

운전 상황에서 휴대폰 음성인터페이스의 사용성 향상에 관한 연구

강윤환¹ · 정성욱¹ · 정가훈¹ · 최재호² · 정의승^{1*}

¹고려대학교 산업시스템정보공학과 / ²대진대학교 산업시스템공학과

Usability Improvement for the Speech Interface of Mobile Phones While Driving

Yun Hwan Kang¹ · Seong Wook Jeong¹ · Ga Hun Jung¹ · Jae Ho Choi² · Eui S. Jung¹

¹Department of Industrial Engineering, Korea University, Seoul 136-792, Korea

²Department of Industrial and Systems Engineering, Daejin University, Kyoungki 487-711, Korea

While driving, the manual use of a mobile phone is heavily restricted due to the interference with the primary driving task. An alternative would be the use of speech interface. The current study aims to provide a guideline to implementation of a speech interface to the mobile phone. To do so, an expert evaluation was made and it revealed that a speech interface requires less workload, less performance degradation of the driving task than that of the keypad interface. To make speech interfaces more usable, new improvements are suggested. Subjective workload can be reduced and user satisfaction can be improved without degrading the primary task performance, for instance, by letting the user interrupt the speech of the phone, eliminating the repetitive words, letting the user know clearly what makes an error, providing a way to go back to the previous state, reducing the usage of keypad buttons and reducing the amount of the information on the screen.

Keywords: Speech Interface, Mobile Phone, Usability, Subjective Workload

1. 서론

운전 중에는 주 태스크인 운전 작업 때문에 다른 작업을 하기
가 쉽지 않기 때문에 꾸준히 음성인터페이스(Speech interface)
의 사용이 제기되어 왔다. 특히, 내비게이션(Navigation)에서
음성인터페이스는 이미 널리 쓰이고 있으며 휴대폰의 경우도
전화걸기 기능 뿐 아니라 메시지 확인, 시간 확인, 부채중 전화
알림 등 다양한 기능에서 음성인터페이스가 적용되고 있다.

음성인터페이스를 사용한 기술은 이미 시각인터페이스와
함께 실용화가 가능한 기술로 인식되고 있으며(Park, 2006), 국
내 사용자들을 대상으로 한 기존 연구에서도 역시 전화기능
및 메시지 전송 기능, 문서 읽어주기 등의 기능에서 음성인터
페이스의 필요성이 제기되고 있다(Kang, 2008). 현재 출시되어

사용되고 있는 국내외 휴대폰들 중 국내에서는 한 회사의 제
품에서 전화걸기 기능이 음성인터페이스를 제공하고 있고, 해
외에서는 여러 회사의 제품들이 전화걸기, 메시지 확인, 전화
번호부 확인, 시간확인 등 다양한 기능에서 음성인터페이스를
제공하고 있다(Kang, 2008).

그러나 이렇게 운전 상황에서 음성인터페이스의 필요성이
부각되고 있음에도 실제 운전 상황에서 휴대폰 음성인터페이
스의 사용성에 대한 연구는 매우 미흡한 실정이다. 기존 연구
를 보면 대부분 주행 중 음성인터페이스 사용 시의 안전 문제
에 대해 시뮬레이션 상황에서의 실험적 평가가 대부분으로,
음성인터페이스가 터치패널이나 키패드를 이용한 인터페이
스보다 운전 상황에서 주 태스크인 운전 방해 주지 않는
다는 것이 주된 연구내용이다. 운전 상황에서 음성 다이얼링

* 연락처 : 정의승, 136-792 서울 성북구 안암동 5가 1번지 고려대학교 산업시스템정보공학과, Tel : 02-3290-3380, Fax : 02-3290-3913,
E-mail : ejung@korea.ac.kr

2008년 11월 12일 접수; 2008년 12월 15일 수정본 접수; 2008년 12월 16일 게재 확정.

시스템에서 튜토리얼 제공과 만족도에 관한 연구(Laurie, 1998), 음성 다이얼링 시스템의 메뉴구조 개선에 관한 연구(Laurie, 1999) 등이 있으나, 정작 음성인터페이스의 사용성에 대한 연구는 부족한 상황이다. 음성인터페이스의 사용성과 관련된 연구로는, 음성인터페이스 디자인 시 고려해야 할 항목들을 태스크 요구사항(task requirements), 기술적 지원 범위(Technological capabilities and limitations), 사용자 기대수준(User expectations and expertise)의 3가지로 정리한 Kamm(1995)의 논문과 음성만을 지원하는 시스템(Voice only system)에서 대화형 인터페이스를 구현하기 위해 필요한 고려사항들을 검토한 Yankelovich(1995)의 연구가 있다. Kamm(1995)의 연구결과는 일반적인 음성인터페이스를 지원하는 시스템에서의 고려사항에 대한 연구를 진행하였으나, 운전 상황과 같이 음성시스템 자체가 보조태스크로 수행될 때에 대한 고려가 되어 있지 않으며, Yankelovich(1995)의 연구 역시 보조태스크로 수행되는 경우에 대한 고려가 되어 있지 않고, 휴대폰과 같은 키패드입력 및 시각디스플레이를 동시에 지원하는 시스템에 대한 내용도 고려되지 않았다.

본 연구에서는 실제 차량을 운전하는 상황에서 현재 사용되고 있는 휴대폰의 음성인터페이스 기능에 대한 태스크 분석 및 주행 시 사용성 평가와 관찰을 통해 휴대폰 음성인터페이스의 사용성 문제점들을 파악하여 기존 연구에서 제시한 음성인터페이스의 고려사항들이 운전 상황에서도 동일하게 적용될 수 있는지 확인하고, 사용성을 향상시킬 수 있는 개선안을 제시하였다.

2. 음성인터페이스 디자인 고려사항

음성인터페이스를 지원하는 시스템에서 인터페이스 디자인

시 고려해야 할 사항을 Kamm(1995)은 태스크 요구사항, 기술적 지원 범위, 사용자 기대수준의 3가지로 분류하였으며, 음성인터페이스를 지원하는 시스템은 위와 같은 속성들을 고려해야 한다. 이에 따라 기존 휴대폰에서 사용하고 있는 음성인터페이스를 분석한 결과는 <Table 1>과 같다.

