한국감성과학회2000추계
학술대회발표논문집

음성인식을 이용한 주관평가의 자동화에 관한 기초연구

한화영*,고한우**,윤용현**,조택동*

*충남대학교 기계설계공학과, **한국표준과학연구원 인간정보그룹

A Basic Study on Automation of the Subjective Evaluation using Speech Recognition

Hwa Young Han, Han Woo Ko, Yong Hyeon Yun, Taik Dong Cho

*Dept. of Mechanical Design, Chungnam Univ.
**Ergonomics Lab, Korea Research Institute of Standards and Science

Abstract

수작업으로 이루어지고 있는 환경의 영향이나 작업의 영향에 따른 정신피로나 신체피로의 주관적인 평가를 자동화하기 위한 방법에 대하여 논하였다. 사람의 가장 자연스러운 의사소통인 평가어를 척도로 하여 평가가 이루어지는 음성인식기술을 응용한 주관평가법에 대하여 연구하였다. 주관평가의 자동화를 위하여 우선, 평가어에 대한 음성 인식을 한 후 인식된 평가 결과 데이터를 이용하여 설문지를 자동 생성시킴과 동시에 파일 형태로 저장시켰다. 음성 인식 알고리즘으로는 DTW(Dynamic Time Warping)인식 알고리즘을 사용하였고. 설문지 질의 내용은 집중도 평가를 이용하였다. 인식실험은 설문에 대한 응답에 필요한 평가어를 대상으로 하였다.

Keywords: 음성인식, DTW, 설문지

1. 서론

인간이 가장 편안한 느낌을 가질 수 있는 환경조성으로는 많이 접하였거나 친숙한 환경이나 일상적인 행동유발을 할 수 있는 여건 등과 같은 것들이 있다. 이런 이유로 주관적인 평가를 위한 설문지 조사를 할 때 심리적으로 안정된 조건을 조성해 주어야 한다.

심리적인 변화나 환경에 민감한 생체신호와 같은 경우 안정적인 환경조성이 필수적이고 모든 사람마다 조건이 동일해야만 한다. 그러나 지금까지의 경우 작업자가 직접 작성하거나 컴퓨터를 통해 입력하는 방식이 사용되어 왔다.

본 논문에서는 편리하면서도 거부감이 없는 방법의 하나로 음성을 이용하였다. 사람에게 가장 일반적인 의사소통 수단인 음성을 인식하여 설문을 진행해 나가는 방법에 착안하여 음성인식을 이용한 감성 평가 시스템을 개발하고 이의 유용성을 검토하였다.

설문유형으로는 5점척도와 7점척도와 11개의 단어(숫자)를 이용하여 평가하는 SD법과 ME법을 사용하였다. 설문에 사용되는 음성의 특징벡터로는 12차 LPC 켑스트럼 계수를 사용하였으며 인식방법으로는 화자 종속 고립어 인

식시스템의 구성에 주로 이용되며 인식률이 높다는 장점을 가지고 있는 알고리즘인 DTW(Dynamic Time Warping)알고리즘을 이용하였다.

2. 주관 평가 시스템의 자동화

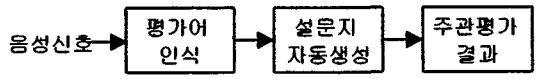

그림 1. 주관평가 자동화 시스템

전체적인 주관평가 시스템의 구성은 그림 1과 같다. 음성신호가 마이크를 통해 입력되면 평가어를 인식하고 설문지를 자동 생성한다. 피험자가 설문을 평가하면 그 평가 결과가 데이터파일로 생성된다.

2.1. 음성인식

2.1.1. 전처리 과정과 특징벡터 추출

음성신호 데이터의 획득은 PC의 사운드카드에 내장된 8비트 A/D 변환기를 이용하여 8kHz로 샘플링 하였다. 그림 2는 특징벡터를 추출하는 과정을 나타낸다. 획득된 음성신호는 $1-0.95z^{-1}$의 전달함수를 갖는 프리엠퍼시스(pre- emphasis)를 거친 후 20ms 크기의 해밍 윈도우(Hamming Window)를 사용하여 프레임 양 끝단의 신호 정보를 보상하기 위하여 10ms씩 중첩을 시켜서 이동시키면서 전처리하여 프레임 단위로 분할시킨다. 전처리된 데이터로부터 12차의 자기상관함수와 12차 선형예측계수를 구한 후 음성 인식 시스템의 특징벡터(feature)로 사용할 12차 LPC 캡스트럼 계수(LPC Cepstrum coefficient)를 추출하였다.

