음성/키 패드를 이용한 한글 단어 입력용 멀티모달 인터페이스

Multimodal interface for Korean inputs using speech and keypad

김원우, Wonwoo Kim*, 전호현, Hohyun Jeon**, 박성찬, Sungchan Park***
*KT 미래기술연구소, **KT 미래기술연구소, ***KT 미래기술연구소

요약 멀티모달 인터페이스(multimodal interface)는 사람과 기계 사이의 통신을 위해 여러 가지 수단을 사용함을 말한다. 본 고에서는 휴대폰 키 패드를 통한 문자 입력과 마이크를 통한 음성 인식의 두 가지 모드를 함께 사용하여 단어를 입력하는 새로운 인터페이스 방법을 제시함으로써 미래지향적 휴먼 인터페이스의 핵심으로 인지되고 있는 음성인식의 한계, 특히 한국어 인식의 문제점을 해결하고자 한다.

핵심어 : Multimodal, Interface

1. 서론

디지털 융합 추세에 부응해 휴대폰, PDA, PMP, 내비게이션, UMPC 등 다양한 휴대용 단말기가 출시되고 있다. 이러한 단말기에서 이용할 수 있는 정보 및 콘텐츠도 매우 다양화, 복합화, 대형화되고 있어서 이를 사용하기 위한 인터페이스도 필연적으로 복잡성이 요구된다.

그러나 출시되고 있는 단말기는 오히려 경량화, 슬림화, 단순화되는 추세여서 사용자들이 단말기를 통하여 정보를 검색하고 콘텐츠를 이용하기가 쉽지 않다.

이러한 불편함을 극복하기 위하여 다양한 사용자 인터페이스들이 나오고 있지만 기존의 비주얼 기반 인터페이스로는 그 한계를 극복하기가 쉽지 않다.

따라서 복잡한 다단계 메뉴 구조를 한 번으로 탐색할 수 있는 음성 인터페이스(VUI: Voice User Interface)에 대한 관심이 높아지고 있지만 낮은 인식 성공률과 새로운 인터페이스에 대한 학습 장벽 등으로 쉽게 자리를 잡지 못하고 있는 실정이다.

본 고에서는 음성인식의 대표적인 기술적 문제로 인식되고 있는 시작점 검출과, 특히 한글 인식을 더욱 어렵게 만드는 파찰음 인식의 문제를 키 패드와 음성인식을 조합함으로써 해결하고자 한다.

2. 문자 입력 인터페이스

2.1 키 패드

대표적인 개인용 휴대단말은 휴대폰으로서 키 패드를 통하여 메뉴를 선택하고, 필요한 문자를 입력한다. 특히 한글을 입력하기 위해서는 미리 정의된 한글 입력 시스템을 사용해야 하는데, 대표적인 한글 입력 체계로는 S사의 천지인, L사의 나랏글(EZ한글), P사의 스카이한글이 있다.

천지인은 그림 1과 같이 12개의 자음을 2~3개씩 묶어서 숫자 4~0에 해당하는 자판에 배열하고 모음은 'ㅣ', '•', 'ㅡ'를 각각 1, 2, 3에 배열한다.

자음은 자판를 두드리는 순서대로 해당하는 자음이 표시된다.(예 : 4번 자판의 경우 ㄱ, ㅋ, ㄲ 순으로 화면상에 표시) 모음은 'ㅏ'의 경우 'ㅣ'와 '•'을 순서대로 두드리면 된다. 받침의 조합은 한글 자동 장치에 의해 처리된다.

그림 1. 천지인 키 패드

이러한 키 패드 류는 하나의 글자를 입력하기 위해 자판을 여러 번 눌러야 하기 때문에 번거롭고 입력시간이 오래 걸린다.

예를 들어, '삼성전자'를 입력하기 위해서는 숫자 자판을 기준으로 볼 때, '8-1-2-0-0'('삼'), '8-2-1-0'('성'), '9-2-1-5'('전'), '9-1-2'('자')를 순서대로 입력해야 한다.

평균적으로 볼 때, 글자 하나 당 약 4번씩 눌러야 하고, 전체적으로는 총 16번이나 눌러야 한다.

2.2 스타일러스 펜과 터치패드

K사의 무선 인터넷 N이 보급되면서, PDA도 함께 활성화되기 시작하였다. PDA의 문자 입력 인터페이스는 스타일러스 펜을 이용한 소프트웨어 키보드 및 필기체 인식이 주를 이루고 있으며, 최근 휴대폰에서와 같은 키 패드가 함께 탑재되고 있다.

