

미디어 처리 기법을 통한 감정 및 심리 상태 분석 방법의 제안

조동욱*, 김봉현**, 이세환**, 가민경**, 박선애**, 오상영***

*충북과학대학 정보통신학과

**한밭대학교 정보통신전문대학원 컴퓨터공학과

***청주대학교 경영정보학과

e-mail : ducho@ctech.ac.kr

A Proposal of Emotion and Psychology Analysis Method Using Media Process Technology

Dong-Uk Cho*, Bong-Hyun Kim**, Se-Hwan Lee**, Min-Kyoung Ka**,
Sun-Ae Park**, Sang-Young Oh***

*Dept. of Information & Communications Engineering, Chungbuk
Provincial University

**Dept. of Computer Engineering, Hanbat National University

***Dept. of Management Information System, Cheongju University

요 약

인간의 감정과 심리 상태를 분석하는 것은 다양한 분야에서 사용되고 있는 응용 기술이다. 이와 같은 분석 분야는 최근 들어 오감의 연계와 결과의 이용에 따른 맞춤형 기술로 다양화 되어 가고 있다. 따라서 본 논문에서는 감정 및 심리 상태를 분석하기 위한 방법 중 흔하게 접할 수 있는 얼굴 영상과 음성을 통한 실험을 진행하고자 한다. 이를 위해 기존의 영상학적, 음성학적 분석 방법을 정립하고 감정 및 심리 상태를 적절한 배경에 맞춰 분석이 가능한 형태로 결과값을 출력하는 기법을 제안하고자 한다. 또한 실험을 통해 추출한 연구 과정 및 방법을 이용하여 실제 사례에 대한 분석 기법의 적용을 도입하고자 한다.

1. 서론

최근 사람의 음성과 얼굴 표정 및 생체데이터를 감지하여 사용자의 감정을 인식하는 관련 기술이 증가하고 있다. 감정인식 기술은 심리학자들이 분류한 기본적인 6개의 감정인 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 인식하는 것으로, 음성 및 얼굴 표정을 비롯하여 뇌파, 맥박, 체온 등의 생체 신호로부터 인간의 감정을 인식한다. 특히, 이 기술은 디지털 가전 및 미래형 주거환경과 밀접한 관련기술이기 때문에 최근 많은 주목을 받고 있다. 사람이 무엇을 원하는지 알아차리고 그에 대한 적절한 행동을 취하는 스마트한 기계가 되기 위해서는 인간과 같은 오감이 필요하다. 지금까지 기계에 의한 감정인식 기술은 주로 시각과 청각에 초점을 맞춰 연구 개발 되고 있다. 사람의 수많은 감정이 시각이나 청각, 혹은 후각이나 촉각 중 어느 하나의 감각만으로는 표출되기

어렵지만 인간의 감정을 표현하는 데 있어 표정과 목소리에 의지하는 면이 많다.

감정인식 기술은 기기가 어떠한 센서를 통해서 인간의 얼굴 표정 변화나 음성 등을 체크해서 현재의 감정 상태를 결정하여 그에 따라 맞춤 서비스를 제공하는 것이 핵심이다. 최근에는 로봇에 감정인식 기술이 적용되어 인간의 감정을 파악해 커뮤니케이션을 하는데 기능을 발휘하고 있다. 이러한 감정인식 기술은 기계들을 논리의 제한을 받는 노예 상태에서 사려 깊고 관찰력 뛰어난 협력자로 변화시킬 수 있다. 감정은 우리가 중요하게 여기는 모든 관계에서 나타난다. 우리는 직장 사람들에게, 친구들에게, 가족에게, 연인이나 배우자에게 감정을 느낀다. 감정은 우리의 목숨을 구해주기도 하지만, 극심한 상처를 입힐 수도 있다. 감정은 우리에게 현실적이면서도 적절하다고 판단되는 행동을 하게 만들기도

하지만, 때로는 나중에 가서 심하게 후회할 행동을 하게 만들기도 한다[1].

2. 감정 및 심리 상태 분석 방법

2.1. 영상 분석 방법

영상을 이용한 감정 인식을 위해서는 얼굴 영상을 분석하여 특징을 찾아내는 알고리즘 개발이 필수적이다[2][3][4]. 이에 음성은 청각에 기반을 둔 가장 효율적이고 자연스러운 휴먼 컴퓨터 인터페이스로 기대되고 있는 분야로, 심리학자인 Ekman과 Friesen에 따르면 사람의 여섯 가지 감정인 행복, 슬픔, 화남, 놀람, 혐오, 공포는 각 문화에 영향을 받지 않고 공통으로 인식되는 기본 감정이라 하였다[5]. 기존의 연구 사례를 살펴보면 레오나르도 다 빈치의 명화 ‘모나리자’의 얼굴은 행복한 표정이었다고 컴퓨터가 판정을 내렸다. 이 흥미로운 연구의 주인공인 네덜란드 암스테르담 대학의 연구진들은 여성들의 표정을 분석해 거기에 담긴 감정을 판단해내는 표정 판독 소프트웨어를 개발, 모나리자의 얼굴을 분석한 결과를 내었다. 그 결과, 모나리자의 얼굴 표정은 감정 요소 중 83%가 행복, 9%가 두려움, 2%가 분노임 것으로 분석됐다.