태스크 요구사항(Task requirements)에 있어서 휴대폰에서 사용자에게 요구하는 정보의 형태는 단어로서, 각 명령을 실행할 때, 'Call', 'Jane' 등과 같이 단어 형태로 정보를 주고받을 수 있다. 단어로 되어 있는 경우에는 애플리케이션의 종류에 상관없이 많은 수의 명령어를 처리할 수 있고, 기억하기 쉬운 장점이 있어 다양한 기능을 제공하는 휴대폰에 적합하다. 지원하는 인터페이스는 주로 음성입력과 키패드입력을 지원하고 있으며, 음성으로 입력을 할 수 있게 하면서 동시에 키패드를 통해 입력할 수 있다. 출력의 경우도, 음성과 동시에 화면으로 내용을 출력하게 되어 있다. 음성인터페이스만 가능한 시스템에서는 단기기억 메모리의 영향을 받게 되지만, 보조적으로 키패드 입력과 화면 출력을 지원함으로써 음성인터페이스만 지원되었을 경우의 한계를 보완해 줄 수 있다. 오류비용의 측면에서는 정보를 확인하는 태스크부터 비복원 기능을 실행하는 태스크까지 다양한 태스크가 있다. 전화를 거는 작업의 경우는 통화요금이 들어가는 작업으로서 오류에 따른 비용이 많이 들어가고, 이름을 말하여 전화번호를 확인하는 작업 같은 경우는 오류가 생기는 경우 시간적인 비용만이 발생하게 된다.

기술적 지원 범위(Technological capabilities and limitations)의 경우, 화자독립(speaker independent)의 단어 수준 음성인식이 가능하다. 태스크의 요구사항이 범용으로 쓰일 수 있어야 하는데, 단어 인식은 이에 적합한 방식이라고 할 수 있다. 그리고 음성 출력의 경우는 자연음성 출력(pre-recorded speech)과 조합 음성 출력(synthetic speech)을 동시에 지원하는데 각 태스크에서 정형화된 부분은 자연음성으로 녹음되어 이 음성을 출력하

Table 1. Current mobile phone's speech interface analysis according to the Kamm(1995)'s speech interface considerations

Kamm(1995)'s speech interface considerations		Mobile phone's speech interface characteristics
Task requirements	Information elements	Word
	Task modalities	Voice, keypad input Voice, display output
	Cost of interaction failures	Information retrieval and non-recovered action mixed
Technological capabilities and limitations	Voice input	Speaker-independent Word recognition
	Voice output	Pre-recorded speech and synthetic speech mixed
	System capabilities	Do not allow to talk-over
User expectations and expertise	Conversational speech behavior	Continuous speech for number only Do not allow to talk-over Do not allow to conversational dynamics Provide novice and expert mode

고, 전화번호부의 사람이름과 같이 특정지어지지 않은 정보에 대해서는 조합음성으로 출력한다. 시스템이 안내를 하는 중간에 사용자가 끼어들기가 가능하게 해주는 인터럽트(talk-over)는 현재 사용되는 휴대폰 음성인터페이스에서는 쓰이지 않는다.

마지막으로 사용자의 기대수준(User expectations and expertise)에 있어서는, 현재의 휴대폰에서도 일부 연속적인 발음(continuous speech)을 허용하는 부분이 있는데 이는 번호를 말할 때이다. 비록 또박또박 말하는 것이 가장 인식률이 좋기는 하지만 연속된 발음에 대해서도 인식하는 것을 확인하였다. 하지만 번호를 제외한 다른 음성입력에서는 연속된 발음을 인식할 수 없었다. 또한, 인터럽트(talk-over)는 시스템이 기술적으로 지원하지 않음으로 인해 지원을 하지 않았고, 휴지기간 중의 음성입력도 허용하지 않았고, 특정 구간에서 사용자 입력이 요구되는 곳에서만 음성입력을 허용하였다. 그리고 초심자와 숙련자모드를 두어 숙련자로 선택한 경우는 사용자가 음성으로 입력한 내용에 대해 확인절차를 거치지 않는 등의 옵션을 제공하였다.

3. 기존 휴대폰의 음성인터페이스 방식 비교분석

2007년부터 최근까지 제품들 중, 음성인터페이스를 지원하는 제품 현황을 보면 국내는 1개 회사의 제품이 음성인터페이스를 지원하며 국외는 4개 회사의 제품이 음성인터페이스를 제공하고 있다. 본 연구에서는 위에 열거한 모든 제품에서 제공하는 기본 기능인 전화걸기 기능의 음성인터페이스를 대상으로 연구를 진행하였다. 현재 판매되는 휴대폰에 대해 HTA(Hierarchical Task Analysis) 방식으로 태스크 분석을 하였는데 음성인터페이스 방식에 따라서 다음과 같은 2가지로 크게 분류할 수 있다.

A-타입의 경우 사용자의 명령어가 필요한 부분에서 경고를 주어 사용자의 피드백이 필요한 부분을 명확하게 알려주고, 현재의 메뉴 내용을 읽어 주어 어느 기능을 실행할 것인지를 명시적으로 알 수 있게 해준다. 그리고 찾는 후보가 없는 경우 찾는 명령이 없다는 안내와 함께 음성인터페이스가 종료된다. B-타입의 경우는 사용자의 명령어가 필요한 경우에 경고음 대신 안내멘트만 제공이 되고 현재 메뉴에 대한 피드백을

Table 2. Speech interface types of the current mobile phones

A-type	Provide the menu name with voice
	Provide the warning beep before user input
	When no candidate is matched, exit the speech interface
B-type	Do not provide the menu name with voice
	Do not provide the warning sound before user input
	When no candidate is matched, provide the way to go to the initial state of the speech interface

제공하지 않는다. 그리고 찾는 후보가 없는 경우에는 찾는 명령이 제대로 확인되지 않았다는 안내와 함께 음성인터페이스의 초기 단계로 이동할 것인지를 선택할 수 있다. 이 2가지 방식에 대해 'Jane이라는 사람에게 전화를 걸고 통화 확인 후 전화를 끊는다.'는 시나리오에 대한 태스크 진행 절차는 <Table 3>과 같다.