소프트웨어 키보드는 터치패드 디스플레이인 PDA 화면 하단에 키보드 모양을 표시하고, 스타일러스 펜을 이용하여 원하는 자판을 클릭하면 이를 인식함으로써 문자가 입력되는 방식이다.

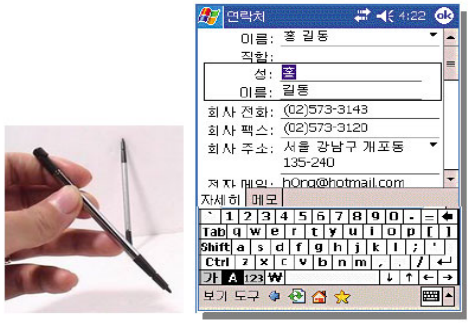

그림 2. PDA 소프트웨어 키보드

그림 2는 스타일러스 펜과 PDA 소프트웨어 키보드를 사용하여 이름을 입력하는 모습을 보여준다. 하단의 키보드에서 원하는 자판을 펜으로 클릭하면 한글 자동 장치에 의해서 조합된 후 상단에 표시된다.

그런데 PDA 디스플레이 크기의 제약 때문에 하나의 키 모양을 표시할 수 있는 영역이 매우 협소하여 원하는 키를 정확히 클릭하기가 쉽지 않다. 특히 운전 중과 같이 이동 중에는 정확성이 현저히 떨어진다.

최근 대중화 단계에 이르고 있는 내비게이션에서도 터치패드 및 스타일러스 펜을 이용한 소프트웨어 키보드를 사용하는데, 크기만 PDA 보다 조금 클 뿐, 위에서 언급한 문제들을 해결하지는 못하고 있다.

소프트웨어 키보드를 이용하여 '삼성전자'를 입력하기 위해서는 'ㅅ-ㅏ-ㅁ'('삼'), 'ㅅ-ㅓ-ㅇ'('성'), 'ㅈ-ㅓ-ㄴ'('전'), 'ㅈ-ㅏ'('자')를 순서대로 입력해야 한다. 평균적으로 한 글자당 약 3번씩의 클릭이 필요하며, 전체적으로는 자판을 총 11번 눌러야 한다

키 패드와 비교해 볼 때 평균 1회의 사용자 작업(누르기/클릭)을 절약할 수 있고, 전체적으로는 5회의 누르기 작업이 줄어든다.

앞에서 언급한 바 처럼 PDA 액정 화면의 크기 제한에 따른 어려움, 즉 각 자판(모양)을 정확히 클릭하기가 쉽지 않다는 문제점은 잔존한다.

한편 그림 3은 필기체 인식 화면을 보여준다. 하단의 3개의 박스가 있고, 그 안에 한 글자를 입력하면(그리면) 이를 인식하여 상단에 표시한다.

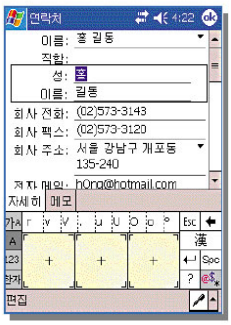

그림 3. PDA 소프트웨어 키보드

'삼성전자'를 입력하기 위해서는 7(삼)-5(성)-5(전)-4(자)의 총 21획을 입력해야 하며, 평균 5획이 필요하다.

또한 각 글자가 입력되면 프로그램에서 이를 인식하는 시간 간격만큼 기다려야 하기 때문에 이를 다시 표시해 보면 7-delay-5-delay-5-delay-4가 된다.

필기체 인식은 키 패드나 소프트웨어 키보드에서와 같이 숫자적으로 비교하기는 무리가 있다. 하나의 주의 소재[3] 내에서의 입력 효용성과 주의 소재의 전환에 따른 효용성에 대한 비교 데이터가 없기 때문이다.

키 패드나 소프트웨어 키보드에서는 하나의 자판을 누르고(클릭하고) 다음 자판으로 이동하는 과정에서 주의 소재의 전환이 이루어지며, 필기체 인식의 경우에는 하나의 주의 소재 내에서 입력이 이루어진다고 할 수 있다.