연구팀은 지중해 지역 혈통 여성 10명의 얼굴을 분석해 중립적 표정의 기준을 정립한 뒤 행복, 놀라움, 분노, 혐오, 공포, 슬픔 등 6가지 감정을 지수화해 비교했다. 분석에 사용된 소프트웨어는 입술 선의 휘어짐과 눈가의 주름을 분석해 여성의 얼굴 표정을 분석하는 일리노이대의 개발 소프트웨어를 부분적으로 이용했다. 전문가들은 얼굴 인식 기술이 급속도로 발전하고 있기 때문에 이번 연구 결과가 최종적인 것으로 굳어지지는 않겠지만 흥미 있는 연구라고 논평했다. 표정분석 소프트웨어는 사용자의 기분을 읽고 그에 맞춰 반응하는 컴퓨터를 개발하는 중요한 응용 사례가 될 것으로 보인다[6].

2.2. 음성 분석 방법

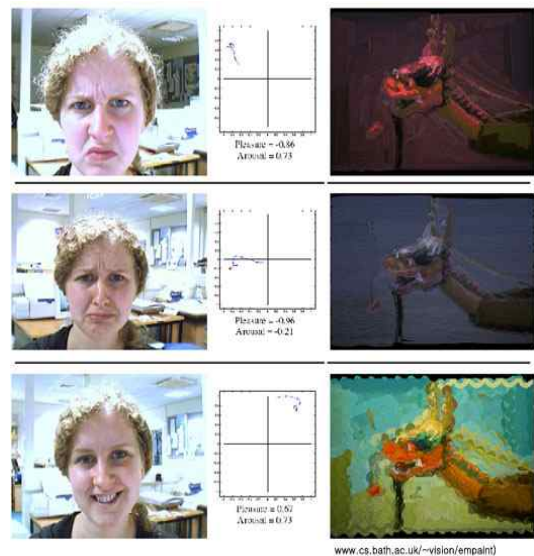
음성은 청각에 기반을 둔 효율적이고 자연스러운 방법으로 여기에 내포된 감정을 추출하려는 연구가 활발히 행해지고 있다. Fukuda는 음성신호의 템포와 에너지를 가지고 여섯 개의 기본 감정에 대한 분류를 시도하였는데, 녹음실과 같은 외부 잡음이 전혀 없는 환경하에서 일본어와 이탈리아어에 대한 음성신호를 녹음한후 감정 추출에 대한 연구를 하였다

[7]. Moriyama는 음성신호의 피치(Pitch)와 전력의 포락선 검출을 통하여 20개의 일본어 샘플에 대하여 실험하였고, 실험 결과 ‘화남’, ‘슬픔’, ‘놀람’ 감정이 다른 감정보다 인식률이 비교적 높은 것으로 나타났다[8]. 표정은 그 사람이 현장을 떠나지 않는 한, 혹은 갈수록 드물어지기는 하지만 가면이나 베일을 써야하는 사회가 아닌 한, 언제나 눈으로 볼 수 있다. 반면에 목소리는 간헐적인 세계로, 언제든지 자유자재로 완벽하게 제어할 수 있다. 사람이 무엇을 원하는지 알아차리고 그에 대한 적절한 행동을 취하는 눈치 빠른 기계가 되기 위해서는 인간과 같은 오감이 필요할 것이다.

3. 분석 사례 및 결과

3.1. 영상 분석 결과

컴퓨터 화면 속의 그림이 내 표정을 따라 변화한다는 내용의 영국 바스 대학 연구진이 개발한 “감정 반응 그림 Empathic Painting”이 BBC 등의 언론에 소개되면서 화제를 모으고 있다. 이 시스템의 소프트웨어는 먼저 웹캠을 통해 사람의 표정을 인식한 후 사람의 감정 상태에 따라 변화된 그림의 모습을 보여준다. 화남 표정이면 컴퓨터 모니터의 그림은 어두워지고, 사람이 행복한 표정이면 그림도 밝은 색상으로 변한다. 그림은 사람의 표정에 실시간으로 반응하는데 소프트웨어가 입과 눈과 눈썹 모양 등을 읽어 분노, 행복감, 무기력 상태 등을 파악하여 그림을 변화시킨다.