4. 사용성 평가 실험

태스크 분석을 통해 검토한 두 가지 방식의 음성인터페이스의 문제점 파악과 사용성 평가를 위해 실제 운전 상황에서 음성인터페이스를 사용하여 전화걸기 작업을 수행하는 실험을 수행하였다. 실험은 입력방식(음성, 키패드), 음성인터페이스 방식(A-타입, B-타입), 태스크 난이도(단순 태스크, 복합 태스크)의 3요인을 독립변수로 하는 Full Factorial, Within Subject Design으로 구성하였으며, 종속변수로는 차량 주행속도의 변화, 작업시간, 작업부하(Workload) 및 주관적인 만족도를 평가하였다.

음성 입력방식은 휴대폰의 버튼을 눌러 음성인터페이스를 동작시키고 난 후, 명령을 말하여 전화를 걸고, 키패드 입력방식은 휴대폰의 키 조작으로 전화를 건다. 키패드 입력방식은 본 연구의 목적은 아니지만 피실험자들이 음성인터페이스의 문제점을 키패드와 비교하여 쉽게 발견하기 위하여 추가하였

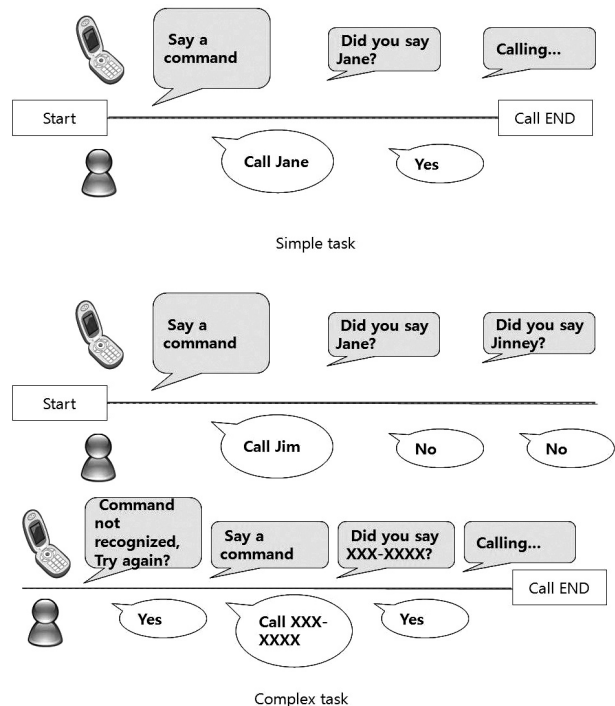


Figure 1. Simple and complex tasks used in the usability experiment

Table 3. Call task process according to the speech interface type

Step	A-type	B-type
1	user : press the keypad button to start speech interface	user : press the keypad button to start speech interface
2	mobile phone : start speech interface	mobile phone : start speech interface
2.1	mobile phone : say 'Call name or number'	
2.2	mobile phone : say 'Say command'	mobile phone : say 'Say command'
2.3	mobile phone : sound warning beep	
2.4	mobile phone : waiting for user's input	mobile phone : waiting for user's input
3	user : do 'Call' command	user : do 'Call' command
3.1	user : say 'Call Jane'	user : say 'Call Jane'
3.2	mobile phone : recognize 'Call' command. When fail to recognize, go to step 2	mobile phone : recognize 'Call' command. When fail to recognize, go to step 2
3.3	mobile phone : find the recipient's name in the phonebook when fail to find candidate, go to step 3.9	mobile phone : find the recipient's name in the phonebook when fail to find candidate, go to step 3.9
3.4	mobile phone : say 'Did you say Jane?'	mobile phone : say 'Did you say Jane?'
3.5	mobile phone : sound warning beep	
3.6	mobile phone : waiting for user's input	mobile phone : waiting for user's input
3.7	user : say 'Yes' When the candidate is not what he/she wants, say 'No' and go to step 3.3	user : say 'Yes' When the candidate is not what he/she wants, say 'No' and go to step 3.3
3.8	mobile phone : recognize user's input	mobile phone : recognize user's input
3.9	mobile phone : no candidate is matched	mobile phone : no candidate is matched
3.9.1	mobile phone : say 'Command not recognized.'	mobile phone : say 'Command not recognized.'
3.9.2	mobile phone : Exit speech interface	
3.9.3		mobile phone : say 'Try again?'
3.9.4		mobile phone : waiting for user's input
3.9.5		user : say 'Yes'
3.9.6		mobile phone : recognize user's input. when 'No', exit speech interface
4	mobile phone : making a call to the matched recipient	mobile phone : making a call to the matched recipient
4.1	mobile phone : say 'Now calling Jane'	mobile phone : say 'Now calling Jane'
4.2	mobile phone : making a call to the matched recipient	mobile phone : making a call to the matched recipient
5	user : making a call	user : making a call
6	user : end a call	user : end a call
6.1	user : press 'END' keypad button	user : press 'END' keypad button

다. 단순 태스크는 전화번호부에 등록된 사람에게 전화를 거는 것이며, 복합 태스크는 전화번호부에 등록되지 않은 사람을 찾게 하고 전화번호부에서 확인이 되지 않으면 번호를 입력하여 전화를 거는 작업으로 구성하였으며 각 단계는 <Figure 1>과 같다.

운전면허를 소지하고 시각 및 청각에 문제가 없는 20~35세 사이의 남자 10명이 실험에 참여하였으며, 실험순서는 학습효

과의 영향을 고려하여 무작위로 배치하였다. 실험은 자동변속기 승용차를 사용하였으며, 차량 내에 비디오 카메라를 설치하여 계기판과 휴대폰 화면을 촬영하였다.

각 피실험자는 실험 전에 설명을 듣고 연습을 통해 음성인터페이스의 사용법에 충분히 익숙해지도록 하였다. 우선 태스크 수행 없이 시속 20km를 유지하며 트랙을 돌고 난 후, 3요인에 변화를 주면서 최대한 시속 20km를 유지하며 태스크를 수

Table 4. ANOVA result on difference between driving speed without secondary task(call) and driving speed with secondary task(call)

Source	SSE	df	MSE	F	p-value
Speech interface type	7.021	1	7.021	4.426	0.065
Input type	17.984	1	17.984	7.565	0.022*
Task difficulty	1.352	1	1.352	0.789	0.398
Speech interface type * Input type	3.616	1	3.616	3.839	0.082
Speech interface type * Task difficulty	0.360	1	0.360	0.298	0.598
Input type * Task difficulty	1.083	1	1.083	0.986	0.347
Speech interface type * Input type * Task difficulty	0.001	1	0.001	0.001	0.979

행하도록 하였다. 안전을 고려하여 다른 차량과 사람이 없는 상황에서 20km의 속도를 유지하는 것을 주 태스크로 설정하였다. 각 태스크를 완료하고 난 후 설문지를 통해 해당 태스크의 작업부하와 주관적 만족도를 평가하였다. 작업부하 평가는 NASA TLX 기법을 사용하였으며, 각 NASA TLX의 항목에 대해 100점 척도로 작업부하를 평가하였고, 만족도의 경우도 100점 척도로 평가하였다. 그리고 비디오 분석을 통해 태스크가 없을 때의 주행속도에서 태스크 수행 시의 주행속도의 차이를 측정하였다. 태스크 수행 시 주행속도는 차량 속도 계기판을 3초 단위로 판독하여 평균으로 계산하였다.

