

청각장애인을 위한 감성자막에 관한 연구

장성모[○], 배영임^{*}, 안영기^{*}, 이상문^{*}

^{○*}충주대학교 컴퓨터정보학과

e-mail: jsmforjism@naver.com, you5613@hanmail.net, gzxxxx@cjnu.ac.kr, smlee@cjnu.ac.kr

A Study on the Emotional Caption System for Auditory Disabled Person

Seung Mo Jang[○], Young Rim Bae^{*}, Young Ki An^{*}, Sang Moon Lee^{*}

^{○*}Dept. of Computer Info. Engineering, Chungju Nat'l University

● 요약 ●

현재 출시되어 있는 대부분의 콘텐츠들은 일반인들의 감각, 감성 욕구를 채워주고 있다. 하지만 사회의 소외계층인 장애인들을 위한 콘텐츠는 그리 많지 않으며 배제되어 왔다. 이 논문에서는 시각 장애인들을 위한 감성 전달 서비스 기법을 제시하였다. 기존의 콘텐츠는 자막 단순화로 인해 소리를 들을 수 없는 청각장애인들에게 매우 불편함을 주고 있다. 따라서 이 논문에서는 청각 장애인은 물론 일반인들에게도 소리를 들을 수 없는 환경에서 최적의 감성 전달 서비스를 제공할 수 있는 감성 지원 캡션 시스템을 구현하여 제시한다.

키워드: Daubechies, 단구간 영교차율(ZCR, zero crossing rate), K-means, 레이블링(Labeling)

I. 서론

인간은 일반적으로 시각, 청각, 촉각 등의 다양한 방법을 통하여 상호간에 정보를 교환한다. 감정/감성의 전달은 공감각적 심상을 이용한다. 예를들면 맛있는 냄새가 나는 주황 리본 그림 같은 느낌을 받게 되면 사람들은 “아! 시탕 맛의 주황색 예쁜 리본이 있구나.” 하고 느끼게 될 것이다.

이런 방식의 특성을 이용하여 영화나 여러 미디어, 쇼 프로그램, 동영상 등을 감상 할 때 활용하고 있는 단일색의 자막이 아닌 감성적인 색깔의 자막과 음성 데이터를 분석하여 음성의 세기에 따라 자막의 크기를 조절함으로써 청각 장애인이나 음악을 켤 수 없는 공공장소에서 일반 사용자들이 전달하고자하는 감정/감성을 좀 더 구체적으로 받아들일 수 있도록 하는 부분의 연구를 진행하였다.

II. 관련 연구

1. 색체에 관한 심리적 작용

색은 그 자체로도, 그리고 형태와 결합이 되었을 때도 사람에게 큰 영향을 주는 힘을 갖고 있으며, 그러한 작용에 대한 사람의 변화를 연구하는 것이 색체심리이다[1]. 색체의 심리적 효과로는 색의 이미지 연관성, 인상, 색의 특성 등의 시각적 분위기를 조성하기 위해 미리 결정되어지는 부분이다. 색체의 형상들은 우리 몸의 눈이라는 시각적 감각기관에서 다양하게 영향을 미칠 뿐 아니라

시각을 매개로하여 감정, 정서에도 영향을 준다. 색체를 지각하는 것은 우리의 마음속 내면에 존재하는 것이기에 심리학적으로도 접근할 수 있다. 만약 특정 색체에 대한 개인적인 선호도를 색체에 대한 감정, 정서적 반응 등 효과를 이해할 수 있게 도와주게 된다면,

표 1. 주요색의 심리적 작용
Table 1. Psychological action of important color

| 색 | 심리적 작용 |
|-----|------------|
| 빨간색 | 불안과 긴장의 증가 |
| 주황색 | 식욕촉진 |
| 노란색 | 방어적인 느낌 |
| 초록색 | 마음의 평온 |
| 파란색 | 인락함 |
| 자주색 | 정서적으로 중립 |
| 흰색 | 균형감 |
| 회색 | 수동적인 느낌 |
| 검은색 | 부정적인 느낌 |

파버 비렌은 표 1과 같이 주요색에 대한 심리적 작용을 정의하였다. 이것은 색체에 대한 심리적 작용이 중요하다는 것을 단적으로 보여 주고 있다.

