Journal of the Ergonomics Society of Korea
Vol. 18, No. 3 1999. 12

음성인식용 인터페이스의 사용편의성 평가 방법론

A Usability Evaluation Method for Speech Recognition Interfaces

한 성 호*, 김 범 수

Abstract

As speech is the human being' s most natural communication medium, using it gives many advantages. Currently, most user interfaces of a computer are using a mouse/keyboard type but the interface using speech recognition is expected to replace them or at least be used as a tool for supporting it. Despite the advantages, the speech recognition interface is not that popular because of technical difficulties such as recognition accuracy and slow response time to name a few. Nevertheless, it is important to optimize the human-computer system performance by improving the usability.

This paper presents a set of guidelines for designing speech recognition interfaces and provides a method for evaluating the usability. A total of 113 guidelines are suggested to improve the usability of speech-recognition interfaces. The evaluation method consists of four major procedures: user interface evaluation; function evaluation; vocabulary estimation; and recognition speed/accuracy evaluation. Each procedure is described along with proper techniques for efficient evaluation.

Keyword: Speech recognition interface, Design guidelines, Usability evaluation method

* 경북 포항시 남구 효자동 산31번지 포항공과대학교 산업공학과 (우) 790-784

1. 서 론

컴퓨터의 일반적인 입력 수단으로는 키보드와 마우스가 주류를 이루고 있으나, 이는 다음과 같은 사용상의 취약점을 가지고 있다(Dillon et al., 1990; Khalid, 1992).

- 타이핑 및 키 입력 기술이 필요하다.
- 화면과 입력 기구 사이의 빈번한 시선의 이동이 필요하다. 이로 인해 작업의 효율성이 감소한다.
- 마우스와 키보드 사이의 손의 이동이 육체적인 피로를 유발한다.

이와 같은 키보드/마우스 사용자 인터페이스의 취약점을 해결하기 위한 대안으로는 여러 가지가 있을 수 있으나 그 중 음성을 이용한 인터페이스가 도입되어, 일부는 이미 상용화 되어 있다. 음성을 인터페이스로 응용한 분야는 의료, 우주항공, 품질 검사, 재고관리, 자동차, 항공 선적작업, 장애자용 휠체어, 속기용 타자기 등 점차 그 분야가 확장되고 있으며 음성 인식을 이용한 제품을 보다 편리하고 성능이 향상된 형태로 개발하기 위한 많은 연구가 진행 중이다. 이와 같은 최근의 연구 추세는 다음에 제시한 음성 사용 인터페이스의 장점을 컴퓨터 사용자 인터페이스에 활용하려는 것으로 이해할 수 있다(Noyes, 1993; Mcleod, 1994; Damper, 1993).

- 수많은 정보를 다루는 컴퓨터 인터페이스에서 음성은 손과 눈의 작업 부하를 줄여준다.
- 음성은 가장 자연스러운 정보교환 수단으로서, 음성의 친밀감과 편리함은 컴퓨터 사용시의 편의성을 도와준다.
- 한정된 공간에서 사용되는 기존의 인터페이스에 비해 공간의 제약이 적다.
- 마우스/키보드 작업과 함께 동시 작업이 가능하므로 작업의 효율성을 증진시킨다.

음성은 사용자에게 보다 편한 인터페이스를 제공하고, 컴퓨터 사용에 익숙하지 않은 사용자에게도 사용상의 편리함을 제공한다. 이러한 장점에도 불구하고, 음성 인식시간과 음성 인식율에 대한 기술적 한계로 인하여, 음성은 마우스/키보드와 같이 일반적인 인간-컴퓨터의 의사전달 수단으로 사용되지 못하고 있다. 그러나, 완벽하지 못한 인식율과 상대적으로 긴 인식시간을 요구하는 현재의 기술만으로도 음성의 적용대상이 될 제품은 많이 존재하며, 이 경우 사용편의성의 확보가 시장에서의 경쟁력을 좌우하는 중요한 요소가 된다. 사용편의성 향상을 위해 무엇보다도 중요한 점은 개발자에 의해 설계되고 평가되는 인터페이스보다는 사용자의 요구사항과 평가를 거친 설계가 이루어져야 하고, 또한 일회성의 평가가 아닌 개발과정에서 반복적으로 평가 보완되어야 한다는 점이다.

본 논문에서는 사용편의성을 체계적으로 수행할 수 있는 방법론을 제시하고자 한다. 구체적인 대상으로는 화자독립(Speaker Independent), 명령어 타입(Command type) 음성인식기를 이용하는 음성인식 시스템을 설정하였다. 이와 같은 기본 목적을 달성하기 위한 세부적인 목표는 다음과 같이 요약된다.

첫째, 음성인식 인터페이스의 평가에 필요한 설계지침(Guidelines)을 제공한다. 본 논문에서는 음성인식에 대한 관련 연구들을 참

조하여, 음성인식 인터페이스를 설계할 때 고려해야 할 사항들을 추출하고, 시스템 설계자와 평가자가 쉽게 사용할 수 있도록 분류하여 제시하고자 한다.

둘째, 음성인식 인터페이스에 대한 사용편의성 평가방법론을 다음과 같은 네 가지의 요소로 정의하였으며, 이들 각각에 대한 평가 수행방법을 제시하고자 한다.

- 사용자 인터페이스에 대한 평가
- 음성인식 시스템의 성능에 대한 평가
- 음성 구현 대상에 대한 평가
- 사용 어휘에 대한 평가

2. 음성인식 인터페이스의 설계지침

사용편의성을 고려한 음성인식 인터페이스의 설계와 효율적인 평가를 위해서는 체계적으로 분류된 설계지침이 필요하다. 이를 위하여 본 연구에서는 기존 국/내외 문헌에서 음성인식에 관련된 연구논문들을 조사하여 설계지침을 추출하였다. 추출된 설계지침은 적용대상, 적용상황, 음성사용의 적합성등을 기준으로 분류되었다. 특히, 현재 음성인식기의 성능이 완벽하지 못하기 때문에 성능을 개선하기 위해 필요한 사용편의성을 이에 부가하였다. 이와 같이 분류된 설계지침은 그림 1에 나타난 바와 같이 크게 사용자 인터페이스, 인식기 성능, 음성명령어 설계, 음성사용, 기타 등 5개의 그룹으로 제시되었으며, 각 그룹에 포함된 상세 설계지침의 수는 총 113개로 나타났다.

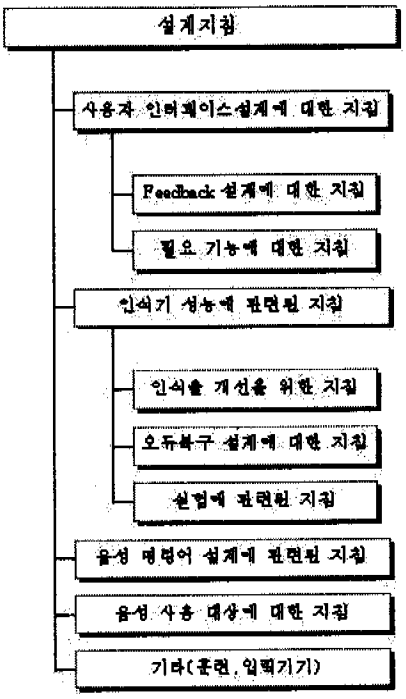


그림 1. 설계지침 분류

사용자 인터페이스 평가에 필요한 설계지침은 인터페이스의 일반적인 구성 요소들(메뉴, 대화상자, 창, Feedback, 아이콘 등) 중 음성인식 인터페이스에서 강조되어지는 Feedback 측면에 관한 지침과 음성인식의 구현을 위해 반드시 필요한 기능에 관한 지침들로 구성되어 있다 (표1참조). 표 1에 나타난 바와 같이 음성인식의 경우 시스템이 음성을 인식하였는지에 대한 Feedback이 매우 중요하며, 이는 여타 인터페이스 (예, 마우스/키보드) 에서 제공되어지는 시각적인 Feedback이 없기 때문이다. 또한, 음성인식 인터페이스의 효율을 최대화하기 위한 지침을 제시함으로써, 사용환경 또는 상황에 따라 음성인식 기능을 선별적으로 사용하는데 도움이 되도록 하였다.

