

음성인식을 이용한 주관평가 시스템 구현

한화영 · 고한우* · 윤용현* · 조택동
충남대학교 기계설계공학과
*한국표준과학연구원 인간정보그룹

The Subjective Evaluation System Implementation Using Speech Recognition

Hwa Young Han, Han Woo Ko*, Yong Hyeon Yun*, Taik Dong Cho
Dept. of Mechanical Design, Chungnam University
*Ergonomics Lab, Korea Research Institute of Standards and Science

Abstract

환경이나 작업부하 등이 인간에게 주는 피로나, stress 또는 쾌, 불쾌감 등의 감성을 평가하기 위한 정신물리학적 방법의 하나로 설문지에 의한 주관적인 평가법이 많이 사용되고 있다. 기존의 수작업으로 이루어지던 설문 방식을 자동화하여 PC 기반으로 설문양식을 자동 생성하고 음성을 통해 응답할 수 있는 프로그램을 개발하였다. 주관평가 자동화 시스템은 주관평가 데이터를 효율적으로 처리할 수 있고 음성을 이용함으로써 피험자의 정신적 부담을 경감시키며 생리신호와 주관평가와의 경시적인 변화를 효과적으로 평가할 수 있다.

설문 형식으로는 5점 척도와 7점 척도를 선택하였으며 평가어는 “매우 아니다” ~ “매우 그렇다”로 구성되었다. 평가어를 인식함에 있어 좋은 인식률을 얻기 위한 특징벡터의 차수와 기본 프레임 개수를 대상으로 인식실험을 하였다.

Keywords: 음성인식, 설문지, 주관평가 자동화

1. 서 론

심리적인 변화나 환경적인 변화가 인간에게 미치는 정서적인 영향을 정확히 평가하기 위해서는 가능하면 안정적인 실험환경과 피험자에게 평가 그 자체가 주는 정신적인 부담을 줄이는 것이 필수적이다.

지금까지 주관평가 방법은 피험자가 설문지를 직접 작성하거나 컴퓨터의 키보드를 통해 입력하는 방식이 사용되어 왔다. 따라서 효율적인 감성평가와 피험자의 정신적 부담을 경감시키고 생리신호와 주관평가와의 경시적인 변화를 관찰할 수 있는 보다 효과적인 방법이 연구되어야 한다.

본 논문에서는 사람의 가장 자연스러운 의사소통 방법인 언어에 의한 주관평가를 하기 위하여 음성인식기술을 이용하여 평가어를 인식할 수 있는 주관평가 시스템을 구현하였다.

구현된 시스템의 유용성을 검증하기 위하여 5점척도와 7점척도를 이용하여 평가하는 SD 법에 적용하였다. 설문에 사용되는 음성의 특징벡터로는 16차 LPC cepstrum 계수를 사용하였으며 인식방법으로는 화자 종속 고표어 인식 시스템의 구성에 주로 이용되며 인식률이 높다는 장점을 가지고 있는 알고리즘인 DTW(Dynamic Time Warping) 알고리즘을 이용하였다.

2. 주관 평가 시스템의 자동화

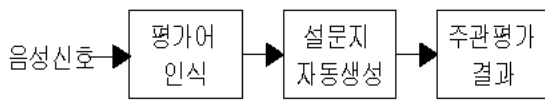


그림 1. 주관평가 자동화 시스템의 구성[1]

전체적인 주관평가 시스템의 구성은 그림 1과 같다. 피험자가 평가하는 음성신호가 마이크를 통해 입력되면 평가어를 인식하고 인식된 평가어는 자동적으로 평가화면에 표시되며, 설문 결과는 자동적으로 file 형태로 저장되어 추후 Excel 등의 spread sheet에서 통계적으로 처리할 수 있도록 하였다.

2.1. 음성검출과 특징 벡터 추출

음성신호 데이터의 획득은 PC의 사운드카드에 내장된 8비트 A/D 변환기를 이용하여 8kHz로 샘플링 하였다. 그림 2는 본 연구에서 구현된 시스템에서 피험자 음성의 특징벡터를 추출하는 과정을 나타낸다. 획득된 음성신호는 $1-0.95z^{-1}$ 의 전달함수를 갖는 프리엠펜시스(pre-emphasis)를 거친 후 프레임 양 끝단의 신호 정보를 보상하기 위하여 20ms 크기의 해밍 윈도우(Hamming Window)를 사용하여

10ms 씩 중첩 시켜서 이동시키면서 전처리하여 프레임 단위로 분할시킨다.

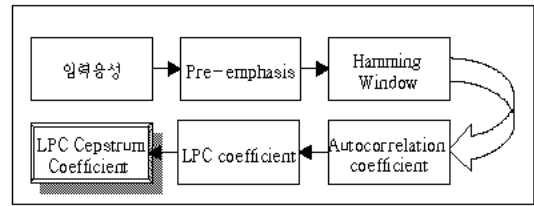


그림 2. 특징벡터 추출 과정[2]

음성신호는 평가어 단위별로 검출을 하게되고 인식 대상이 단어이므로 음성데이터가 입력됐을 때 음성신호의 유무를 확인하여 검출하여야 한다.

