

인간-기계 인터페이스를 위한 감성인식 기술

이 연 주 한국과학기술연구원 바이오닉스연구단 연구원 | e-mail : younjoo@kist.re.kr
 윤 인 찬 한국과학기술연구원 바이오닉스연구단 책임연구원 | e-mail : iyoun@kist.re.kr

이 글에서는 인간-기계 인터페이스(man-machine interface)의 가장 중요한 요소 중 하나인 감성인식 기술(emotion recognition technology)에 대해 소개하고 그 연구 동향에 대해 소개하고자 한다.

‘포스트 스마트폰(Post-Smartphone)’ 시대가 열리면서 사물인터넷(IoT: Internet of Things), 웨어러블(wearable), 스마트홈(smart home), 스마트 자동차(smart car) 등과 관련된 다양한 스마트 기기(smart device)들이 개개인의 일상에 더욱 깊숙이 들어오게 되었다(그림 1). 이 때문에 인간-기계 인터페이스 기술은 사용자의 편의(user-friendly)를 향상시키는 것에서 나아가 사용자를 이해하는(user-comprehensive) 수준으로 발전하고 있으며, 이를 실현하기 위해 감성인식 기술이 반드시 필요하게 되었다.

최근 스마트 모바일 기술 및 웨어러블 기술의 핵심 기술로 급부상하고 있는 감성 ICT(Information & Communication Technology) 기술은 일상생활에서 인간의 감성을 자동으로 인지하고 사용자 환경에 맞게 감성정보를 처리하여 감성맞춤형 서비스를 제공하는 기술이다. 이 기술은 감성신호를 센싱하는 기술과 감성인지 기술, 감성서비스 기술로 나눌 수 있다. 여기서 감성신호센싱 기술은 인간의 감성변화에 의한 자율신경계의 활동에 의해 나타나는 생체신호 및 환경/상황신호, 영상신호, 음성신호 등을 일상생활의 무구속/무자각 상태에서 센싱할 수 있는 초소형/초정밀 센서 소자 기술을 말하며, 감성인지 기술은 센서에서 취득된 신호를 처리/분석하고 이를 기반으로 인간의 감성을 인식, 검증, 규격화하여 정보화 하는 기술이다. 마지막으로 감



그림 1 포스트 스마트폰 시대의 디지털 디바이스

성서비스 기술은 사용자의 상황에 맞게 정보를 처리하고 감성맞춤형 제품이나 서비스를 제공하는 기술이다.

사람의 감정은 다양한 형태로 표출되어 나타난다(그림 2). 가장 대표적인 두 가지는 얼굴표정과 음성이다. 사람의 얼굴표정은 기쁨, 슬픔, 화남 등의 표면적인 감정상태 뿐만 아니라 긍정, 부정 등의 의사표현과 피로 등의 신체적 상황도 표시할 수 있다. 얼굴의 표정 변화는 안면의 각 부위의 근육을 제어함으로써 이루어지는데, 주로 눈과 눈썹, 입 주위의 근육을 움직여 서로 다른 감정을 표현하게 된다. 음성은 문자로 표현되는 정보를 소리로 변환한 것을 의미하는데, 사전적 의미뿐만 아니라 감정도 내포하고 있다. 의성어의 사용과 발성의 속도, 각 발음 사이의 묵음구간의 길이 등이 화자의 감정

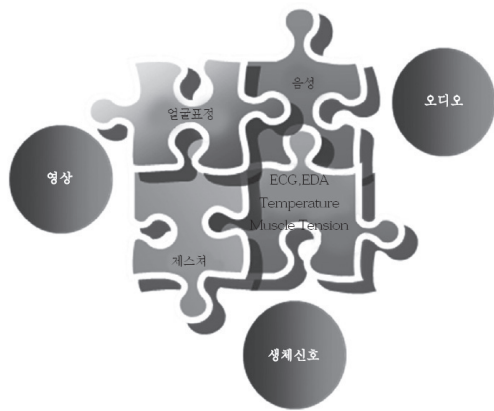


그림 2 인간의 감정표현 수단과 신호의 형태(김남수, Telecommunications Review 2009)