표1. 사용자 인터페이스 평가를 위한 지침

구 분	소 구 분	설 계 지 침
Feedback 설계	적용대상	· 어휘 선택에 도움을 줄 수 있는 Feedback을 제공한다(한성호 et al, 1997a). · 탈락수준(Rejection Level)과 같은 인식기(Automatic Speech Recognizer)의 설정치(Parameter)를 조절하는데 도움을 주는 Feedback을 제공한다(Moody et al, 1988). · 행동(인식 중, 인식가능 상태)에 대한 Feedback을 제공 한다(Jones et al, 1989). · 인식기의 명령어 인식결과를 Feedback으로 제공 한다(Jones et al, 1989). · 오류를 찾을 수 있는 Feedback과 오류를 복구할 수 있는 Feedback을 제공한다(Jones et al, 1989). · 명령 인식에 대한 불확실성이 있는 시스템에서는 모든 명령에 대한 확인창(Yes/No, OK)을 나타내는 Feedback을 제공한다(Frankish et al,1990). · 단기 기억의 부담을 줄이기 위해, 인식된 단어의 축적(History)을 보여 주는 Feedback을 제공한다(Schurick et al, 1985). · 안전을 요하는 작업에서는Feedback이 반드시 필요하다(Hapeshi et al, 1990). · 오류가 매우 적으면서 시간이 중요한 상황에서는 Feedback을 제공하지 않는다(Craft, 1982).
	Auditory Feedback	· 화면(Visual Display)을 관측할 필요가 없는 상황에서 사용한다(Jones et al, 1989). · 음성 Feedback은 기억이 필요 없는 상황에서 유용하다(Jones et al, 1989). · 음성 Feedback은 음성의 기억부하(Verbal Memory Load)가 필요한 상황에서는 사용하지 않는다(Jones et al, 1989). · 오류가 결정적인 상황일 경우, 음성이 주위를 집중시킬 수 있는 성질 때문에 의해 음성 Feedback이 적절하다(Noyes et al, 1994). · 음성 Feedback은 일시적으로 제공되므로 Terminal Feedback에 의한 오류 복구에는 적당하지 않다(Noyes et al, 1994). · 매우 높은 인식 정확성을 요구하는 응용분야에서는 음성 Feedback을 사용한다(Frankish et al, 1990). · 시각적인Feedback이 주작업에 방해가 되는 상황에서 사용한다(Schurick et al, 1985). · 시간이 중요한 작업에서는 음성Feedback을 사용하지 않는다(Schurick et al, 1985). · 사용자가 어떻게 발음해야 하는지에 대한 지표(Reference)를 제공할 필요가 있는 작업에서 음성 Feedback을 사용 한다(Hapeshi et al, 1990). ? 청각적인 Feedback은 소음 환경에는 적합하지 않다(Hapeshi et al, 1990). · 눈에 과부하가 있는 경우, 음성Feedback을 사용한다(Schurick et al, 1985). · 음성 Feedback은 디지털화 된 음성보다는 합성된 음성을 이용한다. · Feedback에 적합한 음성합성의 질은 귀에 거슬리는 정도로서 Head size가 110%, Pitch가 90Hz가 적당하다(Rosson, 1986). · 음성 Feedback은 단순하고, 짧은 메시지 및 즉각적인 행동을 요구할 때 사용한다(Cowley et al, 1990). · 청각적인 부하가 없고 주위환경이 너무 어둡거나 밝을 때 사용한다(Cowley et al, 1990). · 작업 시 사용자의 움직임이 많은 상황에서 사용한다(Cowley et al, 1990).

표1. 사용자 인터페이스 평가를 위한 지침

구 분	소 구 분	설 계 지 침
Feedback 설계	Auditory Feedback	· 작은 어휘, 간단한 문법구조를 가진 음성 인식 시스템을 위한 Feedback으로는 간단한 청각적인 신호(Auditory Tone)로도 충분하다(Cresswell et al, 1993). · Feedback 인식의 실패(Feedback Monitoring Failure)에 의한 사용자 오류를 감소하기 위해서는 시각적인(Visual) Feedback보다는 음성 Feedback이 적당하다(Frankish et al, 1993).
	Visual Feedback	· 긴 문장의 Feedback이 필요한 경우에는 사용자의 기억 부하를 감소시키기 위해 시각적인(Visual) Feedback을 사용한다(Schurick et al, 1985). · 열악한 환경(소음, 진동, High G)하에서는 사용자가 읽기 쉽고 단기 기억에 의존하지 않도록 시각적인(Visual) Feedback을 적정 시간동안 제공한다(Baber, 1996).
	Concunent Feedback	· 많은 기억 부하(High Memory Load)가 있는 상황에서 사용한다(Jones et al, 1989). · 한 작업이 여러 부-작업(Sub-Task)으로 구성된 작업의 경우, 오류복구나 Feedback이 주작업(Primary Task)에 방해가 되므로 동시적인(Concurrent) Feedback과 즉각적인 오류복구는 사용하지 않는다(Frankish et al, 1990). · 연속 (Continuous) 음성인식 시스템에는 사용하지 않는다(Noyes et al, 1994).
	Terminal Feedback	· 주 작업(Primary Task)이 빠른 수행을 요구할 경우에는 Terminal Feedback을 사용한다(Schurick et al, 1985). · 인식율이 높은 인식시스템에서는 Terminal Feedback이 작업 수행 시간상 이득이 있다(Noyes et al, 1994).
	메시지의 내 용	· Feedback 메시지에서는 적절한 행동을 취할 수 있는 대안이 포함되어야 한다(Schumacher, 1995). · 시스템과의 대화(Dialogue)는 시스템이 인식한 것을 보여 주고, 정확성의 여부를 질의한다(Leggatt, 1992).
	기 타	· 자료와 명령(Command)을 위해서는 서로 다른 타입의 Feedback이 필요하다(Frankish et al, 1990). · 현재 명령 가능한 단어나 어구를 나열한다. 이 방법은 가능 어휘가 적을 때 효과적이다(한성호 et al, 1997a; 1997b). · 대단히 높은 인지적 부하(Cognitive Load)가 있는 상황에서는 복잡한 오류복구 대화상자보다는 현재 작업을 중단하고 다시 입력하는 것이 더 좋다(Frankish et al, 1990). · 작업 환경이 많은 양의 시각적인 Feedback을 수용할 수 있고, 음성 입력에 대한 즉각적인 Feedback이 필요한 작업에서는 과거행적(History)을 가진 시각적인 동시(Concurrent) Feedback이 유용하다(Schurick et al, 1985). · 학습할 때와 실제 작업 사이의 음성패턴을 일관적으로 유지하기 위해 인식을 추세를 보여 준다(Hapeshi et al, 1990). · 음성의 파형을 보여줌으로써, 훈련(Training)과 실제 음성 사용의 일관성을 유지할 수 있다(Hapeshi et al, 1990). · Feedback은 텍스트 보다는 부호(Symbol)를 사용하는 것이 좋다(Baber et al, 1992). · 상태표시Feedback은 사용자가 하는 일에 영향을 주지 않아야 하며, 쉽게 신속히 알아볼 수 있어야 한다(한성호 et al, 1997a; 1997b).

표1. 사용자 인터페이스 평가를 위한 지침

구 분	소 구 분	설 계 지 침
필요기능		· 상황을 회복할 수 없거나 인식 결과가 결정적인 시스템에서는 인식 도중에서도 명령을 취소할 수 있는 명령이 있어야 한다(Frankish et al, 1990). · 한 작업을 수행하는데 많은 명령어나 자료(Data String)가 필요하다면, 작업자가 전체 명령을 다시 확인할 수 있는 명령어가 필요하다(Schurick et al, 1985). · 탈락역치(Rejection threshold)를 조절할 수 있는 인터페이스가 필요하다(Hapeshi et al, 1990). · 되돌림(Undo) 기능이 있어야 한다(한성호 et al, 1996).