평가어에 대한 음성인식에는 이미 잘 알려진 DTW 알고리즘을 사용하였다. 동일한 사람이 동일한 단어를 발음한다 하여도 음성의 길이의 축약 및 비선형적인 확장이 일어나므로 두 개의 음성의 유사도를 측정하기가 어렵다. 그림 2에서와 같이 DTW는 이와 같이 음성의 발음길이의 변화가 발생하더라도 시간축에서 발생하는 차이를 보상하면서 기준 음성신호와 입력된 음성신호간의 유사도(Distance)를 동적 프로그래밍을 이용하여 구하는 방법이다[1].

전체적인 음성인식 시스템의 구성은 그림3과 같다.

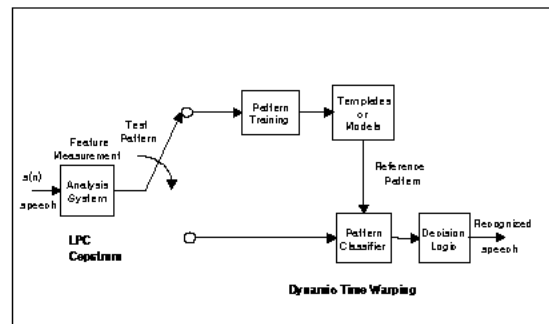


그림 3. 음성인식 시스템의 구성[3]

2.2. 설문지 자동 생성

주관평가의 자동화를 위해서는 평가어의 자동 인식 뿐만 아니라, 설문지 자체도 쉽게 작성하고 변경할 수 있어야 만이 다양한 조건 및 평가어에 대한 실험을 유연성있게 할 수 있게 된다. 따라서 본 연구에서는 평가어에 대한 음성인식 뿐만 아니라 평가어를 결정하여 이를 텍스트 파일로 작성하여 입력하면 5점 척도나 7점척도 또는 magnitude estimation 법에 의한 주관평가를 실시하기 위한 척도를 자동적으로 생성시켜 주는 평가설문지 자동 생성용 소프트웨어도 개발하여 손 쉽게 설문지를 생성하여 PC 상에서 모든 작업이 이루어 질 수 있도록 하였다.

본 시스템을 이용하여 생성된 설문지는 그림 4와 같다.

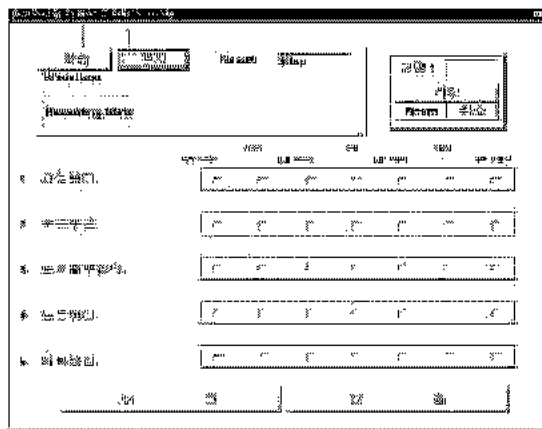


그림 4. 주관평가 자동화 시스템

3. 인식 실험 및 결과

3.1 프레임 수 및 특징벡터 차수에 따른 인식률 변화

본 논문의 주관평가 설문지 음성인식실험에 사용된 평가어로는 SD법에서 사용되는 표 1과 같은 “매우 아니다” ~ “매우 그렇다”의 7가지 평가어를 사용하였다.

음성신호는 평가어단위별로 검출을 하게되고 인식 대상이 단어이므로 음성데이터가 입력됐을 때 음성신호의 유무를 확인하여 검출하여야 한다.

표 1. 평가어 및 척도

척도	발음단어
-3	매우 아니다
-2	아니다
-1	조금 아니다
0	보통이다
1	조금 그렇다
2	그렇다
3	매우 그렇다

효율적인 음성검출과 좋은 인식률을 얻기 위해 알고리즘의 파라미터 결정을 위한 실험을 하였다. 우선, 특징 벡터의 차수에 따른 인식 실험과 기본 프레임에 따른 인식률의 변화를 평가하였다. 각각의 인식률 평가는 5명의 화자의 발음을 데이터베이스로 하여 특징벡터의 차수는 10 - 18 까지, 기본 프레임의 수는 10 - 25로 설정하고 각 피험자별로 25회씩 발음한 데이터를 획득하여 시뮬레이션하였다.

우선 음성 신호 부분만을 추출하기 위해서는 획득된 데이터에 대한 최적 길이 즉 프레임 수를 결정하여야 한다. 획득된 데이터들은 평가어의 어휘길이가 서로 다르므로 발음된 어휘 부분만을 우선 검출하기 위하여 한 프레임당 165 개의 데이터를 갖는 프레임 수를 결정하기 위하여 10, 15, 18, 20, 25 프레임 대항 평가어 인식실험을 실시하여 가장 인식률이 높은 15 프레임을 선정하였다.

