 운전자의 시각적 주의를 빼앗는 경쟁 과제들의 수준이다(Wickens et al., 2004). 운전하면서 지도를 보거나 라디오의 다이얼을 조절하는 것, 햄버거를 먹는 것 등도 일차시각주의 영역과 시각주의를 놓고 경쟁하는 과제들이다(Dingus et al., 1988). 최근 이러한 과제들 중의 하나인 차내 정보시스템(IVIS: In-Vehicle Information Systems)은 대부분 터치 디스플레이의 형태를 많이 띠고 있다. 운전 중 이 시스템을 조작하기 위해서 운전자는 디스플레이에 시선을 두고 손을 뺀어 원하는 버튼 혹은 아이콘을 정확하게 터치해야 한다. 이때 운전자의 시선은 도로로부터 벗어나게 되어 운전 부주의로 이어질 가능성이 커지며(이대영 외, 2010), 이러한 선택주의(Selective attention)의 실패는 자동차 사고의 주요한 원인 중 하나이다(Malaterre, 1990).

따라서 터치 조작의 오류를 줄이고 운전자의 운전 에 대한 집중도를 높이기 위해 터치가 필요 없는 음성인식을 조작에 적용함으로써 사용자의 편의성 및 안정성을 높이려는 연구들이 활발히 진행되었다. 음성은 사람이 가장 편리하게 느끼고 정보전달을 위한 주된 모달리티(Modality)이다(Hong, 2004). 하지만 이러한 음성인식 방법은 차량 내·외부에서 발생하는 다양한 노이즈(예. 차체 소음, 음악, 동승자간 대화 소리 등)에 의해 인식의 정확도가 떨어질 수 있다는 약점을 가진다. Lee and Billinghamurst(2008)의 연구에 따르면 12명의 피실험자 중 75%는 시스템과의 음성대화가 자연스럽게 못하다고 여겼으며 시스템과의 대화를 선호하지 않았다.

음성인식 외에 터치 조작을 대체할 수 있는 방법으로 비접촉 제스처를 활용한 조작에 대한 연구들도 진행되었다(Jung, 2007). 비접촉 제스처는 앞에서 언급한 음성인식의 한계를 극복하면서도 운전자의 시각 분산을 줄여줄 수 있다. Alpern and Minardo (2003)의 연구에 따르면 비접촉 제스처 조작이 물리적인 인터페이스를 손으로 직접 조작하는 것보다 낮은 오류율을 보였으며, 운전자들의 선호도도 높게 나타났다. 또한, 사람과 시스템의 상호작용에 음성과 제스처를 동시에 사용하는 멀티모달 인터페이스(Multi-modal Interface)는 사용의 편리성과 자연스럽게다는 장점에 의해 많은 연구가 이루어지고 있다.

Epps et al.(2004)의 연구에 따르면, 18명의 피실험자들을 대상으로 사진관리 어플리케이션을 조작하는 태스크들을 음성 조작, 비접촉 제스처 조작, 음성과 제스처의 혼합 조작 중에서 자유롭게 사용하도록 하였는데 그 중 63%가 멀티모달인 혼합 조작을 사용하였고 혼합 조작 중 95%는 음성과 제스처를 중복으로 활용하는 동시 조작이었다. Lee and Billinghamurst(2008)의 연구에서도 피실험자 12명을 대상으로 AR(Augmented Reality) 인터페이스를 음성과 제스처로 조작하도록 하였을 때, 전체 중 63%가 멀티모달을 사용하였다. 그러나 이 연구에서는 앞서 설명한 Epps et al.(2004)의 결과와는 달리 혼합 조작 중 94%가 제스처가 음성 명령보다 선행하는 순차적 조작형태를 보였다.

하지만 음성과 비접촉 제스처의 혼합 조작에 대한 위 연구들은 (1) 사용자가 선호하는 조작방법에 관해서만 분석하였으며, (2) 음성과 제스처간의 혼합 조작을 같은 의미의 조작에 대한 중복의 개념으로만 다루었으며, (3) 음성과 제스처 간의 선후 관계에 대한 결론도 서로 달랐다는 한계를 가진다. 특히, (4) 위 두 연구에서는 차내 정보시스템의 조작에 필수적인 양을 조절하는 조작(예. 볼륨을 7로 조절하기)을 일반 조작(예. 음악 재생하기)과 분리하여 조사하지 않았다. 이에 본 연구에서는 비접촉 제스처와 음성인식을 차내 정보시스템의 조작 인터랙션에 활용할 때, 효율성과 안전성의 관점에서 그 효과를 살펴보고 두 모달리티의 혼합 시 양 조절에는 어떤 조작이 더 적합한 지에 대해 알아보는 것을 목표로 이를 알아보기 위한 실험을 수행하였다.