을 인지하는 데 매우 유용한 단서로 사용된다. 얼굴표정과 음성 외에도 손짓, 머리의 움직임, 안구의 움직임 등의 제스처(gesture) 또한 감정을 표출하기 위한 중요한 수단이 된다. 얼굴표정과 제스처는 카메라를 통해 영상신호의 형태로 취득되고, 음성은 마이크를 통해 오디오 신호의 형태로 취득된다. 이 외에도 여러 가지 생체신호의 형태로 감정이 표출되기도 한다. 예를 들어 심전도, 뇌전도, 피부전기전도도, 근육의 강도 등은 사람의 감정이 변할 때 각 신호의 특성이 달라져 사람의 감정 상태를 판단하는데 사용된다. 이 글에서는 감성인식에 사용되는 여러 가지 신호들을 기반으로 현재까지 진행되고 있는 연구들의 신호처리방법, 인식방법, 한계점 등에 대하여 소개하고자 한다.

영상 기반 감성인식기술

영상을 이용한 감성인식은 표면적으로 드러나는 감성신호들 중 얼굴표정이나 몸짓 등과 같이 카메라를 통해 획득될 수 있는 감성정보를 이용하여 사람의 감성을 인식하는 것이다. 영상을 이용한 감성인식에 관한 연구들은 대부분 얼굴표정을 이용하고 있으며, 손짓, 머리의 움직임 등과 같은 제스처가 차지하는 비중은 비교적 적은 편이다.

일반적으로 얼굴표정에서 구분할 수 있는 사람의 감성은 기쁨, 슬픔, 놀람, 화남, 혐오, 공포의 6가지 감성으로 알려져 있으며, 감성을 인식하기 위한 과정은 신호의 형태에 상관없이 전처리-특징추출-인식의 세 단계로 구성된다. 영상을 이용하는 경우에는 관찰하고자 하는 관심영역을 영상 내에서 분리하는 일을 주로 전처리 단계에서 수행한다. 예를 들면, 얼굴표정을 이용할 때는 입력된 영상에서 얼굴을 검출하고, 얼굴영역만 분리하는 작업을 전처리에서 수행하게 된다. 감성인식에서 핵심이 되는 부분은 특징추출과 인식 단계이다. 특징추출 단계에서는 감성변화에 따른 신호의 변화를 가장 잘 대표할 수 있는 특징을 추출하는 것이고, 인식 단계에서는 추출된 특징을 이용하여 감성을 정확하게 분류할 수 있는 적합한 분류기(classifier)를 사용하는 것이기 때문이다.

얼굴표정을 이용한 감성인식에는 추출하고자 하는 특징을 기준으로 기하학적 특징 기반 방법(geometric feature based method)와 외관 특징 기반 방법(appearance feature based method) 두 가지가 있다. 기하학적 특징 방법은 얼굴의 구성요소인 눈, 코, 입, 입술, 눈썹 등의 위치와 이들 사이의 관계를 특징 값으로 변환하여(그림 3(a)) 서로 다른 감성에 따른 각 구성요소들의 형태와 위치변화를 분석하여 해당되는 감성을 판단하는 방식이다. 이러한 방법은 얼굴 특징을 소수의 인자로 간단하게 표현이 가능하고 위치, 크기, 회전 변화에 쉽게 대응이 가능하지만, 얼굴의 구성요소를 정확하게 검출하기 어려운 문제가 있다.

외형 특징 기반 방법(그림 3(b))은 영상의 화소 값을 직접 이용하는 방식으로 분석 영역에 따라 전역특징기반 방법(holistic feature based approach)과 지역특징기반 방법(local feature based approach)으로 다시 구분된다. 전역특징기반 방법으로는 PCA(Principal Component Analysis), LDA(Linear Discriminant Analysis), ICA(Independent Component Analysis) 등이 연구되고 있으며, 지역특징기반 방법으로는 얼굴영