음성인식 시스템의 성능은 음성입력에 대한 인식율과 인식시간에 의해 좌우된다. 본 연구에서는 인식율을 증가 시킬 수 있는 방법 및 사용자 오류가 발생했을 때 이를 복구할 수 있는 방법을 위주로 설계지침을 제시하였다(표2참조). 표 2에 제시된 지침은 다음과 같은 두 가지 측면에서 매우 중요한 지침이 된다. 첫째, 현재의 음성인식기 성능이 완벽하지 못한 점을 들 수 있다. 둘째, 사람과 사람과의 대화에서도 전달되는 정보의 오인식(Misunderstanding)은 존재하며, 실제 인식기가 입력된 음성을 정확하게 인식하기를 기대하는 것은 현실적이지 못하기 때문이다. 따라서, 인식된 명령어의 오류복구에 대한 설계지침은 필수적으로 요구된다.

표 2. 인식기 성능 평가를 위한 지침

구 분	설 계 지 침
인 식 율 개 선	· 시스템에 없는 명령어휘의 사용, 잘못 발음된 명령에 의한 사용자 오류를 감소시키기 위해 사용자와 친밀한 어휘의 사용, 한 명령을 위해 여러 어휘를 대응, 또는 명령어 리스트를 화면에 제시하는 것이 필요하다(한성호 et al, 1997a). · 스트레스에 의한 발음속도의 증가로 발생하는 오류를 감소시키기 위해 연속(Continuous) 음성 인식기를 사용한다(Baber, 1996). · 잘못된 Feedback에 대한 반응과, 오류복구의 실패, 기억의 상실에 의한 사용자 오류를 감소시키기 위해 주 작업(Primary Task)에 방해되지 않고, 사용자가 집중할 수 있는 Feedback을 제시한다(Frankish et al, 1990). · 사용 어휘를 최소화하여 인식율을 증가 시킨다. 예를 들어 어휘(Vocabulary)의 한계를 극복하는 방법으로 소수의 단어를 조합하여 다수의 어휘를 만들어 사용하는 것을 제안한다(Moody et al, 1988). · 어휘의 개선(비슷한 단어, 한 문자로 구성된 명령, 단축어 등은 사용하지 않는다)을 통해 인식율을 극대화한다. · 두 번째 제품을 출시할 때, 사용자가 인식율이 개선되었다는 것을 인식하게 하기 위해서는 처음 상품보다 5% ~10%정도의 인식율 개선이 필요하다(Buskirt et al, 1995).

표 2. 인식기 성능 평가를 위한 지침

구 분	설 계 지 침
인식율 개선	· 인식율을 증가 시킬 수 있는 방법으로 입술 읽기(Visual Lip Reading), 제스처(Gesture), 눈동자 방향(Eye direction) 을 함께 사용한다. 그리고 이는 소음이 높은 환경에서 유용하다(Nishida, 1986). · 오류는 Feedback, 오류복구, 인식단어 확인대화(Word Confirmation Dialogue)를 통해 감소시킬 수 있다(Schurick et al, 1985). · 피로 및 스트레스, 소음에 의한 오류를 감소시키기 위해 각 상황에서의 표준파형(Template)을 훈련(Training)하여 표준파형 군(Template Set)에 저장한다. · 개인의 음성 변화에 따른 오류를 최소화 하기 위해서는 화자 적응 음성인식을 사용한다(Hapeshi et al, 1990). · 오류를 줄이기 위해 천천히 발음하며, 각 단어 사이에 간격(Pause)을 준다(KVS Manual, 1985; Hapeshi et al, 1990). · 인식 오류를 감소시킬 수 있는 최선의 방법으로 여러 개의 인식기를 동시에 사용한다(Barry, 1993). · 차량전화(Car-phone) 같은 소음이 있는 환경에서 사용되는 인식기는 삽입오류(Insertion Error)를 줄이는 것이 중요하다(Ross, 1990). · 소음이 있는 환경에서는 소음 제거기(Noise Filter)나 작동/중지 스위치(On/Off Switch) 가 있는 마이크(Microphone)를 사용하거나, 또는 소음제거기능(Noise Canceling Algorithm)을 사용한다(Baber, 1996). · 음성입력의 변화에 의한 오류를 최소화 하기 위해 마이크(Microphone)의 위치를 고정하거나 머리부착 마이크(Head-set Microphone)를 사용한다(한성호 et al, 1997a; 1997b). · 입 전체를 뒤집어 쓰는 마이크(Microphone)는 되도록 이면 사용하지 않는다(숨소리가 방해가 된다.) (한성호 et al, 1997a; 1997b).
오류 복구	· 자동 오류복구 사용 시에는 선정된 단어 중 다음 최선의 단어로 대체하는 방법을 사용한다(Noyes et al, 1994). · 오류가 적은 상황에서는 나중에 오류를 수정하는 방법(Delayed Error Correction)을 사용한다(Hapeshi et al, 1990; Schurick et al, 1985). · 시스템 설계 시, 오류복구의 2가지 모드(Delayed/Immediate Error Correction)를 선택할 수 있는 옵션을 사용자에게 제공한다. · 사용자 오류 복구로는 오류가 발생하면 잘못 인식된 표준파 형(Template)을 제거하고 다음 명령 인식 후에 즉시 제거된 표준파형(Template)을 원상태로 복구하는 방법을 사용한다(Murray et al, 1993). · 문서편집기(Word Process)등 많은 양의 어휘(Large Vocabulary)를 요구하는 인식기에서는 작업이 끝난 다음에 오류를 복구하는 지연복구(Delayed Error Correction)를 이용하며, 조사, 자료입력 작업에서는 한 단어 다음에 즉시 복구(Immediate Error Correctio)n를 이용한다(Hapeshi et al, 1990).
실 험	· 음성인식 시스템의 성능을 평가 하기 위해서는 다른 종류의 사용자, 다른 훈련(Training)환경, 다른 타입의 어휘, 다른 종류의 작업으로 실험이 이루어져야 한다(Brown, 1988). · 음성과 키보드 비교 평가는 실제 상황과 비슷하고 가장 사용하기 쉬운 조건하에서 평가가 이루어 져야 한다(예: 키보드/마우스 사용의 경우에도 단축키나 단축어의 사용을 대상으로 이루어져야 한다.) (한성호 et al, 1997b).

음성인식 시스템에서 음성명령어로 사용된 어휘에 대한 평가를 위해 명령어 설계에 필요한 지침을 제시하였다. 음성으로 구현되는 대상에 대한 평가와 관련하여, 음성입력 인터페이스에서 음성 적용 대상(기능, 작업) 선정 시에 고려해야 사항들을 음성 입력이 적합한 작업과 적합하지 않은 작업으로 구분하여 제시하였다(표3, 표4 참조). 또한, 사용자가 어떤 작업을 수행하고자 할 때 음성명령어로 선택할 수 있는 어휘의 수가 복수 개 존재하므로 어떤 어휘를 인식 가능하게 할 것인가에 대한 설계지침이 중요하다. 예를 들면, 윈도우 화면 내용을 다음 쪽(page)으로 이동할 때 (즉, Scroll Down), 사용될 수 있는 명령어로는 "아래로", "밑으로", "다음 쪽", "다음 페이지", "아래로 이동", "밑으로 이동" 등 많은 수가 존재한다. 이들 중 어떤 것을 실행 가능한 명령어로 선택하느냐는 매우 중요한 사용편의성 연구대상이 되며, 선택된 어휘의 종류 및 수에 따라 개발기간, 인식속도, 메모리 비용 등 여타 설계사항에도 지대한 영향을 미치게 된다.