5. 실험 결과

태스크가 없을 때의 주행속도에서 태스크 수행 시의 주행속도 차이에 대한 분산분석 결과가 <Table 4>와 같다.

주행속도 차이는 입력방식 간 유의한 차이가 있었으며($p = 0.022$), 음성인터페이스 사용 시 속도 저하가 키패드를 대비 적은 것으로 나타났다. 태스크 난이도의 차이와 음성인터페이스 방식 간에는 유의차가 없었다. 이러한 결과는 음성인터페이스가 키패드 방식에 비해 운전 에 미치는 영향이 작은 것으로 해석되며, 음성인터페이스 사용 시 태스크의 난이도나 음성인터페

이스 방식의 차이가 운전 에 미치는 영향에는 유의한 차이가 없다는 것을 보여준다. 특히, 음성인터페이스를 사용할 때보다 키패드를 사용할 때 작업속도 차이가 크다는 것은, 두 인터페이스의 차이점이 음성과 키패드를 사용한 입력이라고 할 때, 키패드를 사용하지 않는 것이 작업속도 차이를 크게 만들지 않는 요인이라고 할 수 있다.

작업시간에 대한 분산분석 결과는 <Table 5>와 같다.

작업시간은 태스크 난이도와 입력방식에 따라 유의한 차이가 있는 것으로 파악되었다. 복합 태스크 수행 시에 더 많은 시간이 걸리며, 음성인터페이스의 경우에 작업시간이 더 늘어났다. 그리고 입력방식과 태스크 난이도 간 교호작용은 어려운 태스크이면서 음성인터페이스를 사용한 것일수록 시간이 더 걸린다. 각 경우에 대한 pairwise t-test 결과 키패드 사용 시 단순태스크와 복합 태스크의 평균 시간차는 7초였으나, 음성인터페이스 사용 시 단순태스크와 복합 태스크의 평균 시간차는 22초로 음성일 경우 복합 태스크의 시간이 더 오래 걸린 것으로 분석되었다.

작업부하에 대한 분산분석 결과는 <Table 6>과 같으며, 키패드가 음성인터페이스보다, 복합 태스크가 단순 태스크보다 작업부하가 큰 것으로 분석되었으며, 입력방식과 음성인터페이스 종류간의 교호작용이 유의하게 나타났다. 키패드를 사용했을 때 복합 태스크의 작업부하가 높게 나타난 것은 당연한 결

Table 5. ANOVA result on the task completion time

Source	SSE	df	MSE	F	p-value
Speech interface type	6.613	1	6.613	0.233	0.641
Input type	1891.512	1	1891.512	139.124	0.000*
Task difficulty	2796.612	1	2796.612	617.467	0.000*
Speech interface type * Input type	4.513	1	4.513	0.358	0.564
Speech interface type * Task difficulty	56.113	1	56.113	3.295	0.103
Input type * Task difficulty	1029.612	1	1029.612	52.722	0.000*
Speech interface type * Input type * Task difficulty	117.612	1	117.612	5.380	0.046*

Table 6. ANOVA result on subjective workload

Source	SSE	df	MSE	F	p-value
Speech interface type	911.250	1	911.250	4.609	0.060
Input type	6125.000	1	6125.000	21.190	0.001*
Task difficulty	1786.050	1	1786.050	24.599	0.001*
Speech interface type * Input type	911.250	1	911.250	13.494	0.005*
Speech interface type * Task difficulty	3.200	1	3.200	0.236	0.638
Input type * Task difficulty	61.250	1	61.250	0.895	0.369
Speech interface type*Input type*Task difficulty	20.000	1	20.000	0.399	0.543

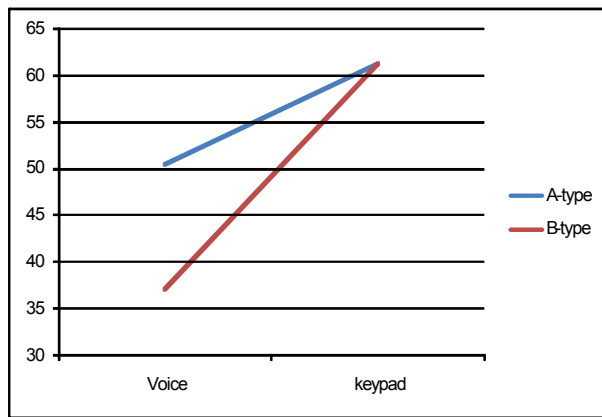


Figure 2. Interaction on subjective workload between input type(voice and keypad) and speech interface type (A-type and B-type)

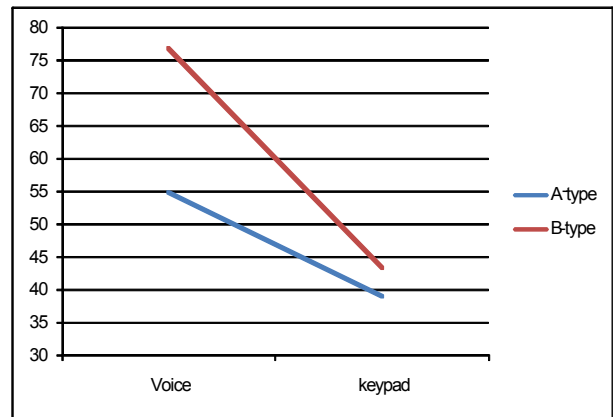


Figure 3. Interaction on user satisfaction between input type (voice and keypad) and speech interface type(A-type and B-type)

과로 생각되나, 입력방식과 음성인터페이스 종류간의 교호작용이 유의하게 나타난 것은 <Figure 2>에 나타난 바와 같이 키패드 방식에서는 A-타입과 B-타입의 작업부하가 각각 61.25로 동일하나(p = 1.000) 음성인터페이스 방식에 따른 작업부하의 차이가 50.5와 37.0으로 크기 때문(p = 0.001)으로 판단된다. 즉, 음성인터페이스 방식에서 B-타입이 A-타입에 비해 작업부하 부담이 적다.