또한 필기체 인식은 음성인식과 마찬가지로 인식 성능의 한계를 지니고 있으며, 오류 수정 과정은 물론, 인식률을 높이기 위해 좌측의 분류자판(한글/영문/숫자/한자)을 클릭해야 하는 불편함이 있다. 특히 인식 시간의 지연으로 인해 하나의 글자에 대한 인식이 끝나기 전에 다음 글자를 입력하는 실수를 유발하게 된다.

2.3 음성 인터페이스(VUI: Voice User Interface)

음성인식 기술은 1984년도에 음성을 텍스트로 전환하는 기술의 발표를 시작으로 음성 다이얼, 음성인식 IVR(자동응답장치), 음성인식 콜 센터, 음성인식 휴대폰 등 다양한 애플리케이션이 개발되었고 있다.

최근에는 유비쿼터스 서비스의 핵심 기술로서 홈네트워크, 텔레매틱스, 지능형 로봇 등의 분야에서 많은 연구개발이 이루어지고 있다.

그림 1은 일반적인 음성인식 과정을 보여주는 데, 사용자로부터 발화된 음성을 마이크를 통하여 검출하고, 음성학적 특징을 추출한 후, 이를 언어모델, 인식모델, 발음사전을 참조함으로써 비교, 분석 및 결과를 생성, 출력한다.

본 고에서 제안한 멀티모달 인터페이스는 기존의 음성인식 처리 과정 중에서 주로 음성 검출과 발음 사전의 생성 및 참조에 관여하고 있다.

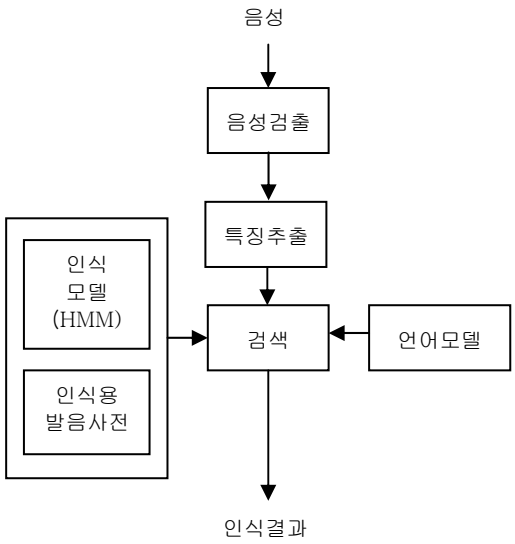

그림 4. 인식 처리 구성도

한편 음성인식 기술은 주변 잡음이 적은 실험실에서는 90% 이상의 좋은 성능을 보이는 반면, 잡음이 심한 복도, 전시장, 회의장, 차량 등에서는 인식률이 현저히 떨어지는 한계를 나타내고 있어서 실용화 및 대중화의 발목을 잡는 원인이 되고 있다.

특히 단어로 구성되어 있는 영어에 비해 한국어는 조사 등의 복잡성과 신호 에너지가 낮아 구분이 어려운 파찰음 등 풀어야 할 숙제가 추가적으로 산재해 있다.

2.1, 2.2에서 언급한 예제에서처럼 16번의 누르기가 필요한 키 패드와 11번의 클릭이 필요한 소프트웨어 키보드에 비해 '삼성전자'라고 말만 하면 문자의 입력이 완료되는 음성인식이 훨씬 효율적이고 편리한 인터페이스임을 알 수 있다.

그러나 음성인식이 실패함으로써 발생하는 수정 노력 및 감성적 불안감을 고려한다면 음성인식이 한 번에 성공한다는 가정 하에서만 위와 같은 비교가 의미가 있다고 하겠다.

3. 멀티모달 인터페이스

3.1 멀티모달 인터페이스

멀티모달 인터페이스란 인간과 기계(컴퓨터, 단말기 등)의 통신을 위하여 음성, 키보드, 펜, 센서 등 다양한 모드를 함께 사용하는 것을 말한다.[1]

그림 5는 내비게이션 단말 상에서 스타일러스 펜과 음성을 사용하여 특정 좌표 주변에 있는 식당들을 찾아주는 멀티모달 응용 사례를 보여준다.

먼저 스타일러스 펜을 사용하여 지도 상의 원하는 지점(범위)에 원을 그리고, "이 근처 원 안에 있는 식당을 찾아주세요!"라고 음성으로 명령을 내리면 시스템은 이 두 가지 정보를 조합하여 "xx 지역에 yy 식당이 있습니다"라는 메시지와 함께 해당 식당정보를 출력한다.