이산 웨이블릿 변환은 고역 통과 부분을 한 단계의 필터 뱅크로 구성하였고, 저역통과 부분은 계속적인 필터뱅크로 확장하는 옥타

브 밴드(octave-band)구조와 고역통과 부분도 필터뱅크로 확장하는 구조를 가지는 웨이블릿 패킷(wavelet packet)구조로 구현될 수 있다[3].

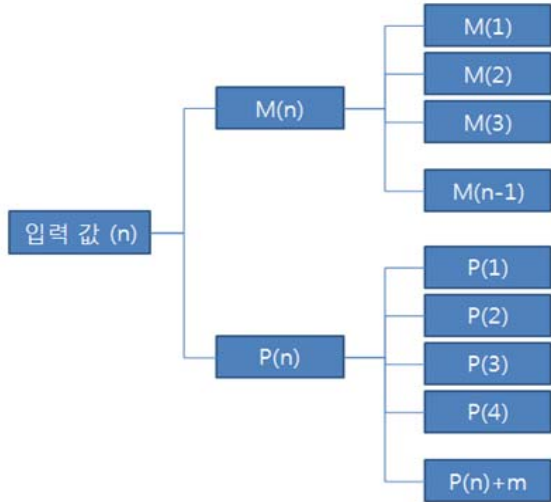


그림 1. 웨이블릿 패킷의 구조
Picture 1. Structure of wavelet packet

그림 1에서처럼 번호를 부여 할 때는 주파수 대역을 균등하게 분할하는 방식인 웨이블릿 패킷 구조 방식을 사용 하였으며, M(n)의 남성음성의 경우 매우 낮은 인식률을 보인 고주파대역 신호인 M(n)을 제외했다.

여성음성의 경우 P(n)은 남성보다 다양한 음의 높낮이를 보였으며 같은 대역의 음을 비교해 보았을 때 높낮이에 따른 감정의 차이가 많았다. 이 논문에서는 남성과 여성의 감정을 DB에 각각 등록하였으며 목소리의 음정에 따라 남자와 여자화자를 따로 구별 하였다. 그리고 가장 널리 사용되고 있는 Daubechies 기저함수를 이용하여 신호를 해석하였다. 또한, 음성의 시작과 끝점을 찾아내는 방법으로 예측계수나 자기상관계수와 같은 음성특징 계수를 사용할 수는 있으나, 이 연구에서는 현재 음성 끝점을 위해 일반적으로 사용되는 방법은 단구간 에너지와 단구간 영교차율(ZCR, zero crossing rate)을 사용하였다.

이렇게 획득된 음성신호는 웨이블릿 필터링 기법을 이용하여 각각의 대역별로 출력 값을 획득 한다. 웨이블릿 필터에서 출력되는 음성신호는 음성 분석부에서 특징벡터로 FFT기반 멜캡스트럼 계수를 구한 후 K-means 알고리즘을 이용하여 독립적인 코드북을 미리 만들어 놓는다. 이때, 감정 인식의 향상을 위해 남성화자와 여성화자의 코드북을 각각 만들어 놓았다.

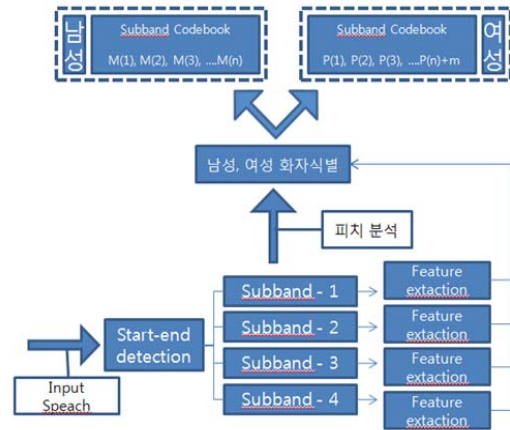


그림 2. 웨이블릿 필터뱅크를 이용한 감정인식기
Picture 2. The sentimental identifier which uses wavelet filter bank