주관적 만족도 분석결과는 <Table 7>과 같으며, 키패드보다

음성인터페이스가, A-타입보다 B-타입 음성인터페이스가, 복합 태스크보다 단순 태스크가 만족도가 높은 것으로 나타났다. 입력방식과 음성인터페이스 종류간의 교호작용이 유의하게 나타났는데, <Figure 3>에 나타난 바와 같이 키패드 방식에서는 제품 간의 만족도 차이가 각각의 평균이 39.0과 43.25로 작으나(p = 0.396) 음성인터페이스 방식에 대해서는 54.75와 76.75로 만족도의 차이가 크기 때문에(p = 0.001) 입력방식과 음성인터페이스 종류간의 교호작용이 유의하게 나타난 것으

Table 7. ANOVA result on user satisfaction

Source	SSE	df	MSE	F	p-value
Speech interface type	3445.313	1.000	3445.313	6.200	0.034*
Input type	12127.813	1.000	12127.813	24.153	0.001*
Task difficulty	2702.812	1.000	2702.812	13.559	0.005*
Speech interface type * Input type	1575.313	1	1575.313	10.529	0.010*
Speech interface type * Task difficulty	2.813	1	2.813	0.018	0.896
Input type * Task difficulty	52.813	1	52.813	0.303	0.595
Speech interface type * Input type * Task difficulty	70.312	1	70.312	0.356	0.565

로 볼 수 있다.

위의 결과를 종합해 볼 때 키패드 방식보다는 음성인터페이스가 운전 상황에 더 안전하고 만족도도 높으며, 키패드 사용 시에는 휴대폰 종류가 다르더라도 방식이 거의 비슷하기 때문에 만족도나 작업부하의 차이가 없으나, 음성인터페이스 사용 시에는 인터페이스 차이가 만족도와 작업부하에 영향을 주는 것으로 음성인터페이스 설계 시 사용성에 대한 고려가 중요함을 의미한다. 특히, 키패드에서는 상대적으로 음성인터페이스에 비해 키패드를 사용하는 빈도가 높기 때문에 음성인터페이스보다 속도 저하가 큰 것으로 볼 수 있다. 반면, 태스크의 난이도 차이에 대해서 운전이 미치는 영향이 미미하다는 것은 오류 상황 즉, 하고자 하는 태스크가 만족되지 않았을 때 추가적인 작업을 하는 것은 운전하는데 별 영향이 없는 것으로 볼 수 있다.

또한 음성인터페이스 A-타입과 B-타입 간에 주행속도의 차이는 없었고, 작업부하와 만족도에는 차이를 나타내고 있는데, 이는 음성인터페이스의 설계가 운전 작업에 큰 영향을 주지는 않지만 작업부하 및 만족도에는 영향을 준다는 것을 의미한다. B-타입의 특징이었던 안내멘트 후 경고음을 제공하지 않는 점, 메뉴를 음성으로 읽어주지 않는 점, 초기단계로 이동하는 방법을 제공한 점 때문에 작업부하를 감소시키며 만족도를 증가시켰다고 점을 알 수 있다.

6. 기존 음성인터페이스 문제점 분석

기존 휴대폰 음성인터페이스의 문제점 분석과 개선안을 찾기 위해 실험 시 촬영된 비디오의 분석과 기존 Yankelovich(1995) 및 Kamm(1995)의 연구에서 제안된 음성인터페이스 가이드라인을 사용하였다. 비디오 분석과 가이드라인 적용을 통해 파악된 문제점과 이에 대한 개선안은 다음과 같다.

6.1 휴대폰의 조합 음성 이해

휴대폰에 음성으로 명령을 실행하는데 의도한 것과 다르게 인식하여 다른 기능을 실행하는 경우 혹은 알아듣지 못하는 경우가 발생하였다. 태스크 단계에서는 <Table 3>에서의 3.2, 3.4, 3.9.1 단계에서 에러가 발생하였다. 3.2부분에서는 Call 명령어 자체를 잘못 인식하는 경우가 있었으나 음성인터페이스의 인식을 문제이므로 본 연구에서는 제외하기로 한다. 3.4의 경우는 음성인터페이스 시스템이 화자가 말한 사람이름에 대해 에러를 줄이기 위해 여러 대안을 내도록 되어 있다. 따라서 일반적으로 2, 3개의 후보 이름을 말하도록 되어 있고, 가장 확률이 높은 이름을 먼저 확인하는 방식이다. 여기서는 휴대폰이 말해주는 이름이 불명확하여 에러를 일으키는 경우가 있었다. 이 경우는 음성인터페이스 시스템의 음성출력이 조합음성이기 때문에 제대로 인식을 못하는 경우이며, 이런 조합음성의 경우는 자연음성에 비해 음의 높낮이와 운율이 다르게 되

어 있으므로 인지에 더 많은 부하가 걸린다(Fuestel, 1983). 이런 경우에는 조합음성에 대해 대화체에서와 같은 발음을 적용하거나(Yankelovich, 1995), 컨텍스트에 맞는 메시지를 함께 주어 해결 가능하다(Kamm, 1995).

6.2 음성 안내 멘트 중 사용자의 명령어 입력

시스템의 에러멘트에 익숙해짐에 따라 휴대폰의 안내가 완료되지 않았는데도 대답을 하려는 경향을 보였다. 특히 같은 안내멘트가 계속 반복되는 부분에서는 미리 대답을 하려는 현상이 두드러졌으며, 에러가 있는 경우에도 미리 대답을 하려는 현상이 있었다. 이는 음성인터페이스가 대화형이 가장 자연스럽게기 때문에 발생하는 문제로서 사용자가 중간에 인터럽트하려고 하기 때문에 생기는 문제이다(Yankelovich, 1995). 전화서비스 상담원과 고객 간의 대화를 분석한 한 연구에 의하면 약 12%의 대화가 상담원과 고객 사이에 동시에 말하여지는 것으로 밝혀졌다(Karis, 1991). 이렇게 안내말 중간에 사용자가 대답하는 문제에 있어서는 안내 중간에도 사용자의 답변을 인식하게 하여 해결할 수 있다. A-타입의 경우, 이전 연구자들이 제안한 방식인 안내 종료 시점을 경고음으로 알려주는 방식(Stifelman, 1993)을 사용하고 있었으나 사용자가 경고음 이후에 대답을 하게 만드는 데는 실패한 것으로 보인다. 오히려 경고음 없이 안내말로 구성된 B 제품의 경우가 오히려 작업부하가 적었고, 만족도도 높았다. 따라서 아예 안내말 중간에 사용자의 간섭을 허용하는 인터페이스 디자인이 자연스러울 수 있다.