이 응용사례에서 사용된 입력 모드는 GPS, 그래픽, 음성이 된다.

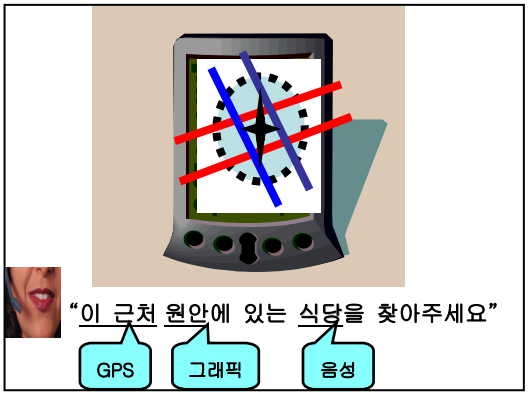

그림 5 멀티모달 응용 예

휴대 단말기에 한정해서 살펴보면, 명령 또는 데이터 입력을 위해 키보드 타이핑, 펜 클릭 및 드로잉(필기체 인식 포함), 음성 인식 등을 사용하고 결과 출력을 위해서 텍스트, 그래픽, 비디오, 음성 및 오디오를 사용한다.

입력 모드의 편리성 측면을 살펴보면, 메뉴 항목을 선택하거나 텍스트 입력을 위해서는 음성-펜-키보드의 순으로 편리한 모드를 제공하며, 기호 입력을 위해서는 펜-키보드-음성, 그림을 그리기 위해서는 펜-음성-키보드 순으로 편리함을 제공한다.

음성(인식)은 손과 눈이 필요하지 않으며 이동 중에도 사용이 가능하고, 다단계 메뉴 상의 명령을 한번에 내릴 수 있다는 장점이 있으나, 소음 환경에 매우 취약하다는 한계를 지니고 있다. 결국 음성인식은 적정 수준의 인식 성공률을 제공한다는 전제를 필요로 한다.

W3C(World Wide Web Consortium)에서는 멀티모달 인터랙션 워킹그룹을 만들어 표준화 활동을 지원하고 있으며, 멀티모달 인터페이스를 이용한 웹 기반 서비스에 필요한 표준안을 개발하고 있다.

그림 6은 W3C에서 제안하고 있는 멀티모달 인터렉션 프레임워크를 보여주고 있는데, 멀티모달 서비스 개발자는 이를 이용하여 서비스, 단말, 상황에 따라 적절한 모드를 선택 또는 복합적으로 사용함으로써 사용자 편의성을 극대화할 수 있다.

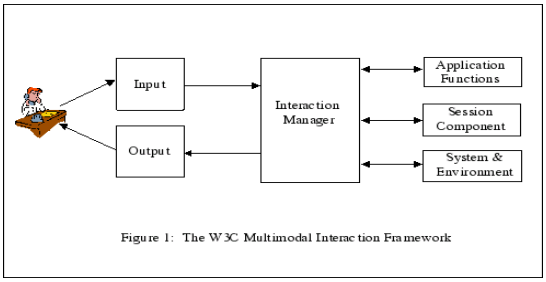

그림 6 멀티모달 인터랙션 프레임워크

여러 모드의 입력 및 출력을 표준화된 형태로 표현하기 위하여 EMMA (Extensible Multimodal Annotation Markup Language) 형식을 사용하고 있으며, 이를 매개 언어로 하여 HTML/SVG, VoiceXML/SALT, InkML 등 각 모드를 표현하는 기존 표준을 통합하는 작업을 진행 중이다.

그림 7는 멀티모달 인터렉션 프레임워크 중에서 입력에 해당하는 멀티모달 컴포넌트 구성도이다.

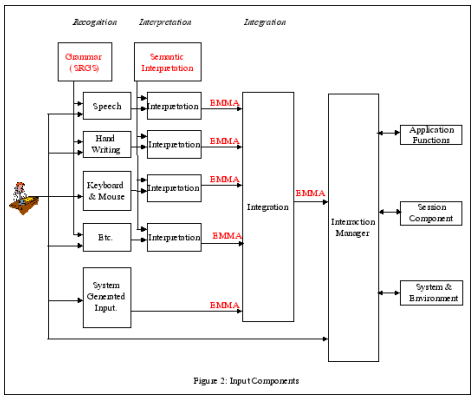

그림 7 멀티모달 입력 컴포넌트

크게 보면 입력장치로부터의 데이터를 인식 및 번역하는 인식(Recognition) 부와 그 데이터가 가지는 의미(Semantic)를 이해하기 위한 해석(Interpretation) 부, 그리고 이들을 통합하는 Integration 부가 있고, 이들 간의 인터페이스는 EMMA를 사용한다.