표 3. 음성 명령어 설계를 위한 지침

설 계 지 침
· 부호(Code)의 입력보다는 의미 있는 이름을 사용한다(Moody et al, 1988).
· 발음하기 쉽고, 작업(Task)을 적절히 반영해야 한다(Moody et al, 1988).
· 긴 어휘는 학습(Learning)하기 쉽고, 일관되게 발음이 가능하며, 더 많은 음성 정보를 제공하므로 짧은 어휘보다 인식되기 쉽다(Moody et al, 1988).
· 자연스럽고(Natural) 친밀한 어휘를 사용한다(Moody et al, 1988).
· 사용자가 자신의 어휘를 선정하도록 하는 것이 좋다(Moody et al, 1988).
· 단어 선택의 유연성을 높이기 위해 'Multiple Mapping'을 이용한다(한 기능을 여러 어휘로 가능하게 한다.) (Moody et al, 1988; Lai et al, 1997).
· 비슷한 발음이나 음소(Phoneme)는 사용하지 않는다(Johns et al, 1989).
· 어휘(Vocabulary)의 한계를 극복하는 방법으로 소수의 단어를 조합하여 다수의 어휘를 만들어 사용하는 것을 제안한다(Hill, 1980; Noyes et al, 1994).
· 한 단어로 표시하기 보다는 다음절어, 짧은 어구를 사용한다(Noyes et al, 1994).
· 어휘는 사용자의 습관과 기대에 따라 설계되어야 한다(Johns et al, 1989).
· 어휘에는 단축어나 두문어를 절대 사용하지 않는다(Johns et al, 1989).
· 명령 문법 설계에도 주의가 필요하다. 예를 들어 '입력, 일, 삼, 오' 보다는 '일, 삼, 오, 입력'이 더 자연스럽다(Johns et al, 1989).
· 상업적인 용도로 가장 적합한 어휘의 크기는 100~200개 정도이다(Casali et al, 1990).
· 많은 양의 어휘(Large Vocabulary)는 적은 양의 어휘(Small Vocabulary)보다 유사한 단어가 많기 때문에 오류가 많다. 그러므로 적당한 Feedback이 요구된다(한성호, 1997b).
· 명령어휘에서 은유적인 표현은 사용하지 않는다(Leggatt, 1992).

표 4. 음성사용 대상 평가를 위한 지침

구 분	설 계 지 침
음성이 적합한 작업	· 손과 눈(Visual Overload)이 다른 일을 하고 있는 상태에서는 음성 입력을 사용한다(예: Material Handling(소포, 메일, 짐의 분류/정리작업)) (Damper, 1993; Damper et al, 1995; Noyes et al, 1993). · 품질 관리 작업(Quality Control Task)에서 음성 사용이 적당하다(예: 전자 제품, 자동차의 검사 작업) (Noyes et al, 1994). · 음성은 명령작업에 적합하고, 숫자나 그래픽자료(Graphical Data)는 수동작이 유리하다 (Khalid, 1992). · 2가지 작업이 동시에 이루어 질 때(Concurrent Task), 작업 중 하나에 음성을 사용한다 (Noyes et al, 1994; Mounfield and north, 1980; Oviatt, 1997). · 불연속적인 작업(Discrete Task)에 음성입력을 사용한다(예: String(단어쓰기), Selection(메뉴 선택) 작업) (Moody et al, 1988; Damper, 1993). · 간격(Pause)이 있는 불연속 음성(Discrete Speech)은 연속 음성(Continuous Speech)보다 20~40 단어/분(Word/Min)의 속도 감소를 유발하므로 시간이 중요한 작업에는 연속 음성 인식기(Continuous ASR)를 이용한다(Moody et al, 1988). · 정해진 표준에 의해 작업이 이루어 질 때 사용한다(Moody et al, 1988). · 적은 어휘를 사용하는 간단한 명령은 음성 사용이 적합하다(Buskirk, 1995). · 복잡한 키 조합 작업과 자료의 그룹을 한번에 입력하는 작업에 음성이 사용된다 (Cresswell, 1993).
음성이 적합하지 않은 작업	· 음성은 오래 사용하면 피로(Vocal Fatigue)를 유발하므로 입력이 빈번한 상황에서는 사용하지 않는다(예: 자료 입력보다는 명령입력에 적합하다.) (Johns et al, 1989). · 정밀한 조절이 필요한 작업에서는 음성이 불편하다(Buskirk, 1995). · 사용상의 안전과 관련된 작업에는 음성을 사용 사용하지 않는다(Mcleod, 1994). · 복잡하고 정신적인 스트레스가 있는 작업에는 음성을 사용하지 않는다(Mcleod, 1994). · 물체를 다루는 공간적인 작업(Spatial Task)은 음성 작업이 부적당하다(Moody et al, 1988). · 연속작업(Continuous Task)에는 음성을 사용하지 않는다. (예: Quantify(Pointer Position), Position(Cursor key Positioning), Path(Curve 그리기), Tracking Task의 경우) (Moody et al, 1988; Foley et al, 1984)

끝으로 인식기의 학습과 음성 입력 수단인 마이크로폰에 관한 설계지침을 제시하였다(표 5 참조). 본 논문에서 제시한 설계지침은 음성인식 시스템을 설계할 때 기초적인 설계 요구사항 (Design Requirements)으로 사용할 수 있음은 물론, 개발 과정 중 제작된 시제품과 완성된 제품의 사용편의성 평가기준으로도 활용이 가능하다.

표 5. 기타(훈련, 입력기기)에 대한 지침

구 분	설 계 지 침
Training	· 화자독립(Speaker Independent)시스템에서 표준파형 군(Template Set)의 구성은 남/녀, 노/소를 대상으로 실제 작업환경의 발음을 표준파형 군(Template Set)으로 구성해야 한다(Moody et al, 1988). · 모든 조건 및 환경(Stress, Fatigue, Noise, Vibration)의 표준파 형(Template)을 저장한다(Hapeshi et al, 1990). · 의미있는 표준파 형 훈련(Template Training)을 제공한다. 즉, 실제작업에서 사용되는 어휘를 제공하거나, 실제 작업과 비슷한 환경에서 훈련이 이루어져야 한다(Johns et al, 1989). · 표준파형 훈련(Template Training)에 필요한 한 단어 당 반복 발음 횟수는 2번~10번이 적당하다(Hapeshi et al, 1990).
	· 화자 적응(Speaker-adaptive) 훈련(Training(=Hidden Training))을 사용할 경우, 사용자 내제적인 비일관성(Stress, Fatigue등)을 제거할 수 있어 음성의 스타일을 유지하는데 유용하다(Hapeshi et al, 1990). · 훈련(Training)할 때는 사용자에게 관심 있는 문장으로 훈련(Training)하여 지루함과 짜증에 의한 음성의 변화를 최소화한다(Christ, 1986). · 훈련(Training)은 화자의 발음 상태가 좋을 때 실시해야 한다. 예를 들어 잠에서 깨고, 충분한 시간 후에 훈련(Training)을 실시한다(Hapeshi et al, 1990). · 각 환경에 따른 표준파형 군(Template Set)을 만들어 둔다. 예를 들어 소음 환경에서는 소음환경 표준파형 군(Template Set)을 이용하여 작업하도록 한다(Hapeshi et al, 1990).
입력 수단	· 음성 특징 중 사용자의 이동성(Mobility)을 최대화하기 위해 유선 마이크(Microphone)보다는 무선 마이크(Microphone)을 사용한다(Usher, 1993). · 사용자 행동이 필요한 입력기보다 계속 입력 방식이 바람직하다. 즉, 사용자가 어떤 행동을 취해야 음성을 인식하는 입력기(예: 버튼을 누르고 있는 상태에서만 입력이 가능한 마이크)보다는 계속 음성을 받아들이는 입력기를 사용한다 (Schumacher et al, 1995).