6.3 에러에 대한 안내 멘트가 부적절함

<Table 3>의 3.9.1의 경우는 찾는 이름이 없는 경우인데 안내멘트가 적절하지 않았다. 'Command not recognized'는 명령어가 인식되지 않았다는 뜻인데 실제로는 이미 지금까지 사용자가 요구한 작업을 충실히 하고 있었는데 사람을 찾지 못한 것이므로 적절하지 않은 안내멘트이다.

오류가 있는 경우에는 오류에 대한 자세한 정보를 제공하도록 한다. 현재 제공되는 휴대폰에서는 'Command not recognized'라는 에러 메시지만을 제공하고 있는데 찾는 사람이 없는 경우에는 'No name found. Try again?' 등으로 무슨 이유로 찾을 수 없는지를 명확하게 알려주어야 한다. Yankelovich(1995)는 에러에 대해 적절한 안내멘트를 제공하는 것이 중요하다고 지적하였고, 특히 에러가 여러 번 반복되는 상황에 대해서는 각 단계에 대해 각기 다른 안내멘트(Progressive assistance)를 주어 시스템이 에러에 대해 적극적으로 대응하고 있다는 인상을 주어야 시스템을 신뢰하게 된다고 하였다. 여러 에러 상황에 대해 똑같은 안내멘트를 주는 것은 사용자의 불만을 가중시키며 잘못했을 때의 좌절감(Frustration level)을 키우게 된다.

6.4 초기 단계 혹은 이전단계 진입 방법이 없음

복합 태스크에서는 전화번호부에서 사람 이름을 찾아서 없 는 경우에는 다시 번호를 부르는 단계로 진입해야 하는데 처음 으로 돌아갈 수 있는 방법을 제공하지 않았다. <Table 3>의 3.8부분이 이런 부분이며, A-타입의 경우는 이 단계에서 명령 이 제대로 인식되지 않는다는 멘트 후에 바로 음성인터페이스 기능을 종료한다. 또한, 3.6부분에서 제품 B-타입의 경우 ‘Cancel’ 이라는 기능을 제공하여 음성인터페이스를 취소하는 기능을 제공한다. 피실험자 중 일부는 이 단계에서 미리 화면 에서 확인하고 ‘Cancel’ 명령을 실행하여 처음부터 다시 시작 하고자 하였는데 종료가 되어버려 당황하는 모습을 볼 수 있었다. 기존 연구에서도 음성인터페이스를 제공할 때 초기 혹은 이전 단계로의 진입을 두는 것을 권장한다(Laurie, 1999; Yankelovich, 1995). 즉, 태스크 중간에 처음으로 돌아갈 수 있는 옵션을 제공하고, 이전 단계로 돌아갈 수 있는 옵션도 제공한다. A-타입은 두 가지 모두 제공하지 않았고, B-타입은 처음으로 돌아갈 수 있는 옵션을 제공하고 있었으나 일부 단계에서만 이를 제공하고 있고 모든 단계에서 제공하고 있지 않았다.

6.5 키패드 버튼 누름

<Table 3>의 태스크 순서를 보면, 음성인터페이스를 시작할 때와 다시 시작하는 경우 키패드의 버튼을 사용하도록 되어 있다. 3.6의 경우나, A-타입의 경우가 그러한 경우였고, 버튼을 누르는 빈도가 증가하면 속도가 느려지고 음성으로 명령하는 것보다 더 많은 작업부하를 가지게 된다. 실제 A-타입의 경우는 복합 태스크에서 음성인터페이스를 종료하고 다시 실행해야 했기 때문에 키패드를 누르는 작업을 더 많이 하게 되었고, 이 때문에 B-타입 보다 작업부하가 높고 만족도가 낮은 원인이 된 것으로 추정된다. 특히, 위의 실험에서 작업 속도의 차가 음성인터페이스를 사용할 때보다 키패드를 사용할 때에 높았다는 점을 보면, 가급적 키패드를 사용하지 않는 단계를 만드는 것이 성능에 영향을 주지 않으면서 사용성을 높일 수 있는 방법임을 알 수 있다. Stiefmann(1993)은 다양한 방식의 입력 방식을 제공하는 것이 서로 단점을 보완할 수 있어 적절하다고 하였으나, 운전 상황에서는 운전을 방해하기 때문에 맞지 않는다. 따라서 사용자가 의도한 태스크가 완료되지 않은 상황에서는 키패드를 사용하지 않은 채, 태스크를 수행할 수 있는 방식을 제공해야 한다.

6.6 키패드에 비해 작업시간이 오래 걸림

정보를 안내멘트를 통해 알려주어야 하기 때문에 음성이 키패드보다 작업시간이 오래 걸려 태스크를 반복하게 되면서 같은 안내멘트를 들려주는 것을 답답하게 생각했다. 이 경우 안내멘트를 가능한 줄여줌으로써 이런 문제점을 해결할 수 있

다. 반복되는 안내말의 경우는 의미 없는 반복 문구를 줄임으로써 사용자에게 실제 의미 있는 안내만 줌으로써 해결할 수 있다(Yankelovich, 1995). 예를 들면, ‘Did you say call XXX?’ 라는 말을 여러 번 반복하게 된다면, 첫 번째 후보이름 이후에는 ‘Call XXX?’로 줄여서 안내를 할 수 있다.