## 2. 연구방법

### 2.1. 사전 실험

본 실험에 앞서 조작하고자 하는 기능에 따라 사용자들이 선호하는 음성 명령어와 비접촉 제스처 조작 방식을 알아보기 위하여 피실험자 28명(서울과기대생 남자 14명, 여자 14명)을 대상으로 실험을 진행하였다. 실험시스템으로는 차내 정보시스템에 일반적으로 포함되는 대표 기능인 MP3 플레이어(Gom Player)와 내비게이션(Atlan)을 선정하였다. 피실험자들에게 표 1과 같은 태스크들을 주고 음성 명령어와 비접촉 제스처를 사용하여 각각 조작해 보도록 하였다. 표 1은 사전 실험을 통하여 도출된 조작법들 중에서 가장 빈도가 높게 나타난 음성 명령어와 제스처 조작 방식을 나타낸 것이며 동일한 조작에 대해 제스처가 중복된 경우 Kim et al.(2011)의 연구결과

를 참고하여 중복되지 않도록 다른 제스처를 선정하였다. 전체적으로 음성에 비해 제스처는 다양한 형태로 결과가 나와서 빈도가 낮음을 알 수 있다. 이렇게 선정된 표 1을 본 실험의 음성 명령어와 제스처 조작 방식으로 적용하였다.

[표 1] 사전실험 및 Kim et al.(2011)의 연구 결과 가장 빈도가 높은 음성 명령어와 제스처 조작 방식

태스크	음성 명령어	제스처 명령어
볼륨 조절	볼륨 업/다운 21.9% 볼륨 올려/내려 21.9%	 69.8%
화면크기조절	화면확대/화면축소 66.7%	 45.8%
빨리감기/되감기	빨리감기/되감기 62.5%	 42.5%
스크롤	위/아래 79.2%	 55.2%
재생	재생 30.6%	 18.0%
일시정지	일시정지 30.6%	 58.0%
정지	음악정지 29.2%	 13.0%
확인	(길)안내시작 54.2%	 19.2%
취소	(길)안내취소 54.2%	 24.1%
명칭 입력	명칭 70.8%	 16.7%
이전/다음	이전/다음 76.4%	 49.0%

## 2.2. 실험 환경

본 실험에서는 피실험자의 음성과 제스처 조작에 대한 확실한 피드백을 디스플레이 상에 제공하기 위해 오즈의 마법사(Wizard of Oz: Gould et al., 1983; Lee and Billinghamurst, 2008) 기법을 사용하였다. 이는 그림 1과 같이 피실험자가 디스플레이를 보면서 주어진 태스크를 음성과 제스처 조작으로 수행하면 마법사 역할을 맡은 진행자(Wizard)가 대신 해당 조작을 컴퓨터에 입력하고 그 결과가 피실험자의 디스플레이에 보이게 하여 피실험자에게 피드백을 제공해 주는 방식이다. 마법사는 피실험자의 조작에 대해 빠르고 일정한 반응을 할 수 있도록 충분한 훈련을 받은 한 사람이 전담하도록 하였다.

[그림 1] 'Wizard of Oz'기법을 활용한 실험 장면



[표 2] 실험에 사용된 태스크들

태스크	MP3 플레이어(Gom Player)
문자입력	'옥주현-서시'를 재생해주세요
스크롤이동	'10cm-아메리카노'를 찾아주세요
볼륨조절	볼륨을 3칸 줄여주세요
재생/정지	음악을 정지시켜주세요
이전/다음음악	바로 이전(다음)음악을 선택해주세요
되감기/빨리감기	30초 빨리(되) 감아주세요
태스크	내비게이션(Atlan)
문자입력	'공릉역'을 입력해주세요
스크롤이동	하늬은행 공릉점에 가려고 합니다
볼륨조절	볼륨을 3칸 올려주세요
확대/축소	화면을 15% 확대(축소)해주세요
확인/취소	길안내를 시작(취소)해주세요
종료	내비게이션 기능을 종료해주세요

## 2.3. 실험시스템 및 태스크

실험시스템으로는 사전 실험과 동일한 기능과 프로그램을 사용하였다. 일반적으로 이러한 시스템의 조작들은 시작, 종료와 같은 단순 명령 조작과 볼륨 조절, 채널 조절과 같이 얼마나 조절할 것인지에 관

한 양의 개념이 포함된 조작으로 나눌 수 있다. 이와 같이 본 연구에서는 차내 정보시스템의 조작 인터랙션을 '단순 명령' 조작과 '양 조절' 조작으로 구분하였으며, 표 2와 같이 총 12개의 태스크를 선정하여 피실험자가 수행하도록 하였다.