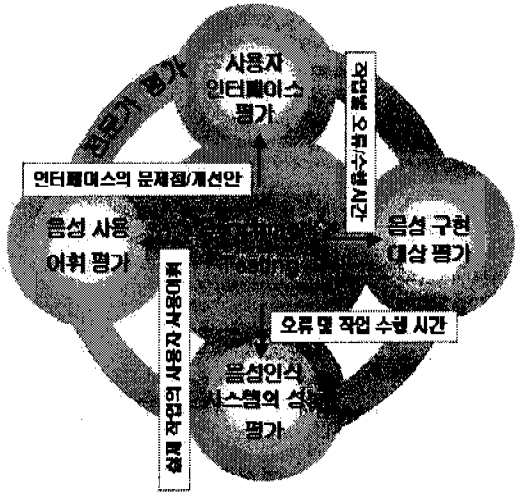
3. 사용편의성 평가 방법론

사용편의성의 평가는 설계의 문제점을 파악하고 개선사항을 도출하는 과정으로 정의되며, 본 연구에서는 음성인식 인터페이스의 사용편의성 평가를 효율적으로 진행하기 위해 필요한 방법론을 제시하고 있다.

3.1. 음성인식 인터페이스의 사용편의성 평가방법론

음성인식 인터페이스는 기본적으로 음성을 인식할 수 있는 장치의 성능 (예, 인식율 등), 음성명령어를 적용할 수 있는 대상 작업 (예, 윈도우 열기, 파일 삭제 등), 대상작업을 수행하기 위해 필요한 명령어휘의 선별,

그리고 사용자 인터페이스의 전반적인 사항 등으로 구성되며, 이들에 대한 체계적인 평가가 필요하다. 전술한 바와 같이 현재 음성인식기의 성능이 완벽하지 못할 뿐만 아니라 시스템에서 수행되는 모든 기능이 음성으로 작동되어 져야 하는지에 대한 문제는 사용편의성은 물론 개발 비용/시간 등에 관련된 문제이다. 본 연구에서는 그림 2에 나타난 바와 같이 음성인식 인터페이스의 사용편의성을 평가하기 위한 대상을 네 가지로 구분하여 위에서 언급한 문제점을 종합적, 체계적으로 고려할 수 있게 하였다 (한성호 et al., 1997a).



(그림 2. 음성인식 인터페이스의 평가방법론)

사용자 인터페이스에 대한 평가에서는 메뉴(Menu), 색(Color), Feedback, 대화상자(Dialogue Box) 등과 같은 일반적인 인터페이스 구성요소에 대한 평가를 수행한다(한성호 et al., 1996). 이는 음성을 사용한 인터페이스에서도 마우스/키보드를 이용한 시각적인 정보 교환(Visual Interaction)이 이루어지므로 위와 같은 요소에 대한 평가가 필요하기 때문이다.

음성인식 인터페이스에 대한 평가에서는 일반적인 사용자 인터페이스에 대한 평가와 함께 음성 사용 어휘에 대한 평가, 음성 구현 대상에 대한 평가, 인식 시스템의 성능에 대한 평가가 추가로 이루어진다. 음성 사용 어휘에 대한 평가와 음성 구현 대상에 대한 평가는 음성을 이용한 명령어 수행 및 화자독립 음성 인식기의 사용편의성 개선을 위해 필수적으로 요구되는 평가이다. 더불어, 음성인식 시스템의 성능에 대한 평가는 평가 대상이 되는 제품을 기존의 마우스/키보드 입력을 사용하는 제품과 비교함으로써 시장 확보의 가능성을 파악하기 위해 필수적으로 요구되는 평가이다.

전술된 네 가지 평가들은 그림2에 나타난 바와 같이 전문가에 의한 평가와 사용자에 의한 평가(Benchmark Testing, 설문조사)를 통해 이루어진다. 전문가 평가에서는 위에서 제시된 평가 요소별 설계지침들을 바탕으로 음성인식 시스템의 각 요소들이 사용편의성을 만족하고 있는지를 평가하는 내용으로 구성된다. 대표작업 평가(Benchmark Testing)에서는 사용자를 참여 시킨 실험을 통하여 음성인식 시스템의 사용편의성을 평가한다(한성호 et al., 1996).

특히, 대표작업 평가(Benchmark Testing)에서는 실험을 통하여 네 가지 평가들의 기초 자료들이 제공된다. 따라서, 음성사용이 적합한 작업을 파악하고, 사용 어휘를 파악하기 위해서는 대표적인 작업의 선정이 중요하다. 대표작업 평가(Benchmark Testing)에

사용되는 작업은 음성인식 시스템에서 사용되는 음성명령어를 대부분 사용할 수 있도록 선택하며, 마우스/키보드 사용이 편리한 작업, 음성이 편리한 작업을 미리 분류하여 제시하여야 한다.

각 평가에 대한 자세한 설명이 다음에 기술되어 있다.

3.2 사용자 인터페이스에 대한 평가

사용자 인터페이스에 대한 평가는 인터페이스의 사용편의성을 최대화하여, 작업의 효율성과 제품에 대한 선호도를 극대화 할 수 있도록 인터페이스 설계상의 문제점을 파악하고, 적절한 개선안을 제시하는 것을 목적으로 한다(한성호 et al., 1996).

사용자 인터페이스에 대한 평가는 사용편의성 전문가에 의한 평가, 대표작업 평가(Benchmark Testing), 자유작업 평가(Hands-on Experiment) 등으로 구성된다(한성호 et al., 1996).

전문가에 의한 평가에서는 음성인식 시스템의 인터페이스 구성 요소(Menu, Window, Icon, Color, Message, Dialogue Box, Help, Terminology, Feedback, Error Recovery)들이 사용편의성 설계지침(Usability Principle)에 따라 적절히 설계되었는지의 여부를 사용편의성 전문가가 평가한다. 특히, Feedback 및 오류(Error)에 관련된 평가는 앞에서 제시한 설계지침을 바탕으로 평가가 이루어 지는 것이 바람직 하다. 이외에도, 음성 인식 인터페이스에서 필요한 기능의 유무를 파악하고, 나아가 음성기능의 시각화 방법의 적정성 등은 음성 인터페이스에서 부가적으로 필요한 중요 평가 대상이다.

전문가에 의한 평가 결과를 보완하고 사용자가 음성인식 시스템을 사용하면서 겪는 문제점 및 불편함을 파악하기 위하여 사용자를 이용한 실험 평가가 필수적으로 요구된다. 이는 사용자의 행동양식이나, 시스템을 사용하는 형태(Human-Computer Interaction)를 전문가가 미리 예측하기 어렵기 때문이다. 이 평가 단계에서는 대표적인 작업을 선정하여 사용자에게 수행해 보도록 함으로써 시스템 사용상의 문제점을 파악하는 대표작업 평가(Benchmark Testing)와 실제 사용자에게 시스템을 자유롭게 사용해 보도록 하여 사용자가 문제점을 스스로 지적하는 자유작업 평가(Hands-on Experiment)가 사용된다.

대표작업 평가에서는 사용자들이 자주 수행하거나 중요성이 높은 대표적인 작업에 대하여 음성 작업과 마우스 작업을 사용자가 수행하도록 함으로써, 마우스와 음성의 수행도를 비교하는 것을 주요 목표로 한다. 또한, 작업 수행과정에서의 불편한 점이나 생각들을 생각말하기 (Thinking Aloud) 형식으로 모두 진술하도록 하여, 인터페이스의 문제점을 파악하는 방법이 유용하게 사용될 수 있다. 이와 함께, 사용자가 시스템을 사용하는 전 과정을 녹화한 비디오 분석을 통하여 사용자의 사용 어휘, 작업 수행 시간, 작업 실수 등을 측정함으로써, 시스템 성능에 대한 평가, 사용 어휘에 대한 평가의 기본 자료로 활용하는 것이 추천된다.

자유작업 평가에서는 시스템 전반에 대하여 사용상의 불편한 점 및 만족스럽지 못한 점을

사용자로부터 직접 파악할 수 있다.

각 평가가 종료된 후에는, 분야별 평가 내용에 대하여 음성 인식 인터페이스의 문제점들을 종합함으로써, 개선안들을 도출한다(한성호 et al., 1996). 제시된 문제점들은 중요도, 파급효과, 수정에 요구되는 노력, 기술적인 난이도 등에 있어서 모두 같은 비중을 가지고 있지 않기 때문에 이러한 측면을 고려하여, 개발팀과 평가팀과의 의견 수렴을 통한 전략적인 수정대상 선정 과정이 필요하다(한성호 et al., 1997b).