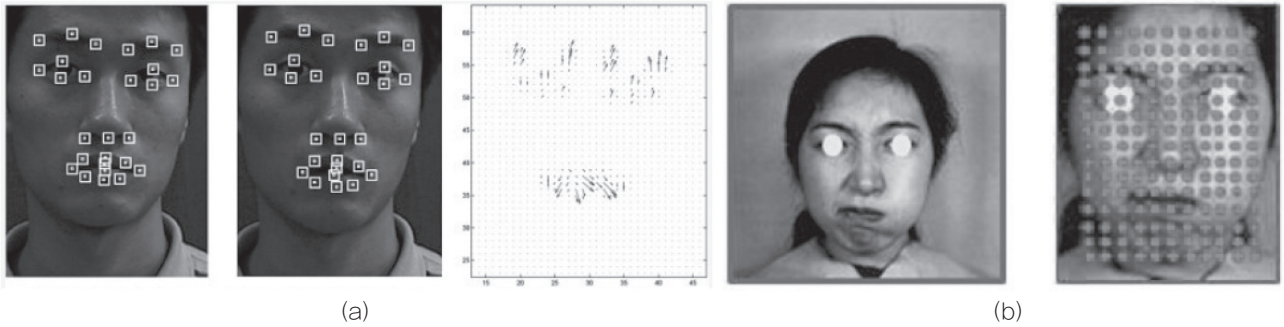


그림 3 얼굴표정에 대한 특징추출 방법: (a) 기하학적 특징추출의 예; (b) 외관 특징추출의 예(이상무, KEIT PD Issue Report 2011)

상의 세부적인 영역을 다루는 Local PCA, 가버 웨이블릿(Gabor wavelet)(그림 3(b)), LBP(Local Binary Pattern) 등이 연구되고 있다. 외형 특징 기반의 방법은 기하학적 특징 기반의 방법과 다르게 영상 자체에서 특징을 추출하기 때문에 얼굴의 구성요소를 찾아야 하는 번거로움이 없고, 기하학적 정보보다 더 풍부한 정보들을 포함하고 있어 최근 외형 특징 기반의 방법들에 관한 연구가 활발히 진행되고 있는 추세이다. 또한 과거의 정지영상(still image)에서 진행되던 연구들이 보다 정확한 감성인식을 위해 비디오(video)에서 분석하고자 하는 움직임이 활발히 일어나고 있다. 감성을 인식하는 분류기로는 Gaussian Mixture Model(GMM), Hidden Markov Model(HMM), Support Vector Machine(SVM)이 사용된다.

음성 기반 감성인식기술

음성은 인간-기계 인터페이스 중에서 가장 효율적이고 자연스러운 방법으로 여기에 내포된 감정을 추출하려는 연구가 활발히 진행되고 있다. 음성을 이용한 감성인식은 현재까지 수많은 연구결과를 얻어온 음성인식에서 그 시발점을 찾을 수 있으나, 특징추출 및 패턴인식 알고리즘의 선택에 있어서 크게 차이가 있다. 특

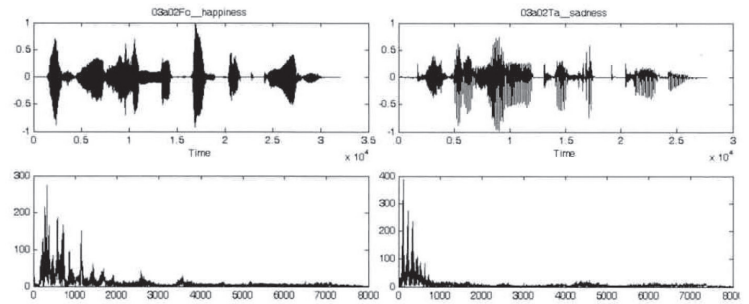


그림 4 기쁨(happiness)과 슬픔(sadness)에 대한 음성신호(정병욱 외, 한 국지능시스템학회 논문지, 2008)

징벡터 선택에 있어서 음성인식의 경우 음소를 모델링하는 요소를 주로 이용하는 반면, 감성인식에 있어서는 운율적 요소를 활용한다. 또한 특징선택과 함께 패턴인식 알고리즘의 선택도 매우 중요한 요소로 추출한 특징을 이용하여 감성을 모델링하는 방식에 따라 패턴인식 알고리즘이 다르게 선택될 수 있다.

그림 4는 한 명의 화자가 기쁨과 슬픔의 상태에서 같은 문장을 발음하였을 때의 음성신호들을 보여준다. 첫 번째 열에 있는 두 그래프는 기쁨의 상태일 때 시간 영역과 주파수 영역에서의 신호들이고, 두 번째 열에 있는 두 그래프는 슬픔의 상태일 때를 나타낸다. 그림 4에서 보는 바와 같이 화자가 기쁨을 표현할 때는 다양한 주파수 영역에서 에너지 값을 가지고, 슬픔을 표현할 때는 특정 주파수 성분들에서만 많은 에너지를 가진다. 또한 시간 영역에서는 슬픔을 표현하는 음성신호가 기쁨을