## 2.4 실험 계획

본 실험은 음성과 제스처 조작 및 그 조합에 따른 효과를 보기 위한 것이다. 조작방법은 제스처만 사용하는 그룹(G: Gesture), 음성만 사용하는 그룹(V: Voice), 음성과 제스처를 조합하여 사용하는 그룹으로 구분할 수 있다. 또한 음성과 제스처 조합 그룹은 양의 조절을 음성으로 하는 그룹(GV: Gesture+Voice)과 제스처로 하는 그룹(VG: Voice+Gesture)으로 나눌 수 있으므로 표 3과 같이 총 4개의 그룹으로 구분하여 실험을 진행하였다. 각 그룹에 따른 조작방법의 차이는 다음과 같다. 예를 들어 단순 명령 조작인 '길안내를 시작해주세요'와 양 조절 조작인 '길안내 음량을 3칸 줄여주세요'와 같은 태스크가 주어질 경우, V그룹은 "시작"을 말하고 "볼륨 감소, 감소, 감소"라고 말하여 태스크를 수행하고, G그룹은 손으로 'OK'라는 제스처를 취하고 검지를 왼쪽으로 3번 치는 제스처를 취함으로써 태스크를 수행한다. 반면에 GV그룹은 'OK'라는 제스처를 취하여 검지를 왼쪽으로 치면서 "3"이라고 말하여 태스크를 수행하고, VG그룹은 "시작"이라고 말한 후에 "볼륨"이라고 말하면서 검지를 왼쪽으로 세 번 치는 제스처를 취하여 태스크를 수행한다.

피실험자가 태스크를 수행하는 동안의 모든 시간을 기록하였으며 이때, 잘못된 명령으로 조작한 시간을 오류시간으로, 명령어를 참조한 시간을 참조시간으로 기록하였고, 조작시간은 위 두 경우를 제외한 순수 조작시간을 의미한다. 또한 잘못된 명령으로 조작한 오류횟수와 명령어를 참조한 횟수도 기록하였다. 모든 실험을 완료한 후에는 피실험자들에게 각각의 조작에 대한 만족도를 5점 척도로 평가하도록 하였다.

[표 3] 실험 계획표

	G	GV	V	VG
남자	4	4	4	4
여자	4	4	4	4

실험에는 서울과기대생 20~30대 총 32명(남자 16

명, 여자 16명)이 참여하였다. 이들을 표 3과 같이 네 그룹 각각에 8명(남자 4명, 여자 4명)씩 랜덤하게 배치하였다. 실험에 사용되는 음성과 제스처는 일관성을 고려하여 피실험자에게 지정된 조작만 사용하도록 하였으며, 실험 수행 중에 조작법이 적힌 문서를 참조할 수 있도록 하였다. 특히 운전 중 환경과 유사하도록 피실험자의 시각주의를 뺀기 위한 시각 차폐 기법(Occlusion technique)을 사용하였다. 시각 차폐 기법이란 피실험자의 시야를 주기적으로 가려 실제 운전환경에서 운전자의 시각주의가 차내 정보시스템으로 뺀기는 상황을 가상적으로 모사하는 기법이다(Young et al., 2003; 박정철, 2009). 본 연구에서는 피실험자의 시각차폐를 위해 3D TV를 보는데 사용되는 셔터글라스(Shutter glasses)의 열림 및 닫힘을 1~4초(0.5초 간격)로 설정이 가능하도록 개발하였으며, 실험 중 차폐시간은 ISO-16673(2007)에 명시된 시각차폐 표준 시간인 열림 1.5초, 닫힘 1.5초 기준을 적용하였다. 셔터글라스의 착용여부에 따른 그룹 내 피실험자의 만족도와 조작시간의 차이를 확인하기 위하여 한 명의 피실험자가 착용, 미착용의 두 조건을 모두 수행하였다. 셔터글라스의 착용 여부, 즉 운전 및 비운전에 따라 요구되는 운전자의 시각주의력을 알아보기 위하여 아래와 같은 공식으로 계산되는 Occlusion coefficient인 R값을 분석하였다.