3.3. 음성인식 시스템의 성능 평가

음성인식 시스템에 대한 성능 평가에서는 사용자 인터페이스의 개선에 의해 성능의 한계를 극복할 수 있는 방법을 제시하고, 마우스/키보드 작업과의 수행도 차이를 비교한 평가 결과를 통해 기존 인터페이스의 음성 대체 가능성을 검토하는 것을 주요 목표로 한다. 본 평가는 대표작업 평가결과를 바탕으로 작업수행시간에 대한 분석, 작업 수행의 정확성에 대한 분석, 마우스/키보드 작업과의 비교분석으로 구성된다.

3.3.1 작업 수행 시간에 대한 분석

일반적으로 음성을 사용하는 작업은 마우스/키보드 작업보다 많은 시간이 소요된다. 이 같은 현상은 다음과 같은 모델에 의해서 설명이 가능하다(식 1 참조).

음성 작업 수행 시간 =
적합한 어휘선택 시간
+발성하는 시간
+음성인식 시간
+ Secondary Feedback제시 시간
+오류복구 시간 (식 1)

적합한 어휘선택 시간은 마우스/키보드 작업에서의 기능 선택(Search and Selection)시간에 대응되는 시간으로, 작업을 수행하기에 적합한 기능(또는 명령어)를 사용자가 추출해 내는 시간과 모든 기능이 음성으로 구현되어 있지 않은 상황에서 음성 구현 기능을 찾는 시간의 합으로 구성된다. 이때 소요되는 시간을 감소시키기 위해서는 적절한 어휘의 사용(사용자가 가장 자주 사용하고, 친밀한 어휘, Label을 그대로 발음하도록 하는 방법)과 음성 사용 기능의 적절한 시각화 방법이 요구된다(한성호 et al., 1997b).

음성을 발성하는데 소요되는 시간은 사용자가 선택한 어휘를 발음하는 시간으로 마우스의 버튼 클릭에 해당하는 작업 시간이다. 이는 음성 사용의 가장 근본적인 문제로 마우스 사용보다 시간을 감소시키기가 사실상 불가능하다. 그러나, 짧은 어휘의 사용과 빠른 발음 속도를 인식 할 수 있는 연속발음(Continuous speech) 음성 인식기의 사용으로 소요 시간을 최소화 할 수 있다.

음성 인식 시간은 사용자가 발성한 단어를 인식기가 가지고 있는 표준파형(Template)과 비교하여 가장 적합한 단어를 찾는 시간으로, 음성인식 알고리즘, 주기억장치(CPU) 속도, 메모리(RAM)의 크기, 어휘의 크기에 영향을 받는다. 이 부분의 소요시간이 마우스 사용과 비슷하게 되기까지는 음성인식 속도 및 정확도에 대한 앞으로의 기술발전에 달려 있다.

Secondary Feedback 제시 시간은 시스템이 인식한 단어를 사용자에게 제시하는 시간으로, 음성합성을 통해 제시할 경우, 사용자가 인식하기까지는 약간의 시간이 소요된다. 이 시간은 100%의 인식율을 가진 시스템에는 필요 없는 시간이며, 음성 Feedback이 반드시 필요한 상황(기억이 필요 없는 상황, 높은 인식의 정확성을 요구하는 상황, 눈에 과부하가 있는 상황, 너무 어둡거나 밝은 환경) 이외에는 시각적 Feedback을 사용함으로써 단축시킬 수 있다.

오류 복구 시간은 오류에 의해 부가적으로 요구되는 시간으로서, 인터페이스와 인식율의 개선에 의해 최소화 될 수 있다. 인터페이스의 개선은 3.2절에 그리고, 인식율의 개선은 3.3.2절에 각 각 자세한 설명이 기술되어 있다.

이상과 같이, 음성 인식 시간과 오인식에 의한 시간은 시스템 성능 개선에 의해 최소화하고, 이외의 시간은 어휘 및 인터페이스의 개선에 따라 소요시간을 단축시켜야 한다. 이를 위해서 작업수행도의 평가는 비디오 분석에 의한 작업 별 시간분석(Time Line Analysis)을 통하여 각각의 시간을 측정함으로써, 어떤 요소가 작업 수행 시간에 가장 큰 영향을 주는지를 파악하고, 앞으로 어느 정도까지 작업 시간을 단축할 수 있는 지를 평가함으로써, 시스템의 개선방향 및 개선전략을 수립할 수 있도록 하여야 한다. 예를 들어, 실험 분석 결과, 사용자가 음성명령 어휘를 선택하는데 소요되는 시간이 전체 작업 시간의 가장 많은 부분을 차지하고 있으면 평가자는 어휘의 개선에 의한 작업 수행 시간의 감소 방안을 가장 먼저 강구하여야 한다.

3.3.2. 작업수행의 정확성 분석

작업수행의 정확성 분석은 작업수행 시 발생하는 오류의 종류 및 빈도에 대한 분석이다. 음성 인식 인터페이스 사용상의 가장 큰 문제점의 하나로 불완전한 인식율에 의한 시스템 오류가 지적되고 있다(한성호 et al., 1997b). 낮은 인식율은 작업 수행 시간과 사용자의 만족도를 감소시키는 주 원인으로 파악되고 있기 때문에 중요한 평가 대상이 된다.

작업실수는 오인식에 따른 시스템 오류와 부적절한 어휘 사용에 의한 사용자 오류로 구분할 수 있다(한성호 et al., 1997a).

시스템 오류는 마우스/키보드 사용 시스템에서는 거의 발생하지 않는 오류로서, 작업환경과 훈련(Training)환경 사이의 비일관성이 가장 주요한 원인이고, 그 이외에 사용자의 음성피로(Vocal Fatigue), 발음속도, 스트레스, 주위소음, 어휘의 수, 단어의 길이, 마이크의 성능 및 위치 등이 원인으로 파악된다(Johns et al., 1989; Hapeshi et al., 1990). 특히, 어휘의 수는 인식율에 큰 영향을 주는 것으로 파악되므로, 적절한 수준을 유지해야 한다(Hill, 1980; Noyes et al., 1989).

사용자 오류의 원인으로는 시스템에 없는 명령어휘의 사용, 잘못된 어휘의 사용, 잘못된 Feedback에 대한 반응과, 오류복구의 실패, 주작업(Primary Task)에 대한 기억의 상실, 잘못 발음된 명령이 주요 원인이며, 각각 어휘의 개선(사용자의 친밀한 어휘의 사

용, Multiple-Mapping 사용, 명령어 List를 화면에 제시), 적절한 음성 사용 기능 시각화를 통해 해결할 수 있다(한성호 et al., 1997b).

3.3.3. 마우스/키보드 작업과의 비교

마우스와 키보드를 음성인식으로 대체하거나 음성인식이 보조수단으로 사용될 수 있는지에 대한 검증을 하기 위해서는 음성 작업과 마우스/키보드 작업의 수행시간, 정확성 및 선호도에 대한 비교 평가가 필요하다. 평가는 대표작업 평가를 이용하며, 작업 내용은 음성 사용이 유리한 경우와 마우스/키보드 사용이 유리한 경우의 작업이 모두 포함되어 있어야 한다. 또한, 음성과 키보드 비교 평가는 실제 상황과 비슷하고 가장 사용하기 쉬운 조건하에서 평가가 이루어 져야 한다(예: 키보드/마우스 사용의 경우에도 단축키나 단축어의 사용을 대상으로 이루어져야 한다).

사용자의 주관적인 평가에서는 학습의 난이도, 요구되는 집중도, 좌절감, 만족도 등에 대한 종합적인 평가가 이루어 져야 한다. 마지막으로 동시 작업이 가능한 작업을 선정하여, 손과 눈에 부하가 있을 경우, 음성 사용의 가능성을 파악하는 것이 필요하다.