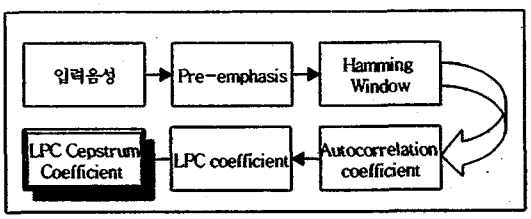

그림 2. 특징벡터 추출 과정

2.2. DTW(Dynamic Time Warping) 알고리즘

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

단어 음성패턴 A, B는 일련의 특징벡터로서 적절한 특징추출에 의하여 다음과 같이 표현할 수 있다[2].

$$A = a_1, a_2, \cdots a_i, \cdots a_I$$

$$B = b_1, b_2, \cdots b_i, \cdots b_J$$

두 패턴 A와 B의 길이는 각각 I, J 로 길이가 서로 다르다. 이와 같은 길이의 차이는 일련의 점인 와핑함수(Warping Function) $c=(i, j)$로 묘사될 수 있으며 전체 와핑함수 F는 다음과 같이 나타낼 수 있다.

$$F = c(1), c(2), \cdots, c(k), \cdots, c(K)$$

여기서 $c(k)=(i(k), j(k))$이며 한 시점 k에서 패턴 A와 B 사이의 거리를 최소로 하는 점이다.

이 와핑함수는 음성패턴 B의 시간축상에 음성패턴 A의 시간축을 매핑하는 것으로 볼 수 있으며 두 특징벡터 사이의 거리는 다음 $d(c)$와 같이 계산된다.

$$d(c) = d(i,j) = \|a_i - b_j\| = \sqrt{\langle a_i - b_j, a_i - b_j \rangle}$$

$$= \sqrt{(a_{i_1} - b_{j_1})^2 + (a_{i_2} - b_{j_2})^2 + \cdots + (a_{i_p} - b_{j_p})^2}$$

와핑함수 F에 관한 거리의 합계는 $E(F)$와 같다.

$$E(F) = \sum_{k=1}^{K} d(c(k)) \cdot w(k)$$

와핑함수 F는 최적으로 시간차이가 조정될 때 최소값을 가진다. 두 음성 패턴 A와 B 사이에 대한 시간축 정규화거리 $D(A,B)$는 다음과 같다.

$$D(A,B) = \underset{F}{Min}\left[\frac{\sum_{k=1}^{K} d(c(k)) \cdot w(k)}{\sum_{k=1}^{K} w(k)}\right]$$

분모는 $\sum_{k=1}^{K} w(k) = N$ 으로 정의할 수 있으며 와핑함수에 관련된 무게값 다음과 같이 표현된다.

$$w(k) = (i(k) - i(k-1)) + (j(k) - j(k-1))$$

$N = I + J$ 가 되어 $D(A,B)$는

$$D(A,B) = \frac{1}{N} \underset{F}{Min}\left[\sum_{k=1}^{K} d(c(k)) \cdot w(k)\right]$$

여기서 와핑함수 F는 모든 프레임에 대하여 구하지 않고 다음과 같은 조건에서 일반적으로 구한다.

① 단조 증가조건

② 경계조건

$$i(1) = 1, \quad j(1) = 1$$

$$i(K) = I, \quad j(K) = J$$

③ 조정창 조건

전역 경로 제한 조건인 조정창 조건은 $|i(k) - j(k)| \le r$로 표시되며 전역 경로 제한에 의해 최적 경로 탐색을 위한 범위를 제한한다.

DTW을 이용한 음성인식 시스템은 고립어 인식시스템 구성에 주로 이용되며 인식률이 높다는 장점이 있는 반면, 단어 수량이 증가하면 계산량이 상당히 방대하고 융통성이 부족해 고립단어 인식 방식 이외에는 적용하기 어려운 단점이 있다.

주관평가에서는 5개 ~ 11개의 평가어가 사용되므로 음성 인식을 하는 실험을 하기에 적합한 DTW알고리즘을 사용하였다.

전체적인 인식 시스템의 구성은 그림 3과 같다. 전처리 과정에서 구해진 12차 LPC 캡스트럼 계수를 이용하여 그림 3의 상단부분에서 기준 패턴을 생성하고 실제 인식과정을 수행할 경우에는 하단부분의 기준 패턴과 입력된 음성과의 유사도를 DTW(Dynamic Time Warping)알고리즘을 이용하여 인식한다.

Decision logic으로는 DTW 알고리즘의 유사도를 이용하여 제한한다.

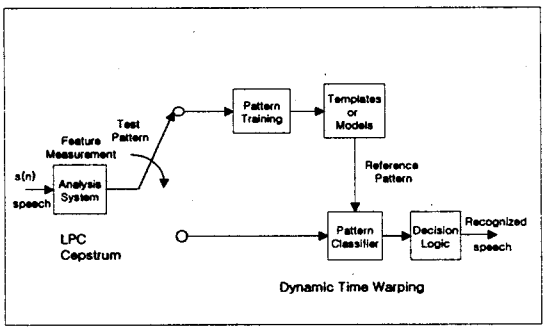

그림 3. 음성인식 시스템의 구성[3]

2.3. 설문지 자동 생성

감성평가를 할 경우 피험자의 환경이나 조건에 많은 영향이 있으며 음성을 인식할 때에도 주변환경의 영향이 크다. 감성 평가 시스템을 구현하기 위해서는 그 설문의 유형이나 환경을 고려해야 하는데 본 실험에서는 실제 사무환경을 대상으로 하였고 설문 유형으로는 집중도 평가에 관한 5점척도와 7점척도를 사용하였다. 생성된 설문지는 그림 4와 같이 표현된다.