6.7 전방을 주시하지 않고 휴대폰 화면을 자주 확인함

음성인터페이스를 사용하는 경우에도 휴대폰 화면을 주시하는 빈도가 높았다. 주로 명령이 제대로 실행되는지 확인하거나 에러가 발생한 경우에 화면을 보게 되었으며 음성으로 응답을 주는 것이 불확실한 경우에 화면을 주시하는 것을 확인할 수 있었다. 특히 <Table 3>에서 3.4 부분의 전화하고자 하는 대상을 확인하는 부분에서 화면을 자주 보는 것으로 확인되었다. 합성음의 부자연스러움과 여러 후보를 말해주는 부분이기 때문에 화면주시 빈도가 높은 것으로 볼 수 있다. Stiefmann(1993)의 연구에서는 화면출력이 음성인터페이스의 단기기억속성 때문에 생기는 문제점을 보완할 수 있는 좋은 방식이라고 하였다. 그러나 운전 상황에서는 화면을 보는데 필요한 시간을 줄여야 하므로 제한적으로 화면 출력을 이용해야 한다. 완전히 화면을 주시하는 것을 줄일 수 없다면 화면 출력의 장점을 극대화하면서 사용자의 불안감을 해소할 수 있는 방안을 찾는 것이 좋을 것이다. 이 문제점을 해결하기 위해서는 화면 구성을 바꾸는 것이 좋은 대안이 될 수 있다. 현재는 A-타입, B-타입 모두 음성인터페이스 메뉴나, 전화 화면 등이 일반 키패드로 조작할 때와 같은 인터페이스 화면을 제공한다. 실제로는 음성인터페이스를 쓰게 되는 운전 상황 같은 경우에는 안전성이 우선이기 때문에 오래 화면을 볼 수도 없고, 자세히 볼 수도 없으므로 일반 키패드를 이용하는 상황과 같은 인터페이스를 보여주는 것은 위험하다. 가급적 화면에 나타나는 정보량을 줄이고 잠시 보더라도 즉시 확인이 가능하도록 중요 정보를 크게 보여주도록 <Figure 4>와 같이 개선할 수 있다.

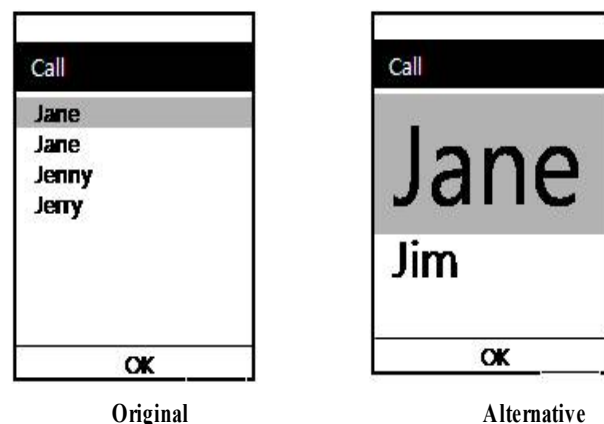


Figure 4. Candidates screen layout

Table 8. Speech interface improvements summary compared to the original interface

Improvement items	Alternative	Original
Reduce the usage of keypad	When choosing the right candidate, 'Cancel' goes to the initial state	When choosing the right candidate, 'Cancel' terminates the speech interface and press the keypad button to start the speech interface again
Reduce the amount of information and enlarge the screen information size	The candidate screen shows 2 candidates in the screen and enlarge the selected candidate line size	The candidate screen shows 7 candidates in the screen and the selected candidate size is same as others
Provide the detailed information of error	When no candidate is matched, provide the message, 'No name found. Try again?'	When no candidate is matched, provide the message, 'Command not recognized. Try again?'
Provide the way to go back to the previous state and the initial state	When the mobile phone is waiting for user input, provide the command, 'Cancel' and 'Exit'	When the mobile phone is waiting for user input, there is no command like 'Cancel' and 'Exit'
Eliminate the repetitive words	When candidates are 2 or more, provide the message, 'Did you say XXX?' first and provide 'Call XXX?' from the second query	When candidates are 2 or more, provide the message, 'Did you say XXX?' first and afterwards
Talk-over	Allow the user to interrupt the mobile phone's speaking	Do not allow the user to interrupt the mobile phone's speaking

위에 제시된 문제점들에 대한 개선안을 기존안과 대비하여 정리하면 <Table 8>과 같다.

7. 개선안 검증실험

위의 개선안에 대한 검증 실험을 앞에서와 동일한 실험환경 및 방식으로 수행하였다. 기존안과 개선안을 비교하였는데 기존안은 앞의 실험에서 작업부하가 적고 만족도가 높은 B-타입의 인터페이스를 사용하였으며, 개선안은 Wizard of OZ 방식 (Klemmer, 2000)으로 구현하였다. 앞의 실험과는 달리 차량에서 음성인식 안내시스템을 실제로 사용하고 있거나 사용해 보았던 사람들을 피실험자로 선정하였다. 새로운 음성인터페이스에 대한 검증실험이므로 사용경험이 없는 피실험자의 경우 기존 인터페이스와의 차이를 느끼지 못할 것으로 판단되어 제

외하였다. 운전면허를 소지하고 시각 및 청각에 문제가 없는 20~35세 사이의 남자 5명을 대상으로 개선안 검증실험을 수행하였다. 기존안과 개선안에 대해 단순 태스크와 복합 태스크를 수행하였고, 주행속도의 차이와 작업시간, 작업부하, 주관적 만족도를 평가하였다. 각각에 대한 pairwise t-test 분석결과는 <Table 9>와 같다.

주행속도 차이는 유의한 차이가 나타나지 않았으며($p = 0.296$), 작업부하 및 만족도는 개선안이 기존안보다 우수한 것으로 분석되었다. 주행속도 차에 있어서 유의한 차가 나타나지 않은 이유는 20km의 저속도 주행이었기 때문인 것으로 판단된다. 이전 실험에서는 키보드 조작방식과 음성인터페이스를 비교하여 차이를 보였는데, 이번 실험에서는 크게 보아 동일한 구조를 가진 음성인터페이스 간에는 속도차이가 없는 것으로 볼 수 있다. 작업시간의 경우, 대안으로 제시된 항목 중 키패드 사용을 최소화하는 부분과 이전, 초기화면으로 이동하

Table 9. Pairwise t-test summary

	Mean of the difference	95% confidence interval		t	df	p-value
Difference of speed between task and no task	0.483	-0.502	1.468	1.108	9	0.296
Task completion time	-9.6	-13.075	-6.125	-6.249	9	0.000*
Workload	9.998	5.621	14.374	5.167	9	0.001*
User satisfaction	-14	-25.778	-2.221	-2.688	9	0.025*

는데 음성인터페이스를 사용하여 기존안보다 시간이 더 걸리는 것을 확인할 수 있었다. 반면, 음성인터페이스의 개선을 통해 작업부하 및 만족도는 향상시킬 수 있었다.