그림2는 이 논문에서 제안한 웨이블릿 필터뱅크를 이용한 감정 인식기를 나타낸 것이다. 인식과정에서는 인식하고자하는 음성신호가 인식되면 웨이블릿 변환하여 주파수별로 음성신호를 분할한다. 그리고 성별을 구분하여 DB에 저장된 코드북과 비교하기 위해 저주파대역에서 피치를 이용한 성식별을 분석한 후 음성 분석부에서 각각의 주파수 대역에 대한 특징벡터를 계산한다. 이와 같이 음성분석부에서 계산된 특징 벡터는 미리 뱅크별로 만들어 놓은 코드북과의 거리를 계산한 후 독립적인 인식률을 산출한다.

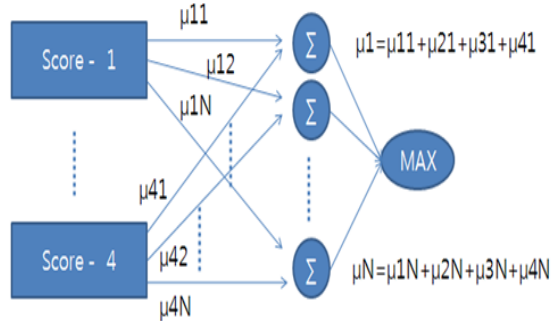


그림 3. 다중 밴드 의사결정 방법
Picture 3. Multiplex band Decision-making method

그림3은 다중밴드 의사결정 방법에 대해 보이고 있다[4]. 각 대역별에서 산출된 인식률은 음성신호를 프레임으로 나누고 각각의 프레임에서 얻어진 특징벡터와 코드북과의 거리 계산에 의하여 산출되기 때문에 어느 특정 감정에 대한 정보만을 가진 것이 아니라 인식하고자 하는 각각의 감정 정도에 대한 소속정도를 모두 가지고 있다.

2. 음성의 세기에 따른 자막 크기 조절

일반적으로 사람이 듣고 말 할 수 있는 음성 주파수 대역의 범위는 200 ~ 3,000Hz 이다. 매체에서 자막에 따른 음성이 나올 때 시스템 상에서는 이 음성을 데이터베이스화 하여 음의 세기를 조절한다.

DB 구하는 공식 = $10 \log I/I_0$

(I: 어떤 지점의 음의 세기를 나타냄, I_0 : 가청문턱 값으로 우리가 들을 수 있는 가장 작은 세기의 소리)

예를들어 $I = 1 \times 10^{-5} \text{ W/m}^2$ 을 대입하여 dB을 구하면

$$\begin{aligned} \text{dB} &= 10 \log 1 \times 10^{-5} / 1 \times 10^{-12} \\ &= 10 \log 1 / 1 \times 10^{-7} = 10 \log 10^7 \quad (1) \\ &\therefore 70 \text{ dB} \end{aligned}$$

이러한 수식에 의해 자막의 크기를 사람의 목소리 역양에 따라 조절해 준다.

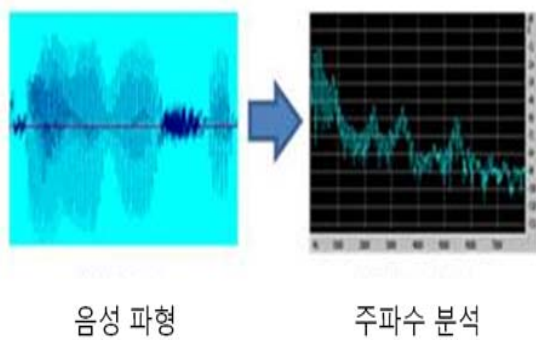


그림 4. 음성파형에 따른 주파수 분석

Picture 4. The frequency analysis which follows in the vocal wave shape

그림4의 매체의 음성을 추출 후 주파수 파형을 분석하고 범위를 DB로 변환하여 DB에 저장된 글자의 크기를 지정해준다. 예를 들어 인식결과가

10Hz면 “야호!” 20Hz면 “야호!” 30Hz면 “야호!” 로 표시된다. 이런 글꼴의 크기로 영화나 동영상의 자막에서 음성의 세기를 대처 할 수 있게 된다. 특히, 청각장애인의 경우 들을 수 없었던 역양의 핸디캡을 여기서 보완 할 수 있다. DB에는 각각의 Hz영역 별로 레이블링을 이용하여 번호를 매겨준다.