$$R_{값} = \frac{\text{셔터글라스 착용 시 조작시간} / 2}{\text{셔터글라스 미착용 시 조작시간}}$$

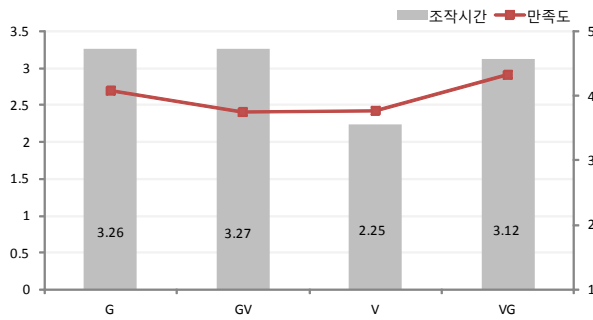
## 3. 실험 결과

### 3.1. 그룹별 분석

측정된 데이터가 정규분포를 따르지 않아 그룹에 따른 차이를 Kruskal-wallis와 Mann-Whitney U 비모수 검정으로 분석하였다.

그림 2는 그룹별 조작시간과 만족도를 나타낸 그래프이다. 조작시간은 그룹 간에 유의한 차이를 보였는데(H(3)=16.976, p=0.001), V그룹의 조작시간이 2.25초로 가장 짧았다. 특히 제스처로만 조작한 G그룹보다 음성과 제스처를 혼합하여 조작한 VG그룹('양 조절'을 제스처로 조작한 그룹)의 조작시간이 조금 더 짧게 나타났다. 그룹 간 만족도에도 유의한 차이가 있었으며(H(3)=42.389, p=0.000), V그룹은 빠른 조작시간에 비해 만족도가 3.77로 비교적 낮게 나타났다(G그룹 4.08, GV그룹 3.75, VG그룹 4.33).

[그림 2] 그룹별 조작시간 및 만족도



[그림 3] 성별과 그룹에 따른 조작시간

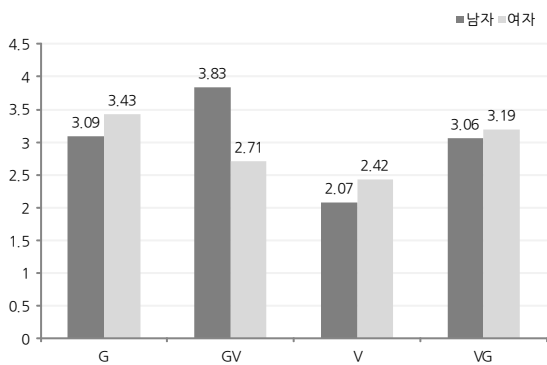


그림 3은 성별과 그룹에 따른 조작시간을 나타낸 그래프로 성별에 따라 조작시간에 유의한 차이가 존재하였다( $U=79,056.0$ ,  $P=0.031$ ). 남자는 GV그룹(‘양 조절’을 음성으로 조작한 그룹)에서, 여자는 G그룹에서 조작시간이 가장 느리게 나타났지만 남녀 모두 조작시간이 가장 빠르게 나타난 그룹은 V그룹이다.

그림 4에서 보여주듯이 오류시간도 그룹 간에 유의한 차이를 보였으며( $H(3)=13.548$ ,  $p=0.004$ ), V그룹이 0.16초로 가장 짧았고, GV그룹(‘양 조절’을 음성으로 조작한 그룹)이 0.78초로 가장 길었다. 참조시간에서도 그룹 간에 통계적으로 유의한 차이를 보였으며( $H(3)=25.804$ ,  $p=0.000$ ), V그룹이 0.08초로 가장 짧았고, GV그룹이 1.52초로 가장 길었다. 즉, 음성명령은 기억하기 용이하여 조작법을 참조하지 않고도 쉽게 수행하였으나 제스처 명령 혹은 혼합된 경우에는 음성보다는 직관성이 떨어져 더 많이 조작법을 참조하였고 오류도 더 많이 발생한 것으로 보인다.