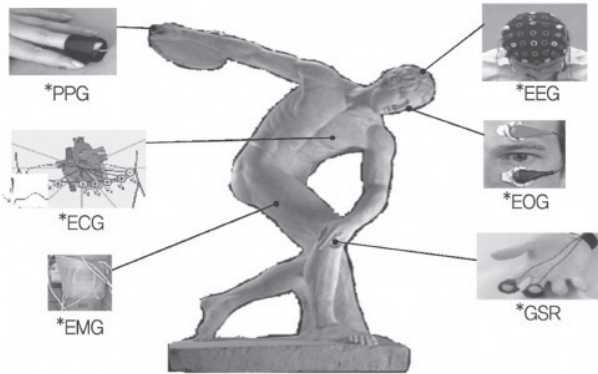


그림 5 감성인지를 위해 사용되는 생체신호(신현순 외, 한국 전자통신연구원 2014)

표현하는 신호보다 거칠게 나타나는 것을 볼 수 있다. 이와 같이 음성을 이용한 감성인식을 수행하기 위해서는 음성인식에서 일반적으로 사용되는 특징이 아닌, 감성을 구분해낼 수 있는 특징을 잘 선별해야 한다.

음성의 음소를 나타낼 때 사용되는 파라미터로는 MFCC(Mel-Frequency Cepstral Coefficient)가 대표적인 특징이고, 감성인식에 사용되는 운율적 요소로는 피치(pitch), 에너지, 발음속도 등이 있다. 이러한 음성 파라미터는 음성신호의 정의된 구간에 대하여 계산된 피치와 에너지 값으로부터 피치 평균, 피치 표준편차, 피치 최대값, 에너지 평균, 에너지 표준편차 등의 통계적 정보를 산출하여 감성인식에 사용한다.

인식 단계에서 활용되는 식별 방법으로는 음성인식 및 화자인식 등에서 사용되는 GMM, HMM, SVM, Artificial Neural Network(ANN) 등이 적용되고 있으며, 이 중 GMM 기반의 식별 방법이 피치나 MFCC와 같은 단구간 특징 파라미터에 적합하다는 연구결과가 있다.

생체 기반 감성인식기술

일반적으로 인간이 느끼는 감성은 외부자극에 의해 나타나며 뇌를 통해 중추신경계 또는 자율신경계의 반응으로 이어져 생체신호로 측정 및 관별이 가능하다. 인간의 감성과 매우 밀접한 관계를 갖고 있는 생체신호

로는 뇌파(EEG: Electroencephalogram), 심전도(ECG: Electrocardiography), 피부전도도(GSR: Galvanic Skin Response), 피부온도(SKT: Skin Temperature) 등이 있다.(그림 5)

뇌파(EEG)는 생체 기반 감성인식에 가장 널리 활용되는 생체신호 중 하나이다. 뇌파는 다른 생체신호들과 달리 중추신경계의 명령을 직접 받으며, 알파파와 베타파가 감성과 연관이 있다고 알려져 있어 감성인식에 주로 사용된다. 알파파는 안정, 긍정감정일 때 나타나고 베타파는 부정감정일 때 나타난다는 특성이 있다. 하지만, 뇌파를 이용한 감성인식 기술을 실용화하기 위해서는 다양하고 복잡한 신호의 잡음을 제거해야 하는 큰 벽을 넘어야 한다.

심전도에서 얻어지는 심박동변이율(HRV: Heart Rate Variability)와 심박수(HR: Heart Rate)를 이용한 연구들도 있다. 심전도에서 HRV를 산출하고 여러 가지 특징 파라미터를 추출해 감성을 분류를 하고 있지만, 정확하게 감성을 인식하는 것이 어렵다는 단점이 있다.

피부전도도는 기본적으로 피부에 땀이 발생할 때 증가되는 피부의 전도성을 의미하는 것으로 외부 자극이나 스트레스에 대한 유용한 지표로 사용되고 있다. 피부전도도의 진폭(amplitude)은 각성 또는 부정감정일 때 커지고, 반응의 속도(latency)와 반응의 민감도(slope)도 민감한 자극일 때 빠르고 크게 나타난다. 따라서 피부전도도는 주로 부정적인 감성을 측정하는 좋은 지표로 활용될 수 있다.