3.4. 음성 구현 대상에 대한 평가

현재의 기술수준이나 사용자의 작업요구사항을 고려하여 볼 때, 장애자용 컴퓨터와 같은 특수 응용분야를 제외하고는, 인터페이스의 모든 기능을 음성으로 구현하기에는 한계가 있다. 즉, 음성 사용이 마우스/키보드 사용보다 편리하고 효율적인 경우에만 음성인식 기술을 사용하여야 한다.

어떤 기능을 음성으로 가능하게 할 것인가에 대한 평가는 전문가에 의한 기능 선정, 사용자를 대상으로 한 설문조사의 두 가지 방법을 이용하는 것이 제안된다.

평가에 앞서, 음성 적용 소프트웨어에서 사용되는 기능과 작업의 조사가 필요하다. 조사된 기능은 특징별(사용 빈도, 중요성, 기능의 위치 등)로 분류하고, 작업 분석을 통하여 어떠한 작업 특징을 가지고 있는 지를 파악하여야 한다.

전문가에 의한 기능 선정 단계에서는 문헌조사를 통하여 기능 선정에 대한 설계 지침과 선정 기준을 정립하고, 이와 함께 작업 분석을 통하여 음성 사용이 편리한 기능 및 작업을 선정한다. 필요에 따라 경쟁사 제품분석(Competitive Analysis)을 통하여 유사 제품에서 음성으로 구현된 기능을 추출하여 분석에 사용하여야 하며, 이 때 기능 선정의 기준은 친밀성 (Familiarity)이 되어야 한다. 즉, 기존에 발표되어 많은 사용자들이 사용하고 있는 기능은 우선적인 선택 대상이 되어 사용자에게 친밀감을 확보할 수 있도록 하여야 한다. 사용자를 대상으로 하는 설문조사에서는 자주 사용하는 기능 및 선호하는 기능에 대한 조사가 이루어 져야 한다.

마지막으로 선정된 기능에 대하여, 음성 사용 적합 여부를 판정하기 위한 대표작업 평가를 실시한다. 마우스/키보드 사용과 음성 사용을 각각 적용하여, 작업 수행도를 측정함으로써, 선정된 기능의 음성 구현여부가 평가되어 진다.

한편, 일부 기능만이 음성으로 작동될 경우 사용자는 새로운 사용자 모델(User Model)을 수립해야 하는 부담과 함께 음성인식이 가능한 기능을 익혀야 하는 기억의 부담을 가지게 된다. 이와 같은 부담들은 음성인식이 가능한 기능들에 대한 적절한 시각화 방법을 통해서 감소시킬 수 있다(한성호 et al., 1997b). 즉, 음성인식이 가능한 기능들을 음성인식이 불가능한 기능들과 차별화 함으로써 쉽게 인식할 수 있도록 하여 사용자 모델의 수립을 돕고, 기억의 부담을 덜어 줄 수 있게 된다.

이 단계에서는 기능의 시각화 방법들을 조사하고, 적용하고자 하는 분야에 적합한 방법을 선정하는 과정이 필요하다. 선정 방법으로서는 일관성 유지 가능성, 화면에서 추가 공간의 필요성, 다른 항목과의 구별 용이성, 정보코딩(Information Coding) 의미 파악의 용이성 등의 기준에 따라 사용자 및 전문가의 의견을 모두 고려하는 방법이 제안된다.

3.5. 어휘에 대한 평가

음성인식시스템의 성능평가 중 음성명령을 찾는 시간, 시스템이 제공하는 어휘의 부족, 사용자의 부적절한 어휘 사용 등이 가장 큰 문제점으로 지적된다(한성호 et al., 1997a). 따라서, 작업수행 시간 및 오류를 감소시키기 위한 어휘의 적정성 및 포괄성에 대한 평가가 이루어져야 한다.

한편, 사용자가 원하는 기능을 수행할 때 예측할 수 있는 모든 종류의 어휘가 구현 되어 있는 것이 이상적인 시스템이다. 그러나, 어휘 수의 증가는 음성인식에 소요되는 시간을 증가 시키게 되므로, 적정 어휘의 선택이 요구 된다. 어휘에 대한 평가는 어휘확장의 한계범위와 관련하여 적절한 어휘의 개수와 기능에 적합한 사용 어휘를 제시 하는 것이 제안된다. 평가는 전문가에 의한 어휘 제시와 대표작업 평가, 그리고 사용자의 선호도에 의한 어휘선정 등 세 가지 방법을 이용한다(한성호 et al., 1997b).

대표작업 평가에서는 실제 작업을 통하여, 시스템이 제공하는 명령어 이외에 같은 기능을 수행하는 명령어를 파악할 수 있다. 그러나 이 과정에서는 일부 기능의 어휘만을 조사할 수 있는 단점을 가지고 있다. 전체기능의 어휘는 다음에 제시되는 전문가 평가와 설문조사의 병행을 통해 파악하는 방법을 사용할 수 있다.

전문가 평가에서는 유사 프로그램(Application Program)에서 사용되는 명령어 및 전문가가 제시한 명령어를 모두 고려하는 것이 바람직하다. 이에 부가하여, 사용자로부터 프로그램을 사용할 때 사용 가능한 명령어를 조사함으로써, 명령어휘의 기본 대상을 추출할 수 있다. 이렇게 수집된 어휘는 표 3에 나타난 설계지침을 바탕으로 부적절한 어휘를 제거해 나가는 방법으로 적절한 명령어를 선택할 수 있다. 예를 들어, 지금까지 조사한 어휘 중에서 기술적으로 구현이 불가능한 어휘(외자, 유사발음)를 명령어에서 제거 할 수 있다. 마지막으로 각 기능에 대한 여러 명령어 중에서 사용자의 선호 어휘를 찾기 위해 설문조사를 통하여 선호되지 않는 어휘는 제거함으로써 최종 명령어를 확정하는

방법이 제안된다.

평가 결과는 시스템의 성능을 고려하여 어휘의 수 및 변화정도를 수용하는 분석에 사용함으로써, 시스템의 사용편의성을 향상시키는 방안을 모색할 수 있다.

4. 결론 및 추후 연구

음성인식기의 성능 (인식율, 인식속도, Feedback 등)이 완벽한 경우 기타 사용자와의 상호작용에 대한 사용편의성만 고려한다면 음성인식기가 마우스나 키보드 만큼 널리 사용될 것으로 기대하나, 이는 현실적으로 불가능하며 또한 불완전한 성능이라고 해서 실생활에 응용할 수 없다는 의미는 아니다. 즉, 부족한 성능을 가진 음성인식 기술로도 활용분야가 많이 존재하며 이럴 경우 시장에서의 성공 여부는 대부분 사용편의성에 의존된다고 해도 과언이 아니다.

이를 위하여, 본 연구에서는 불완전한 성능을 가진 화자 독립 명령어 타입 음성인식기를 사용하는 음성인식 인터페이스의 사용편의성을 제고하기 위한 설계지침 및 평가체계를 제시하고 있다. 설계지침은 사용자 인터페이스에 대한 지침, 인식기 성능에 관련된 지침, 음성 명령어 설계에 관련된 지침, 음성 사용대상에 대한 지침, 훈련 및 입력 기기에 대한 지침 등 크게 5개로 분류하여 총 113개의 설계지침을 제시하였다. 제시된 지침은 설계초기 단계에서 뿐만 아니라 개발과정 전체에 모두 적용될 수 있을 것으로 기대된다.

한편, 사용편의성 평가체계는 전문가 의견, 사용자를 활용한 실험, 주관적인 의견 등을 종합적으로 사용할 수 있도록 설계되었으며, 제시된 평가체계를 개발과정에 사용한다면 사용편의성을 체계적으로 제고 시킬 수 있을 것으로 기대된다.