[그림 4] 그룹별 오류시간 및 참조시간

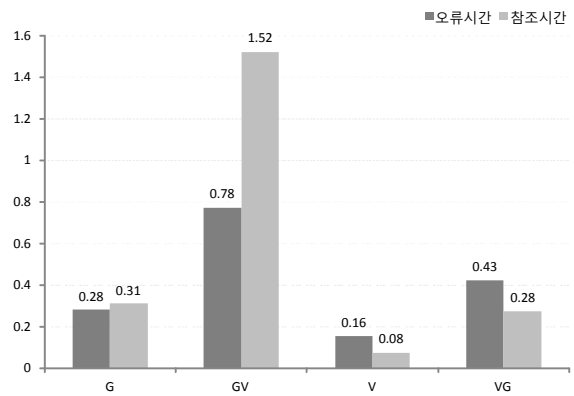


그림 5는 그룹별 오류횟수 및 참조횟수를 보여주는데, 그룹 간에 유의한 차이를 보였다( $H(3)=30.623$ ,  $p=0.000$ ;  $H(3)=37.296$ ,  $p=0.000$ ). V그룹의 오류횟수 및 참조횟수가 각각 0.019회, 0.024회로 네 그룹 중에서 가장 적게 나타났다. 그림 5에 따르면 G그룹과 VG그룹(‘양 조절’을 제스처로 조작한 그룹)의 오류시간 및 참조시간은 비슷하지만 그림 6을 보면 오류횟수 및 참조횟수는 G그룹보다 VG그룹이 더 적었던 것을 알 수 있다. 이 또한 VG그룹에 비해 G그룹에 속한 피실험자가 기억해야 할 제스처의 종류가 더 많았기 때문에 오류나 참조횟수가 많았던 것으로 보인다.

[그림 5] 그룹별 오류횟수 및 참조횟수

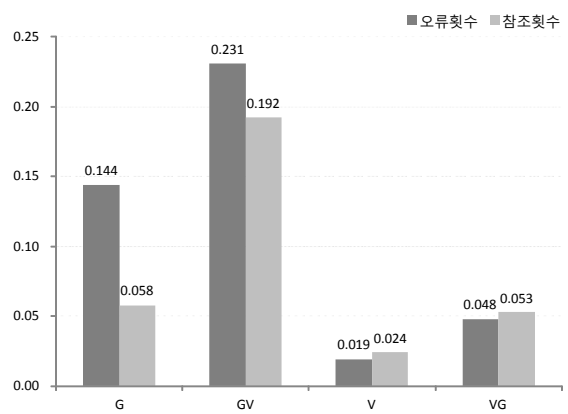
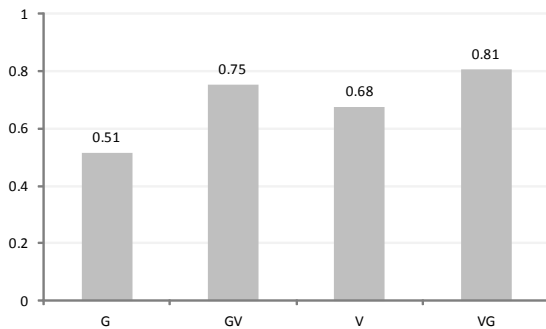


그림 6은 그룹별 R값(시각주의력 요구수준)을 나타낸 그래프로 그룹 간에 통계적으로 유의한 차이를 보였다( $H(3)=17.602$ ,  $p=0.001$ ). R값은 커질수록 더 많은 시각주의력을 요한다는 것을 의미하는데 그 값이 1에 가깝다는 것은 서터클래스의 단합(시야가 가린 상태 즉, 운전자가 도로를 주시한 상황)에 더 많은 영향을 받았다는 것을 나타낸다. VG그룹(‘양 조절’을 제스처로 조작한 그룹)의 R값이 0.81로 네 그룹 중

에서 1에 가장 가까운 값을 보였으며, G그룹의 R값은 0.51로 네 그룹 중에서 가장 낮게 나타났다. 또한, 음성에 비해 제스처 조작이 더 적은 시각주의를 요했으나 음성 및 제스처의 혼합조작 시에는 R값이 오히려 증가했다.

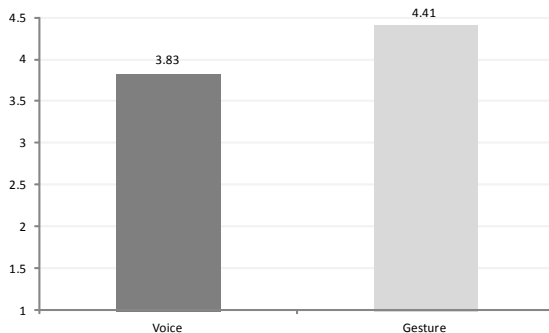
[그림 6] 그룹별 R값



### 3.2. '양 조절' 조작 분석

그림 7은 '양 조절' 조작에 한하여 조작시간을 나타낸 것인데, 음성과 제스처 간에 유의한 차이가 존재하였다( $U=14,100.5, p=0.000$ ). '양 조절'을 음성(V그룹, GV그룹) 또는 제스처(G그룹, VG그룹)로 조작한 경우, 평균 조작시간이 음성은 3.83초, 제스처는 4.41초로 0.58초 정도의 차이를 보였다.