3.2 음성/키 패드를 이용한 멀티모달 I/F

본 고에서는 음성과 키 패드를 조합한 멀티모달 인터페이스를 사용하여 단어 단위의 텍스트를 입력하는 방법을 제시한다.

사용 방법은 '삼성전자'를 입력한다고 할 때, 먼저 키 패드를 사용하여 첫 번째 자음인 'ㅅ'을 입력하고, 음성으로 '삼성전자'를 발화한다.

첫 번째 자음의 입력 시점을 기준으로 사용자 음성의 시작점을 찾기 때문에 음성인식의 주요 문제점인 시작점 검출 오류를 효과적으로 줄일 수 있고, 해당 자음으로 시작하는 단어들로 인식의 범위를 미리 제한해 둠으로써 음성 인식률을 향상시킬 수 있다.

휴대폰의 경우에는 자판의 제약 때문에 하나의 자판에 몇 개의 자음이 묶여 있는데, 제시된 방법을 사용하여 하나의 자음을 입력하려면 할 수 없이 여러 번의 키 입력을 수행해야 한다.

예를 들어, 천지인 입력 시스템의 경우 'ㅆ'을 입력하기 위해 숫자 8번 키를 세 번 눌러야 한다.

이러한 불편함을 줄이기 위하여 하나의 키 패드에 묶여진 자음들을 하나의 그룹으로 해서 사용할 수도 있다. 즉, 천지인의 경우, 'ㅅ', 'ㅎ', 'ㅆ' 각각 대해 숫자 키 '8'을 한번만 누르도록 하는 것이다.

물론 첫 번째 자음을 정확히 입력하게 되면 인식률은 높아지고 사용의 편의성은 떨어지며, 하나의 키 패드에 묶여진 자음을 그루핑하면 인식률은 낮아지는 반면 편의성은 높아지게 된다.

본 고에서는 두 번째 방법을 채택하여 설명하기로 한다.

이러한 방법은 음성 인식률을 대폭 향상시킬 수 있을 것으로 기대하며, 4장에서 시뮬레이션을 통한 음성인식 성능 향상 정도에 대한 실험 결과를 보여준다.

4. 음성 인식률 시뮬레이션

4.1 시뮬레이션 개요

이 장에서는 음성/키 패드 조합의 멀티모달 인터페이스를 적용한 시뮬레이션 결과를 소개한다. 단, 실제적으로 음성인식률을 대폭 향상시킬 것으로 예상되는 시작점 검출은 포함하지 않고, 자음 군으로 음성인식 대상 범위를 축소한 결과만을 포함한다.

또한 신호특성이 미약한 파찰음으로 시작하는 자음을 포함하는 단어의 인식에 효과적임을 내포한다.

시뮬레이션 절차는 일반적인 음성인식 방법으로 인식 테스트용 음성(녹음된) 세트에 대한 인식성능을 평가한다. 이 음성세트는 실제 철도예약 서비스를 개발하기 위해 녹음했던 음성을 담고 있다.

비교를 위해 먼저 '인식용 발음사전'과 '언어모델'은 전체 인식대상 단어(전체 열차역명)에 대해 생성하고, 그 다음에 특정한 자음(예로 들면, 'ㄱ', 'ㅈ' 등)으로 시작하는 단어에 대해서 각각 '인식용 발음사전'과 '언어모델'을 생성한다.

동일한 음성세트에 대하여 음성 인식기를 실행함으로써 음성인식 성능을 각각 측정한 후, 두 가지 방법에 의한 음성인식률을 비교 분석한다.

4.2 시뮬레이션 환경

시뮬레이션을 위한 음성인식기는 다음과 같이 설정되었다.