이와 더불어 ME법 설문지인 "그렇지 않다" ~ "매우 그렇다"의 구간으로 각각 0 ~ 100의 득점을 하도록 되어있는 설문지[4]를 10점씩 증가하여 11개의 평가어를 대상으로 설문유형을 형성하였다.

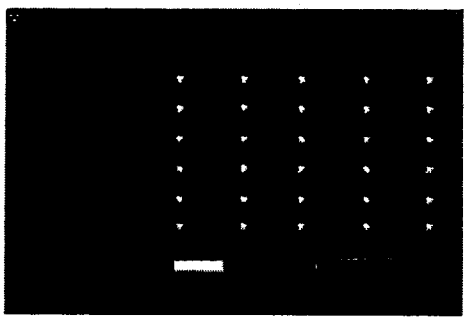

그림 4. 5점척도 설문지

3. 인식 실험 및 결과

논문의 설문지 음성인식실험에 사용된 평가어로는 ME법에서는 "그렇지 않다" ~ "매우 그렇다"의 구간으로 각각 0 ~ 100의 득점을 하도록 되어있는 설문지를 10점씩 증가하여 발음한 표 1와 같은 숫자단어를 사용하였다. 또한 SD법에서는 표 2와 같은 "매우 아니다" ~ "매우 그렇다"의 7가지 평가어를 사용하였다.

표 1. 숫자 단어

숫자음	영, 십, 이십, 삼십, 사십, 오십, 육십, 칠십, 팔십, 구십, 백

표 2. 발음에 따른 척도

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

본 실험에서는 5점척도와, 7점척도, 11개의 숫자음을 인식하는 3가지의 인식실험을 하였다. 인식에 사용된 평가어는 표 1과 표 2와 같다. 피험자 5명(여자 2명, 남자 3명)에 대하여 실험 시작 전 피험자로부터 평상시의 발음으로 표 1과 표 2와 같은 고립어를 20회씩 발음하게 하고 그 중 임의로 하나씩 취하여 기준 패턴으로 하였다.

5점척도의 경우 평균적으로 표 3에서 보는 것과 같이 의 96%의 인식률을 보였고 7점척도의 경우에는 96.31%의 인식률을 보였다.

피험자를 chamber내에서 Task 수행을 하고 음성인식을 이용하여 주관적인 감성평가를 하였다. 피험자의 평가가 잘못 인식될 경우를 대비하여 실험자가 외부에서 피험자를 감시하면서 실험을 실시하였다.

표 3. 인식률 결과

	5점척도	7점척도
인식률	96%	96.31%

숫자음의 경우 좋은 결과를 내지 못했다. 그 이유로는 본 논문의 인식시스템에서 무성음 단위인 칠십이나 팔십과 같은 단어를 제대로 인식하지 못하는 경향을 나타내었다. 이는 음성의 녹음에서 울림현상을 방지하지 못했기 때문이며 음성신호 자체의 문제이므로 앞으로 이에 대한 개선이 필요하다.

4. 결과 및 고찰

궁극적인 연구목적인 음성인식기술을 이용한 주관적인 감성평가 시스템의 개발을 위한 기초연구로서 5점척도와 7점척도의 인식실험을 하였고 그 인식률은 어느 정도 만족할 만한 수준인 96%정도의 인식률을 보였으나 숫자음의 경우 음성데이터를 녹음하는 과정에서부터 시작하여 전체적으로 인식하는 데에 어려움이 있었다.

따라서 주관적인 평가시스템의 구현에서는 간단한 설문지 유형일 경우 음성인식을 이용하기에 적합하였다. 피험자가 직접 주관평가를 하였을 때 만족할 만한 수준의 결과를 보였다.

수작업으로 이뤄지고 있는 주관평가를 자동화하기 위해 인식알고리즘의 개선과 인식률 향상에의 지속적인 연구가 필요하다.

참고문헌

[1] 김 문열(1998), 음성인식을 이용한 스태핑 모터 구동에 관한 연구, 석사학위논문.

[2] HIROAKI SAKOE and SEIBI CHIBA(1978), Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 1.

[3] Lawrence Rabiner, Biiling-Hwang Juang (1993), Fundamentals of Speech Recognition, Prentice Hall PTR, 51.

[4]이창미,고한우,윤용현(2000), 단조작업시 정신피로도 측정을 위한 한국어판 질문지에 관한 연구, 한국감성과학회2000년 춘계학술대회 발표논문집, 195-202.