[그림 7] '양 조절' 조작시간



반면, '양 조절' 조작의 오류시간에서는 음성과 제스처 간에 통계적으로 유의한 차이가 없었으나( $U=19,302.5, P=0.135$ ), 참조시간에는 유의한 차이가 존재하였다( $U=19,464.5, 0.044$ ). '양 조절' 조작의 오류시간과 참조시간 모두 제스처 조작이 음성 조작보다 각각 0.2초, 0.27초 더 짧게 나타났다(그림 8). 이는 앞서 살펴본 전체 오류 및 참조시간에 있어 제스처가 포함된 그룹이 음성만으로 조작하는 그룹보다 더 오래 걸렸던 것과 반대되는 결과이다(그림 4). 즉, 전체 오류 및 참조시간에 있어서는 제스처가 더 좋

지 않으나 '양 조절' 조작에 대해서는 제스처가 오류시간과 참조시간이 더 작았음을 알 수 있다. 차내 정보시스템의 조작 시 음성과 제스처 둘 다 요구하는 시각주의력이 크지 않으므로(그림 6 참고) 조작시간보다는 조작의 직관성에 관련된 오류나 참조시간이 안전운전에 미치는 영향이 더 클 것이다. 따라서 '양 조절' 조작에 제스처가 음성보다 다소 더 오래 걸렸음에도 불구하고 오류와 참조시간이 더 적었던 제스처 조작이 음성보다 '양 조절' 조작에서는 더 직관적이었던 것으로 판단된다.

[그림 8] '양 조절' 조작의 오류시간 및 참조시간

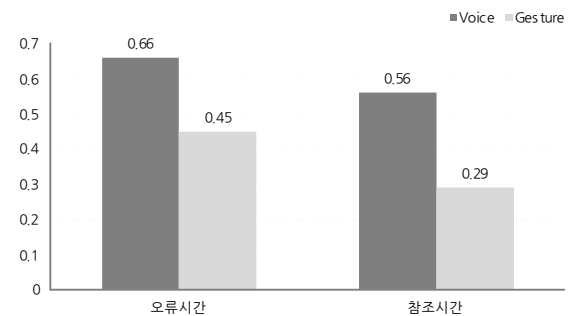
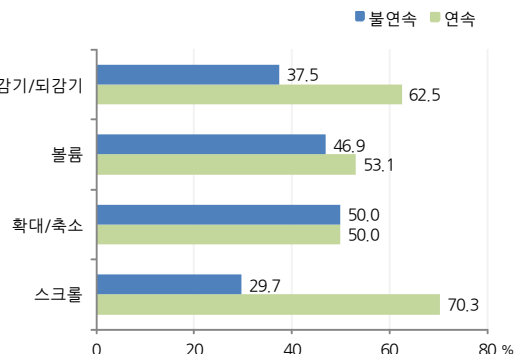


그림 9와 그림 10은 각각 음성과 제스처를 활용한 '양 조절' 조작에서 태스크 별로 피실험자들의 조작방법(연속/불연속)의 사용빈도를 조사하여 백분율로 나타낸 그래프이다. 그림 9를 보면 음성으로 조작하였을 때 확대/축소를 제외한 대부분의 태스크에서 많은 양을 한 번에 연속적으로 조작하는 피실험자들이 많았다. 예를 들어, '길안내 음량을 3칸 올려주세요.'와 같은 태스크가 주어질 경우, 많은 피실험자들이 "볼륨 업~~~"이라고 말하여 연속 조작으로 태스크를 수행하였으며, 일부 피실험자들은 "볼륨 업, 업, 업"이라고 말하여 불연속 조작으로 태스크를 수행하였다.