- Viterbi 검색 알고리즘 인식기
- 39차 MFCC 음성특징 , mixture=7 HMM 파라메터 사용
- 파라메터는 97,389개 훈련용 음성DB로부터 생성

시뮬레이션은 Xeon Dual CPU(3.80GHz), 2GB RAM, Windows XP Professional이 탑재된 하드웨어 상에서 이루어졌고, 시뮬레이션에 사용될 어휘는 철도청의 기차역명 전체를 포함하였다.

4.3 시뮬레이션 방법

3.2절에서 설명한 두 가지 방법 중, P사의 휴대폰 단말기를 기준으로 각 숫자판에 할당된 자음군 별로 분류하여 적용하였다. 즉, 1,260개의 시험 세트를 인식 어휘 첫 음절의 초성별로 1번-[ㄱ,ㅋ,ㄲ], 5번-[ㄴ,ㄹ], 4번-[ㄷ, ㅌ,ㄸ], 7번-[ㅁ,ㅅ,ㅆ], 8번-[ㅂ,ㅃ,ㅍ], 0번-[ㅇ,ㅎ], '*'번-[ㅈ,ㅉ,ㅊ]으로 분류하였다.

표 1은 그룹별 어휘의 일부를 보여준다.

표 1 그룹별 어휘(일부)

[ㄱ,ㅋ,ㄲ]	[ㄴ,ㄹ]	[ㄷ, ㅌ,ㄸ]	[ㅁ,ㅅ,ㅆ]	[ㅂ,ㅃ,ㅍ]
가수원	나원	다산	마사	반곡
가수원역	나원역	다산역	마사역	반곡역
가야	나전	다솔사	마산	반성
가야역	나전역	다솔사역	마산역	반성역
가은	나주	다시	마석	반야월
가은역	나주역	다시역	마석역	반야월역
가좌	나한정	단성	마성	백마

실험을 위해 먼저 모든 인식 어휘에 대한 인식 입력 데이타(인식 클래스 데이타, 인식용 발음사전, 인식 목록)를 생성하고, 전체 테스트용 음성 세트에 대하여 음성인식기를 실행함으로써 음성인식률을 시뮬레이션 한다.

두번째로 초성별 인식 어휘에 대한 인식 입력 데이타를 생성한 후, 동일한 테스트 음성 세트에 대해서 음성 인식 시뮬레이션을 수행하였다.

4.3 시뮬레이션 결과

일반 음성인식 방법의 시뮬레이션 결과는 표 2와 같다.

표 2 음성인식 시뮬레이션 결과

그룹	인식율	실패	시험세트
[ㄱ,ㅋ,ㄲ]	87.81	190	1,447
[ㄴ,ㄹ]	93.52	16	247
[ㄷ, ㅌ,ㄸ]	86.32	327	2,391
[ㅁ,ㅅ,ㅆ]	88.37	419	3,603
[ㅂ,ㅃ,ㅍ]	79.81	259	1,283
[ㅇ,ㅎ]	87.11	263	2,041
[ㅈ,ㅉ,ㅊ]	80.7	322	1,668
평균인식율	85.84%	1,796	12,680

12,680개의 음성이 녹음된 테스트 세트에 대해 음성인식기를 실행시킨 결과, 1,796개가 인식에 실패하였고, 평균 인식률을 85.84%를 기록하였다.(본 테스트 음성 세트는 서비스 개발을 위하여 정상/비정상적인 음성을 포함하였기 때문에 실제 음성인식기의 성능을 대표하는 것은 아님을 밝혀둔다.)

음성/키 패드 조합의 멀티모달 인터페이스 시뮬레이션 결과는 표 3과 같다.

표 3 멀티모달 시뮬레이션 결과

그룹	인식율	실패	시험세트	성능향상
[ㄱ,ㅋ,ㄲ]	91.57%	122	1,447	3.76%
[ㄴ,ㄹ]	96.76%	8	247	3.24%
[ㄷ, ㅌ,ㄸ]	88.5%	275	2,391	2.18%
[ㅁ,ㅅ,ㅆ]	91.31%	313	3,603	2.94%
[ㅂ,ㅃ,ㅍ]	86.59%	172	1,283	6.78%
[ㅇ,ㅎ]	91.67%	170	2,041	4.56%
[ㅈ,ㅉ,ㅊ]	87.05%	216	1,668	6.35%
평균인식율	89.94%	1,276	12,680	4.10%

동일한 12,680개의 테스트 음성에 대하여, 1,276개가 인식에 실패하였고, 평균 인식률을 89.94%를 기록하였다

4.4 결과 분석

일반적인 음성인식 방법에 의한 모의 실험 결과는 85.84%의 인식 성공률을 나타냈고, 음성/키 패드 조합의 멀티모달 인터페이스 시뮬레이션 결과는 89.94%의 인식 성공률을 보임으로써 평균적으로는 4.1%의 인식률 향상이 이루어졌다.