8. 결론 및 토의

본 연구는 차량을 이용한 실제 운전 상황을 통해 휴대폰 음성 인터페이스의 사용성을 분석하고, 개선안을 도출했다는 점에 의의를 가진다. 또한, 기존 연구결과를 통해서 알려진 음성 인터페이스의 설계가이드(Yankelovich, 1995; Kamm, 1995)가 운전 상황의 휴대폰에 접목된 음성인터페이스에서도 적용될 수 있는지 확인할 수 있었다. 본 연구의 결과를 요약해 보면, 음성 인터페이스 디자인 시 사용성 개선을 위해 사용자의 안내말도중 인터럽트를 허용하고, 의미 없는 반복 문구를 생략하며, 자세한 오류정보를 제공, 처음 및 이전단계로 돌아가는 방법을 제공, 휴대폰 화면의 정보량 최소화/크기 확대, 키패드 버튼 사용 최소화 등을 고려할 수 있음을 알 수 있다. 특히, 키패드 입력의 경우는 Stifelmann(1993)의 연구에서는 음성인터페이스를 보조하는 수단으로 효과적으로 사용될 수 있음을 전제하였으나, 운전 상황에서는 사용이 적합하지 않은 것으로 보이므로 사용을 자체해야 할 것이며, 동일 연구 내용 중 화면 출력의 경우도 일반적인 음성인터페이스의 보조수단으로 활용하는데 있어서 정보량을 줄이는 노력을 기울여야 할 것이다. 키패드입력보다 음성인터페이스가 운전성능에 적은 영향을 미치지만 일정 정도 운전성능에 영향을 주므로(Lee, 2001), 이와 같이 사용성을 개선하기 위한 고려는 더욱 필요하다.

여러 인터페이스 고려사항 중 작업부하 및 만족도를 개선하는데 있어서 그 주요인은 오류정보를 자세하게 제공하는 것과 처음 및 이전단계로 돌아갈 수 있는 방법을 제공하는 것으로 생각된다. 이는 음성인터페이스의 기본적인 속성이 단기기억에 의존하고 있으며, 한 번 실행된 명령은 반복하기 힘들기 때문에 에러를 잘 제어하지 못하면, 사용자가 음성인터페이스를 잘 제어하고 있다고 생각하기 힘들고, 인지부하를 증가시킨다. 때문에 이런 에러를 이해하기 쉽게 하고, 복구할 수 있게 해주는 것은 음성인터페이스를 사용함에 있어 에러가 발생하더라도 목적하고자 하는 태스크를 포기하지 않고 수행하여 음성인터페이스를 사용하기 쉽게 만든다고 볼 수 있다. Laurie (1998)의 연구에서도 이를 확인할 수 있는데, 음성인터페이스를 사용하기 전 충분한 소개를 받은 집단이 그렇지 않은 집단보다 음성인식의 인식률이 떨어지는 상황에서 더 높은 만족도를 보여주었다. 향후에는 이와 같이 음성인터페이스가 에러를

복구하기 쉽게 디자인되어야 할 것이다.

실험환경 디자인 시에 실제 차를 실험에 사용하면서 사고의 위험을 줄이기 위해 시속 20km 로 제한을 두었는데, 실험결과로 보았을 때는 속도가 더 빠른 상황이 되면 음성인터페이스의 대안을 어떻게 만드느냐에 따라 더 큰 차이가 있을 것으로 예상된다. 추후 일반적인 운전 상황에 더 가깝게 환경을 구성하여 조사할 필요가 있다.

앞서 본 현재의 제품들의 음성인터페이스는 사용성 측면에서 발전의 여지가 많은 것이 사실이다. 본 연구결과를 통해 향후 휴대폰의 음성인터페이스의 사용성을 향상시키는데 도움이 될 수 있을 것이다. 그리고 본 연구는 전화걸기 기능에 대해서만 분석을 하였으나, 기본적인 음성인터페이스의 가이드로서 휴대폰의 다른 음성인터페이스 기능에도 확대해서 적용해 볼 수 있을 것이다.

참고 문헌

- Fuestel, Luce, P., et al. (1983), Capacity demands in short-term memory for synthetic and natural speech, *Human factors*, **25**(1), 17-32
- Kamm, Candace (1995), Voice interfaces for voice applications, *Proc. Natl. Acad. Sci. USA*, **92**, 10031-10037.
- Kang, Yun Hwan, et al. (2008), Research on mobile handset's optimal voice recognition functions while driving, *Proc. 2008 Spring Conference of Ergonomics Society of Korea*.
- Karis, D. and Dobroth (1991), Automating services with speech recognition over the publicswitched telephone network : human factors considerations, *IEEE Journal on Selected Areas Communications*, **9**, 574-585.
- Klemmer, Scott R, et al. (2000), SUEDE : A Wizard of Oz Prototyping Tool for Speech User Interfaces, *CHI Letters*, **2**(2).
- Laurie, Nancy E., et al. (1998), The role of instruction sets in operator satisfaction while using a voice activated dialing system, *International Journal of Industrial Ergonomics*, 307-312.
- Laurie, Nancy E, et al. (1999), Case study : Increasing usability of voice activated dialing systems, *International Journal of Industrial Ergonomics*, 621-627.
- Lee, John D., et al. (2001), Speech-based Interaction with In-vehicle Computers : The Effect of Speech-based E-mail on Drivers' Attention to the Roadway, *Human Factors : The Journal of the Human Factors and Ergonomics Society*, **43**(4), 631-640.
- Park, Jun Serk (2006), Five senses' information processing technology of the next generation human interface, *Weekly Technology Report 1252*, ETRI.
- Stifelmann, Lisa J., et al. (1993), VoiceNotes : A Speech Interface for a Hand-Held Voice Notetaker, *ACM INTERCHI*, **93**, 179-186.
- Yankelovich, Nicole, et al. (1995), Designing SpeechActs : Issues in Speech User Interfaces, *ACM SIGCHI '95 Proceedings*, 369-376.