(그림 1) 감정 반응 연구 결과



(그림 2) 얼굴 심리학 표정 연구 결과

기쁨과 슬픔, 걱정과 두려움 등 인간의 감정은 표정으로 나타난다. 비언어 의사소통 전문가인 폴 에크먼(샌프란시스코 의대 심리학 교수)에 따르면 인간은 얼굴의 2개 근육만으로 무려 300가지 표정을 만들어낼 수 있다. 3개 근육으로는 4000가지, 5개 근육을 조합하면 1만개 이상의 표정을 만들 수 있다고 한다. 에크먼은 이처럼 다양한 조합 중에서 3000개의 표정에 사용된 근육에 번호를 붙이고 1978년 세계 최초로 얼굴지도인 '얼굴 움직임 해독법'(FACS)을 만들었다. '얼굴 심리학'은 표정연구의 대가인 에크먼이 얼굴 움직임 해독법을 통해 사람이 어떻게 표정으로 감정을 드러내는지 설명해주는 책이다. 얼굴 근육의 조합임을 과학적으로 밝혀낸 에크먼은 "표정은 언어와 달리 민족이나 문화권에 관계 없이 보편적이며 조작된 표정으로 감정을 숨길 수는 없다"고 설명한다. 예를 들어 진정으로 기분이 좋아서 웃을 땐 광대뼈로부터 입가 쪽으로 이어진 큰 광대근과 눈둘레근이 함께 수축한다. 그러나 억지로 웃을 땐 큰 광대근만 수축된다. 큰 광대근은 의지에 복종하지만 눈둘레근은 그렇지 않다는 것이다. 따라서 얼굴 움직임 해독법을 알면 거짓말 탐지기보다 정확하게 마음을 읽어낼 수 있다고 한다.

교육을 통해 표정 해독 전문가가 되면 15분의 1 초밖에 되지 않는 '미세표정'을 읽어 거짓 표정을 잡아낼 수 있다는 얘기가. 에크먼은 이 책에서 그동안 실시한 수많은 실험을 통해 슬픔과 괴로움, 분노, 놀라움, 두려움, 억겨움, 업신여김, 기분 좋음 등 여덟 가지 감정이 일어나는 과정과 그에 따른 얼굴 표정을 사진과 함께 자세히 설명하고 있다.

3.2. 음성 분석 방법

본 논문에서의 음성분석 기법을 통한 사례 분석 실험은 2가지로 수행하였다. 첫 번째로 통합민주당 박세승 공천심사위원장의 음성과 두 번째로 신은경 후보, 나경원 후보의 음성을 분석하였다.

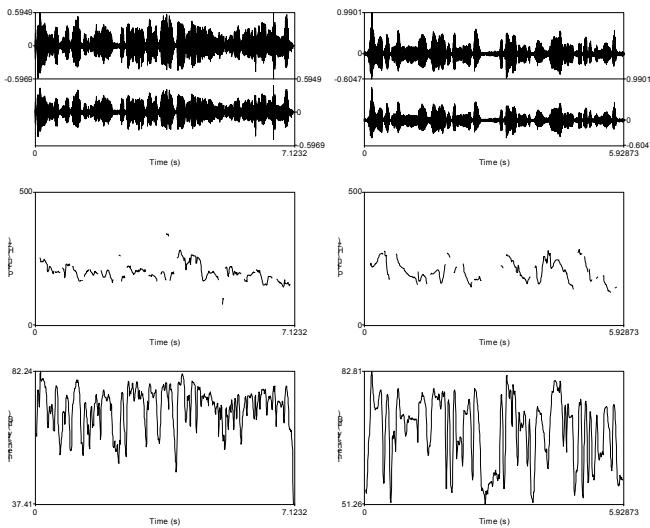
<표 1> 통합민주당 공천심사위원장 임명받은 날

<p>음성과형 : 간결하면서 규칙적인 파형을 나타냄으로써 굳은 의지가 엿보인다.</p>	
<p>피치 : 피치 파형이 전체적으로 높은 편으로 활기찬 음성임을 분석 결과로 알 수 있다.</p>	
<p>강도 : 말소리의 세기에서 최대/최소의 범위가 작게 형성되어 있으며 전체적으로 고루 분포되어 있어 극도로 감정의 자제하는 형태이다.</p>	