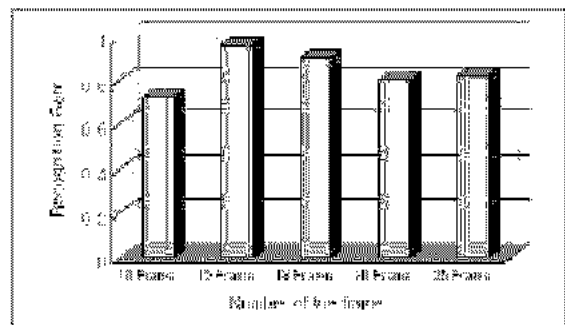


그림 5. 음성 검출 기본 프레임에 따른 인식률

그림 5는 “매우 아니다“ 부터 “매우 그렇다“ 까지 7 개의 평가어에 대한 프레임 수별 인식률을 대한 피험자 5명에 대한 실험 결과이다.

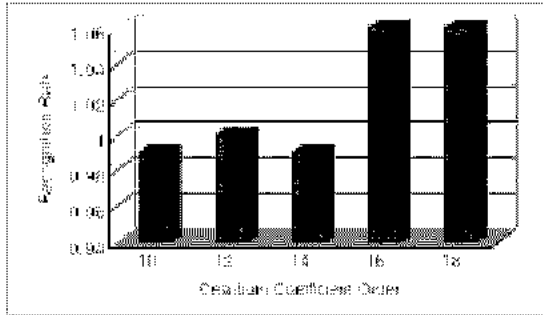


그림 6. Cepstrum 계수의 차수에 따른 인식률

한편 전처리된 데이터로 높은 음성인식율을 얻기 위해서는 적절한 LPC 켈스트럼 및 자기상관함수의 차수를 결정하여야 한다. 각 평가어에 대한 시물에선 결과 그림 6과 같이 16차가 가장 좋은 인식률을 나타내었으므로 차수는 16차로 하였다.

3.2 음성인식 평가

위의 실험결과를 바탕으로 결정된 프레임 수와 특징벡터의 차수를 이용하여 전체 인식 알고리즘을 구현하고 7점척도를 기준으로 하여 7가지 평가어를 대상으로 인식평가 실험을 하였다. 설문내용으로는 작업자의 피로도를 평가하는 내용이었다[4]. 인식에 사용된 평가어는 표 1과 같으며 피험자 10명(여자 5명, 남자 5명)에 대하여 실험시작 전 기준패턴을 저장한 후, 피험자로부터 평상시의 발음으로 표 1과 같은 고립어를 50회씩 발음하여 인식 실험을 하였다.

피험자는 표준과학연구원 내의 사무환경 평가실(소음 약 55dB 조건) 내에서 Task를 수행하고 음성인식을 이용하여 주관적인 감성평가를 하였다. 피험자의 평가가 잘못 인식될 경우를 대비하여 실험자가 외부에서 피험자를 감시하면서 실험을 실시하였다.

실험결과 96.4%의 인식률을 보였다.

4. 결과 및 고찰

주관평가의 자동화를 위하여 설문지의 자동 생성과 피험자 실험용 PC에서 실험용 Task의 제시와 주관평가를 함께 실시할 수 있는 음성인식기술을 이용한 주관평가 자동화 시스템을 구현하고, 실제 환경에서 7점 척도를 갖는 평가어에 대한 인식실험을 통하여 그 유용성을 확인하였다.

실시간으로 음성인식 평가를 한 결과 인식률 100%는 얻을 수 없었으나 실험환경에서 주관평가 데이터를 효율적으로 처리할 수 있고 음성을 이용함으로써 피험자의 정신적 부담을 경감시키며 생리신호와 주관평가와의 경시적인 변화를 관찰할 수 있다는 측면에서 만족할 만한 수준의 인식률을 보였으므로 효과적인 주관평가 방법의 하나로 생각할 수 있었다.

그러나 추후 보다 강건하고 인식률이 높은 알고리즘의 구현을 위한 연구는 계속되어야 하리라 판단된다.

*본 연구는 G-7 감성공학 기반기술개발사업에 의해 지원되었음(2000-J-ES-02-A-01)

참고문헌

- [1]한화영,고한우,윤용현,조택동(2000), 음성인식을 이용한 주관평가의 자동화에 관한 기초연구, 한국감성과학회2000년 추계학술대회 발표논문집, 113-117.
- [2]Lawrence Rabiner, Biiling-Hwang Juang (1993), Fundamentals of Speech Recognition, Prentice Hall PTR,
- [3]김 문열(1998), 음성인식을 이용한 스테핑 모터 구동에 관한 연구, 석사학위논문.
- [4]이창비,고한우,윤용현(2000), 단조작업시 정신 피로도 측정을 위한 한국어판 질문지에 관한 연구, 한국감성과학회2000년 추계학술대회 발표논문집, 195-202.