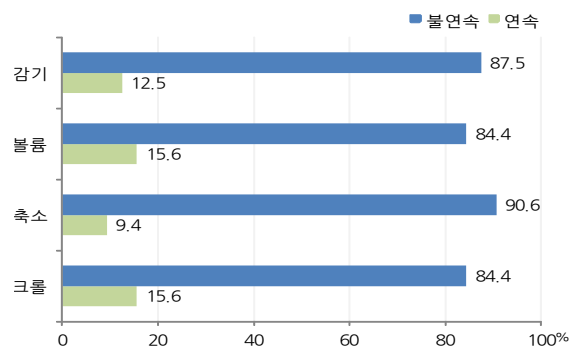
[그림 9] 음성 조작 시 태스크별

'양 조절' 조작방법의 사용빈도



반면, 그림 10을 보면 제스처로 조작하였을 때는 작은 양을 몇 번에 걸쳐 이동(불연속)하는 피실험자들이 확연하게 많은 것을 확인할 수 있다. 예를 들어, '길안내 음량을 3칸 올려주세요.'와 같은 태스크가 주어질 경우, 일부 피실험자들만이 검지를 오른쪽으로 길게 한번 움직임으로써 연속적인 조작으로 태스크를 수행하였고, 대부분의 피실험자들은 검지를 오른쪽으로 세 번 치는 불연속적인 조작으로 태스크를 수행하였다.

[그림 10] 제스처 조작 시 태스크별 '양 조절' 조작방법의 사용빈도



4. 결론

운전 시 차내 정보시스템의 조작은 도로로 향한 운전자의 시각주의를 뺏기 때문에 사고를 유발하는 중요한 요인 중의 하나이다. 따라서 시각주의를 덜 요구하는 조작방법인 음성이나 비접촉 제스처는 차내 정보시스템 조작에 그 활용가치가 높다. 본 연구에서는 음성과 비접촉 제스처, 그리고 그 둘의 혼합 조작에 기반한 조작 인터랙션의 유용성과 혼합의 적절한 방식을 알아보기 위한 실험을 수행하였다.

표 4는 실험 결과 중 각 척도 측면에서 차이가 나게 가장 좋았던 그룹을 표시하여 요약한 표이다. 본 연구결과에 따르면, 운전 중 음성, 제스처, 그리고 두 조작의 조합에 따른 차내 정보시스템의 조작 방식은 음성만 사용할 때(V)가 조작시간이 가장 짧고 오류도 적게 일어나는 것을 알 수 있었다. 그 다음으로는 명령 조작은 음성으로 하고 '양 조절' 조작은 제스처로 하는 혼합조작 방식(VG)이 만족도가 가장 높았고 오류시간도 짧았으며, 오류횟수 및 참조횟수가 적은 조합이었다. 반면에 '양 조절'을 음성으로 조작한 음성과 제스처의 혼합 조작 방식(GV)은 모든 면에서 가장 나쁜 수행도를 보였다. 특히 '양 조절' 조작은 음성으로 조작하는 것보다 제스처로 조작하였을 때 조작시간은 짧았으나 오류를 범하

는 시간과 조작 방식을 참조하는 시간이 더 짧게 나타났다. 이는 음성 제어장치들이 '양 조절'과 같은 연속적인 동작을 제어하는 데 적합하지 않다는 Wickens et al.(1985)의 연구결과와도 동일하다.

[표 4] 실험 결과 요약(좋은 결과 보인 그룹만 표시)

	V	G	VG	GV
조작시간	●			
만족도		●	●	
오류시간	●	●	●	
참조시간	●	●	●	
오류횟수	●		●	
참조횟수	●	●	●	
R값	○	●		
양 조절 조작시간	●			
양 조절 오류/참조시간		●		
양 조절 방식	연속	불연속		

●: 상대적으로 매우 좋은 결과, ○: 대체로 좋은 결과

하지만 비접촉 제스처를 활용한 '양 조절' 조작은 한 번의 제스처 동작으로 연속적인 이동이 일어나도록 할 경우 운전자는 변하는 양을 확인하기 위해 차내 정보시스템으로 지속적으로 시선을 빼앗기게 될 것이다. 따라서 '양 조절' 조작의 제스처는 한 번의 제스처 동작에 한 번의 이동씩만 일어나는 불연속적인 조작 인터랙션을 적용하는 것이 더 적합한 것으로 나타났다. 그리고 운전환경을 모사한 서터글라스의 영향을 확인하는 R값이 G그룹에서 가장 낮게 나타났다. 이는 운전 및 비운전에 따른 영향의 차이가 가장 적은 것을 의미하며 나아가 제스처 조작이 운전 시에도 적용가능한 조작방법이라는 것을 알려준다. 하지만 비접촉 제스처는 음성 명령어에 비해 직관성이 떨어지기 때문에 모든 조작을 제스처로 하게 되면 운전자는 학습과 기억에 대한 부담이 커질 것이다. 작업 시 기억 실패는 누구에게나 그리고 비교적 자주 발생할 수 있으며(Wickens et al., 2004), 이는 매우 큰 사고를 초래할 수도 있다(Rolt, 1978). 따라서 차량 내·외부에서 발생하는 다양한 노이즈로 인한 음성인식의 한계를 극복하고 제스처 조작의 학습 부담을 줄일 수 있도록 모든 명령 조작(예. 불륨)은 음성과 제스처를 함께 중복적으로 사용할 수 있고 '양 조절' 조작(예. 불륨 3으로 감소)은 불연속적인 제스처 조작으로 한정짓는 혼합 설계가 차내 정보시스템에 가장 적합한 인터랙션 설계로 보인다.