본 연구에서 제시된 지침과 평가방법론은 음성인식의 경우 뿐만 아니라 음성합성의 경우로 확장이 필요하다. 특히, 음성인식과 합성이 동시에 사용되는 경우가 증가할 것으로 예상되는 바, 통합된 사용자 인테페이스의 설계를 위한 설계지침 및 사용편의성 평가방법론의 개발도 필요하다. 이를 보다 확장 시키면 멀티미디어의 사용이 대중화 됨에 따라 어떤 상황 및 작업에 어떤 상호작용 매체(Interaction Modality)가 사용되어야 하는지에 대한 다중 매체 사용자 인터페이스(Multi-modal User Interface)에 대한 연구도 피요하다. 마지막으로, 설계지침이 제안되어 있기는 하지만 많은 경우에 서로 상충되는 지침이 발생할 수 있으며, 또한 어떤 상황에서 어떤 지침이 가장 효율적인지에 대한 연구도 수행되어야 개발과정에서 보다 더 유용하게 사용될 것으로 판단된다.

참 고 문 헌

[1] 한성호, 곽지영, 어홍준, 김병주, 홍상우, 김범수, "미래로 사용자 인터페이스의 사용편의성 평가", 연구보고서, 포항공과대학교, 포항, 1996.

[2] 한성호, 윤명환, 어홍준, 김범수, 이세훈, "음성인식을 이용한 Internet Navigator

의 사용편의성 평가", 연구보고서, 한국 전자 통신 연구원, 대전, 1997a.

[3] 한성호, 윤명환, 어홍준, 김범수, "음성 명령어 및 Human Interface 개선 사항", 연구보고서, 한국 전자 통신 연구원,대전, 1997b.

[4] Baber, C., "Automatic speech recognition in adverse environments.", Human Factors, 38(1), 142-155, 1996.

[5] Barry, T. P., "Enhanced recognition accuracy with the simultaneous use of three automated speech recognition systems", Proceedings of the human factors society 37th annual meeting (pp.288-292), 1993.

[6] Brown, T. J., "Human factors assessment of automatic speech recognition system performance.", Proceedings of the human factors society 32nd annual meeting (pp.288-290), 1988.

[7] Casali, S. P., Williges, B. H., and Dryden, R. D., "Effect of recognition accuracy and vocabulary size of a speech recognition system on task performance and user acceptance", Human Factors, 32 (2), 183-196, 1990.

[8] Christ, K. A., "Comparison of isolated-word and connected-word voice recognition systems for a data entry task", Proceedings of Speech Tech86(pp.262-264), 1986.

[9] Cowley, C. K., Miles, C., and Jones, D. M., "The incorporation of synthetic speech into the human-computer interface", In Lovesey, E. J.(Ed), Contemporary Ergonomics, Taylor and Francis, London, 1990.

[10] Cresswell, A. F., "Is control by voice the right answer for the avionics environment?" In Baber, C., and Noyes, J. M.(Ed), Interactive speech technology: human factors issues in the application of speech input/output to computers, Taylor & Francis, London, 1993.

[11] Damper, R. I., "Speech as an interface medium: how can it best be used?", In Baber , C., and Noyes, J. M.(Ed), Interactive speech technology : human factors issues in the application of speech input/output to computers, Taylor & Francis, London, 1993.

[12] Damper, R. I., and Wood, S. D., "Speech versus keying in command and control applications", Int. J. Human-Computer Studies, 42, 289-305, 1995.

[13] Dillon, R. F., Edey, J. D., and Tombaugh, J. W., "Measuring the

true cost of command selection: Techniques and results", Proceedings of CHI Conference on Human Factors in Computing Systems(pp. 19-25), 1990.

[14] Foley, J. D., and Chan, P., "The human factors of graphics interaction techniques", IEEE Computer graphics and Applications, 4, 13-48, 1984.

[15] Frankish, C., Jones, D., and Hapeshi, K., "Maintaining recognition accuracy during data entry tasks using speech input", In Lovesey, E. J.(Ed), Contemporary Ergonomics, Taylor and Francis, London, 1990.

[16] Frankish, C., and Noyes, J., "Feedback in automatic speech recognition: who is saying what and to whom", In Baber, C., and Noyes, J. M.(Ed), Interactive speech technology: human factors issues in the application of speech input/output to computers, Taylor & Francis, London, 1993.

[17] Hapeshi, K., and Jones, D. M., "The ergonimics of automatic speech recognition interfaces", In Lovesey, E. J.(Ed), Comtemporary Ergonomics, Taylor and Francis, London, 1990.

[18] Hill, D. R., "Spoken language generation and understanding by machine: a problems and applications oriented overview", In Simon, J. C.(Ed), Spoken language generation and understanding, Dordrecht, D. Reidel, 1980.

[19] Jones, D., Hapeshi, K., and Frankish, C., "Design guidelines for speech recognition interfaces", Applied Ergonomics, 20(1), 47-52, 1989.

[20] Khalid, H. M., "Assessment of integrated speech-manual input system: Towards optimising behavior and performance", In Lovesey, E.J.(Ed), Contemporary Ergonomics, Taylor and Francis, London, 1992.

[21] Lai, J., and Vergo, J., "Medspeak: Report creation with continuous speech recognition", Proceedings of CHI Conference on Human Factors in Computing Systems(pp.431-438), 1997.

[22] Leggatt, A., and Baber, C., "Optimizing the recognition of digits in automatic speech recognition through the use of "Minimal dialogue" ", In Lovesey, E. J.(Ed), Contemporary Ergonomics, Taylor and Francis, London, 1992.

[23] Mcleod, R. M., "Human factors

issues in speech recognition", In Lovesey, E.J.(Ed), Contemporary Ergonomics, Taylor and Francis, London,1994.

[24] Moody, S. T., and Joost, M. G., "System Design for Automated Speech Recognition", In Hendler, M.(Ed), Handbook of Human-Computer Interaction, Elsvier Science, New York, 1988.

[25] Mountfield, S.J., and North, R. A., "Voice entry for reducing pilot work-load", Proceedings of the Human Factors Society24th annual meeting(pp.185-189), 1980.

[26] Murray, A. C., Frankish, C. R., and Jones, D. M., "Data-entry by voice: facilitating correction of misrecognition", In Baber, C., and Noyes, J. M. (Ed), Interactive speech technology: human factors issues in the application of speech input/output to computers, Taylor & Francis, London, 1993.

[27] Nishida, S., "Speech recognition enhancement by lip-information", Proceedings of CHI Conference on Human Factors in Computing Systems(pp.198-204), 1986.

[28] Noyes, J. M., "Speech technology in the future", In Baber , C., and Noyes, J.M.(Ed), Interactive speech technology : human factors issues in the application of speech input/output to computers, Taylor & Francis, London, 1993.

[29] Noyes, J. M., and Frankish, C. R., "Errors and error correction in automatic speech recognition systems", Ergonimics, 37(11), 1943-1957, 1994.

[30] Oviatt, S., Deangeli, A., and Kuhn, K., "Integration and synchronization of input modes during multimodal human-computer interaction", Proceedings of CHI Conference on Human Factors in Computing Systems (pp.415-422), 1997.

[31] Ross, T., "Carphone dialing by direct voice input: The effect of feedback mode and recognition errors on performance", In Lovesey, E. J.(Ed), Contemporary Ergonomics, Taylor and Francis, London, 1990.

[32] Rosson, M. B., and Cecala, A. J., "Designing a quality voice: An an alysis of listeners' reactions to synthetic voices", Proceeedings of CHI Conference on Human Factors in Computing Systems (pp.192-197), 1986.

[33] Schumacher, R. M., Hardzinski, M. L., and Schwartz, A. L.,

"Increasing the usability of interactive voice response systems: Research and guidelines for phone-based interfaces", Human Factors, 37(2), 251-264, 1995.

[34] Schurick, J. M., Williges, B. H., and Maynard, J. F., "User feedback requrements with automatic speech recognition", Ergonimics, 28(11), 1543-1555, 1985.

[35] Usher, D. M., "automatic speech recognition and mobile radio", In Baber, C., and Noyes, J. M.(Ed), Interactive speech technology: human factors issues in the application of speech input/output ot computers, Taylor & Francis, London, 1993.