III. 설 계

기존의 동영상, 영화, 쇼프로그램 등의 인간의 감성을 전달하는 매개체를 분석하여 자막의 색상, 크기를 이용하여 청각장애인들의 정서적인 풍요로움을 전달 할 수 있다. 이 연구에서는 타 미디어 플레이어와는 다른 서비스를 제공 할 수 있다.



그림 5. 감성 자막의 구현

Picture 5. reali Emotional Caption System

그림 5에서는 구현이 완료되고 난 후의 완성된 프로토타입을 적용하였다. 이 동영상에서는 감성의 기록에 따라 자막의 색깔이 의미구로 나누어 적용되며 각각의 구에는 음성의 세기에 따라 자막의 크기가 변화한다.

자막을 이용하여 감성을 전달하기 위해 음성의 세기 빠르기 톤을 이용하여 감성을 수집, 판별하여 DB에서 파형을 분석하게 된다. 이 분석한 자료를 레이블링을 이용하여 각각의 영상의 픽셀에 고유한 번호를 매기는 작업을 하고 영상과 DB의 매칭을 통해 각 군집에 번호를 부여한다. 레이블링 할 때 레이블 번호는 DB에서 지정한 컬러 범위에 해당하는 번호를 부여하게 된다.

각 영상의 자막의 크기를 판별하여 DB에서 음의 세기를 통한 화자(캐릭터)들의 상태나 어떤 것을 강조하는 것인지 직접 듣지 않고서도 사용자는 느낄 수 있으며 전달하고자 하는 내용의 핵심을 짚을 수 있게 된다.

IV. 결론 및 향후과제

이 논문에서는 다양한 동영상 및 쇼 프로그램, 영화를 감상하는 청각 장애인들과 음악을 직접 들을 수 없는 상황에서나 공공장소의 일반인들에게 까지 다양한 형태로 도움을 줄 수 있는 콘텐츠를 대상으로 연구를 시작해 보았다. 현재 음성인식 기술은 스마트폰과 유비쿼터스 그리고 내비게이션 까지 다양한 분야에 실제적으로 적용되고 있으며 그 범위도 기하급수적으로 확장 되고 있다. 하지만 아직도 적재적소에 다양하게 개발·적용이 된 것은 그리 많지 않다. 향 후 스마트 폰이나 아이패드 등에도 적용이 가능하도록 연구되어야 할 것이다.

참고문헌

- [1] 이세리, “음향오해를 통한 색채 심리 방법론 제시”, 한국색채 디자인학회, 한국색채디자인학연구 Vol 1 no1 통권 제1호, 129~136쪽, 2005.
- [2] 송영민, “노인여가복지시설에있어서 환경색채의 영향에 관한 연구”, 홍익대학교 석사학위논문, 2009.
- [3] Stephane Mallat, “A wavelet tour of signal processing”, Academic press, 1999.
- [4] 이대중, 곽근창, 유정웅, 전명근, “웨이블릿 필터뱅크에 기반을 둔 강인한 화자식별 기법”, 정보처리학회 논문지 C 제9-C권 제4호, 2002.
- [5] 신성윤, 이경주, 이양원, “영상에서의 감성인식과 아동 색채 심리”, 한국컴퓨터정보학회 논문집 18권 2호, 133~136쪽, 2010.
- [6] 고현주, 이대중, 전명근, “얼굴표정과 음성을 이용한 감정인식”, 정보과학회논문지 제31권 제6호, 799~807쪽, 2004.