표 3의 그룹별 성능 향상을 보면 각 그룹별로 6.78% ~ 2.18%로서 비교적 고른 개선이 이루어 졌음을 알 수 있다.(그림 8 참조)

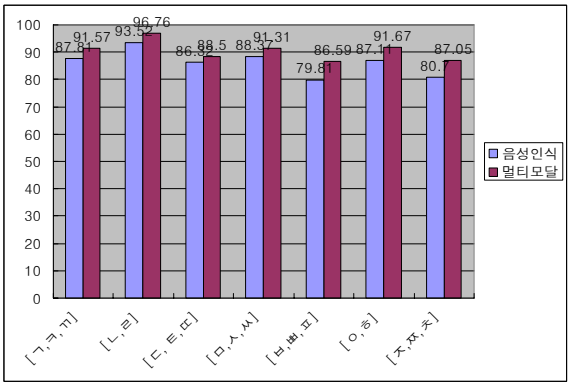

그림 8 시뮬레이션 결과 비교

이번 실험은 초성 자음만 알려주고 인식을 수행한 결과이므로, 실제 환경에서는 좀 더 높은 인식률 개선을 기대해 볼 수 있다. 첫 자음(군)의 입력 시점을 기준으로 사용자 음성의 시작점을 찾기 때문에 주변 소음 등의 영향으로 사용자가 언제 역명을 발화할 것인가를 찾는 어려운 문제를 어느 정도 해결할 수 있기 때문이다.

한편 기차역명 인식은 기차역명 간의 유사도가 비교적 높기 때문에, 유사도가 낮은 인식 애플리케이션에서는 좀 더 나은 성능을 나타낼 것으로 예상된다.

4. 결론

디자인의 슬림화, 기능의 복잡화 및 다양화 추세에 따라 음성인식 기술에 대한 관심이 높아져 가고 있으며, 그 적용 분야 또한 날로 다양해 지고 있다. 이는 음성 인식의 성능 문제만 해결된다면 우리 생활의 모든 분야에서 활용될 수 있다는 믿음 때문일 것이다.

이와 같은 배경에 비추어 볼 때 여기서 제안한 멀티모달 인터페이스를 통하여 음성인식의 한계를 극복하고 새로운 인터페이스의 가능성을 제시하고자 한다.

본 고에서는 음성/키 패드 조합의 멀티모달 인터페이스를 제시하고, 기차역명을 대상으로 음성인식률 향상을 모의 실험하여 인식률의 향상 정도를 확인하였다.

특히 음성인식의 대표적인 기술적 문제로 인식되고 있는 시작점 검출과, 특히 한글 인식을 더욱 어렵게 만드는 파찰음 인식의 문제를 어느 정도 해결할 수 있으며, 음성인식 대상을 대용량화하는 데에도 기여할 것으로 보인다.

향후에는 시작점 검출을 포함하여 실제 음성인식률의 향상을 가늠해 볼 수 있도록 시뮬레이션 해 보고, 다른 휴대폰 단말에서의 자음 조합 등에 대한 시뮬레이션도 수행해 볼 필요가 있다.

또한 보다 다양한 모드의 멀티모달 인터페이스에 대한 연구를 통하여 유비쿼터스 시대의 새로운 인터페이스 개발을 위한 다양한 시도가 필요하다.

참고문헌

[1] 구명완, “음성 인터페이스와 멀티모달 인터페이스 “, ITFIND 주간기술동향 통권 1193 호(2005.4.27)

[2] Ho-Hyun, Jeon, et al., “A Speech Operated Railroad Information & Reservation Service With Multistatage Dialogue”, SST-2000

[3] Jef Raskin(이건표 옮김), “인간 중심 인터페이스(Human Interface)”, 안 그라픽스, 2003

[4] Xuedong Huang, et al., “Spoken Language Processing - A Guide to Theory, Algorithm, and System Development”

[5] Multimodal Architecture and Interfaces, W3C Working Draft 11 December 2006, http://www.w3.org/TR/2006/WD-mmi-arch-20061211/