본 연구는 차량 내 운전환경을 모사하기 위해 실험실 환경에서 서터글라스를 사용한 시각차폐 기법



을 사용하였다. 그러나 실제 운전 환경은 음성의 경우는 소음이나 동석자와의 대화, 제스처의 경우는 운전, 기어 조작 혹은 대화 중 손동작 등에 의해 영향을 받을 것이므로 이러한 영향 하에서의 추가 연구가 필요하다. 또한 본 연구는 실제 음성과 비접촉 제스처 인식이 가능한 시스템이 아닌 오즈의 마법사 기법을 사용했다는 한계를 가지고 있다. 하지만 시각차폐와 오즈의 마법사 기법은 많은 연구들에서 실제 시스템의 설계를 모사하는데 그 유용성이 입증된 것이므로, 본 연구의 결과는 추후 자동차 차내 정보 시스템의 음성과 제스처의 혼합조작 연구를 위한 기초 자료로 활용될 수 있을 것으로 기대된다.

## 참고문헌

- Alpern, M. and Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. *CHI'03 extended abstracts on Human Factors in computing systems*.
- Dingus, T.A., Antim, J.F., Hulse, M.C. and Wierwille, W. (1988). Human Factors issues associated with in-car navigation system usage. *Proceedings of the 32nd Annual meeting of the Human Factors Society*, 1448-1453.
- Epps, J., Oviatt, S. and Chen, F. (2004). Integration of Speech and Gesture Inputs during Multimodal Interaction. *Proc Aust. Int. Conf. on CHI*.
- Gould, J.D., Conti, J. and Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter, *Communication of the ACM*, Vol.26, No.4, 295-308.
- Hong, K.H. (2004). Speech Based Multimodal Interface Technologies and Standards. *Journal of the Korean Society of Speech Sciences*, Vol.51, 117-135.
- ISO (2007). Road Vehicles - Ergonomic Aspects of Transport Information and Control Systems - Occlusion Method to Assess Visual Demand due to the Use of In-vehicle Systems. *ISO International Standard 16673*.
- Jung, Y.G., Lee, J.S. and Han, M.S. (2007). Development of recognizer using fusion of Voice and Gesture. *Conference of the Korean Institute of Information Technology*.
- Kim, H., Kang, S., Eom, M., Jung, D. and Cho, K. (2011). What are Korean Stereotypes on Non-Touch Operation Gestures?, 4th International Conference on Advanced Science and Technology for Materials, Manufacturing and Measurement, August 17-19, 2011, Seoul, Korea.
- Lee, M. and Billinghurst, M. (2008). A Wizard of Oz Study for an AR Multimodal Interface. *In Proceedings of ICMI'08. ACM, New York* 249-256.
- Loehmann, S., Diwischek, L. Schröer, B. Bengler, and K. Lindemann, U. (2011). The User Experience of Freehand Gesture.
- Malaterre, G. (1990). Error analysis and in-depth accident studies, *Ergonomics*. vol33, 1403-1421.
- Park, J.C. (2009). A User-driven Visual Occlusion method for Measuring the Visual Demand of In-Vehicle Information Systems (IVIS). *Journal of the Ergonomics Society of Korea*, Vol.28, No.3, 49-54.
- Rolt, L.T.C. (1978). Red for danger. *London: Pan Books*.
- Wickens, C. D. and Seidler, K.S. (1985). Information access, representation and utilization. In R. Nickerson(ed.). *Emerging needs and opportunities for human factor research. Washington, DC: National Academy of Sciences*.
- Wickens, C. D., Lee, John D., Liu, Yili D., Gordon Becker and Sallie E. (2004). An Introduction to Human Factors Engineering, 2nd Edition. *Pearson Education*.
- Wightman, D.C. and Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, 27(3), 267-283.