피부온도는 신체의 특정 부위들에 대한 온도 지표로써 체온과는 구별된다. 일반적으로 피부온도신호의 진폭이 커지면 긍정감정일 수가 있고, 작아지면 긴장 또는 불쾌한 부정감정일 수 있다. 하지만 피부온도는 외부환경에 의해 크게 영향을 받기 때문에 외부적 요소도 함께 고려해야 보다 정확한 결과를 기대할 수 있다.

근전도(EMG)를 이용한 감성인식은 주로 얼굴 근육의 움직임에 이용하는데, 얼굴표정 근육의 움직임을 측정하고 이를 이용하여 스트레스가 높은 긴장도 등을 측

정하여 감성인식에 활용한다. 즉, 긴장/각성 상태에서 EMG는 증가하고, 이완상태에서는 감소하는 특성을 이용하는 것이다. 하지만, 근육 긴장의 절대 수치는 측정되는 근육 부위와 사람에 따라 다르기 때문에 세심한 측정이 요구된다.

앞서 살펴본 바와 같이 감성인식에 사용할 수 있는 생체신호는 다양하게 존재하나 주위의 환경조건이나 개인에 따른 심리 기저상태에 따라 감성의 변화가 달리 측정될 수 있기 때문에 단일 생체신호만으로 감성을 정확히 인지하는 것이 매우 어렵다. 따라서 단일 생체신호만을 이용한 감성인식을 지양하고 여러 생체신호를 복합적으로 활용하거나 추가적으로 얼굴표정 및 음성 등의 정보들도 함께 활용하는 방법이 시도되고 있다.

이 글에서는 인간-기계 인터페이스에서 가장 중요한 기술 중 하나인 감성인식 기술에 사용되는 여러 가지

방법들에 관하여 알아보았다. 최근 들어 영상, 음성, 생체, 뇌파 및 신체의 정보를 감지하고 감성을 추출하여 감성서비스로 제공하고자 하는 시도가 다방면에서 활발히 이루어지고 있다. 게다가 감성서비스 기술은 엔터테인먼트, 헬스케어, 시장조사, 온라인교육, 자동차, 고객 마케팅 및 일반 가정 등 다양한 분야에 적용될 수 있는 잠재력을 가지고 크게 성장하고 있다. 이와 같은 성장에 발맞춰 감성인식 기술을 실용화하기 위해서는 우선 자연스럽게 유도된 많은 양의 감성 데이터베이스가 구축되어야 하고, 여러 가지 주변 환경의 잡음의 영향에 강인하도록 멀티모달(multimodal) 방식의 처리기법이 적용되어야 한다. 이러한 다양한 노력들을 통해 감성인식 기술은 향후 ICT 기술과 점진적으로 융합 정착되어 진정한 꽃으로 피어날 것이라 예상해본다.



기계용어해설

열증기압축기(Thermal Vapor Compressor)

고압증기를 이용하여 저압증기를 중간압으로 증압/이송시키는 일종의 이젝터 장치

유전자 알고리즘(Genetic Algorithm)

실세계의 문제를 풀기 위해 잠재적인 해들을 컴퓨터 상에서 코딩된 개체로 나타내고, 여러 개의 개체들을 모아 개체군을 형성한 뒤, 세대를 거듭하면서 이들의 유전 정보를 서로 교환하거나 새로운 유전 정보를 부여하면서 적자생존의 법칙에 따라 모의 진화를 시킴으로써, 주어진 문제에 대한 최적의 해를 찾는 계산 모델.

유도초음파(Guided Wave)

길이에 비해 두께가 얇은 시편에서 발생하는 파형으로 구조물의 벽면을 따라 전파하는 파인. 신호가 복잡하여 해석하는데 어려움이 있지만 장거리 탐상에 적합한 파인.

열차폐코팅(Thermal Barrier Coating)

가스터빈 고온 부품의 온도를 낮추기 위해 적용되는 2중 코팅

영구자석(Permanent Magnet)

강한 자화상태를 오래 보존하는 자석으로 외부로부터 전기에너지를 공급받지 않아도 자성을 안정되게 유지함.

위상최적화(Topology Optimization)

설계 영역을 이산화하여 각 부분의 존재 유무를 판정함으로써 최적의 구조를 추출해 내는 설계 방법

인공 추간판(Artificial Disc)

척추의 퇴행성 질환의 치료에 사용되는 척추 유합술의 문제점을 보완하기 위한 체내 인공삽입물으로써 척추의 운동성을 유지할 수 있는 장점을 가짐.