<표 2> 통합민주당 공천 1차 발표 직후

<p>음성과형 : 파형이 물려있으며 높낮이가 불규칙적으로 형성되어 있는 것으로 격한 감정을 드러내고 있다.</p>	
<p>피치 : 피치 파형이 전체적으로 낮은 편으로 슬픈 음성임을 분석 결과로 알 수 있다.</p>	
<p>강도 : 말소리의 세기에서 최대/최소의 범위가 크게 형성되어 있으며 전체적으로 불규칙적인 분포를 나타내어 격한 감정을 나타내고 있다.</p>	

위의 <표 1>과 <표 2>에서 알 수 있듯이 통합 민주당 박재승 공천심사위원장의 경우 처음 공천심사위원장을 임명받는 날의 소감 인터뷰에서는 자신의 목소리를 상당히 간결하게 표현하면서 감정의 노출을 극도로 자제하여 심사위원장으로서의 꾀꾀한 의지를 표출하는 듯한 음성 분석 결과를 보인 반면 1차 공천 발표 후 인터뷰에서는 반발 세력에 다소 흔들리는 듯한 격한 감정을 표현하고 있는 음성 분석 결과를 나타내고 있다.



(그림 3) 신은경(左)과 나경원(右) 후보의 음성 분석

<표 3> 신은경과 나경원 후보의 음성 내용

신은경	“주민들과 부대껴 왔던 그곳에 어떤 문제점이 있고 어떤 미래를 원하는지 그분들의 바람을 잘 알고 있습니다.”
	분석결과 : (43자/7.12초=6.039자/s), 다소 화가 나 흥분한 상태
나경원	“당을 같이 한 분인데 매우 안타깝습니다. 유권자들로부터 능력으로 평가받겠습니다.”
	분석결과 : (34자/5.93초=5.734자/s), 침착하게 감정을 조절하고 있는 상태

위의 (그림 3)과 <표 3>에서 알 수 있듯이 신은경(左) 후보는 음성 파형이 크고 분석 결과도 매우 불규칙하며 감정이 격한 부분에서 파형이 크게 나타나는 것으로 다소 화가 났음을 알 수 있고, 나경원(右)후보는 음성 파형이 작고 고루 분포되어 있으며 분석 결과도 일정한 수준을 유지하고 있는 것으로 감정을 최대한 조절하고 있는 것을 알 수 있다.

4. 결론

얼굴에 나타나는 여러 표정을 분석해보면 심리 상태 및 성격을 알 수 있다. 이는 우리의 감정이나 성격이 얼굴 표정 변화에 의해 나타나기 때문이다. 얼굴 표정 중 가장 크고 자세한 감정을 나타내는 부위는 눈과 눈썹 입으로 이 부분에 대한 분석을 통해 큰 감정을 읽어 낼 수 있으며 전체적인 얼굴 표정을 통해 여러 가지 분석을 행할 수 있는 것이다.

감정인식 기술은 근육의 긴장 상태, 전기저항, 피부 표면 온도, 혈액 순환 등의 생체 신호를 이용해서 다양한 감정 상태를 데이터베이스화한 지표와 연계해서 파악해내는 원리이다. 이러한 음성인식 기술에서 발전해 휴대폰에 센서가 부착되면서 바이오리듬을 체크하거나 현재 감정 상태를 인지해 다른 부가서비스와 연계하는 모델도 생각해 볼 수 있다. 공상과학 영화에서나 나오는 인간의 음성을 듣고 분노, 기쁨, 슬픔, 화남, 놀람 등의 인식하는 장면이 실제로 가능한 시대가 올 것으로 전망된다.

참고문헌

- [1] Paul Ekman, 얼굴의 심리학, pp. 6, Oct. 1960.
- [2] G. yang and TS. Huang, "Human face detection in a complex background", pattern recognition, Vol 27, No 1, pp 53-63, 1994.
- [3] Lantis, C. J. Taylor, and T. F. Cootes, "An automatic face identification system using flexible appearance models", Image and Vision Computing, Vol 13, No 4, pp. 393-400, 1995.
- [4] T. Leung, M. Burl and P. Perona, "Finding faces in clustered scenes using labelled random graph Vision, 637-644, MIT, boston, 1995.
- [5] P.Ekman and W.V. Friesen. Emotion in the human face System. Cambridge University Press. Sanfrancisco, CA. second edition, 1982.
- [6] <http://blog.naver.com/ararikim/70005305394>
- [7] V.Kostov and S.Fukuda, "Emotin in User Interface, Voice Interaction System." IEEE Intl Conf. on Systems, Man, Cybernetics Representation, no.2, pp. 798-803, 2000.
- [8] T. Moriyama and S. Oazwa, "Emotion Recognition and Synthesis System on Speech." IEEE Intl. Conference on Myltimedea Computin and Systems, pages 840-844, 1999.