음성신호 분석 기반의 영화 장르별 감정변화 특성 연구

유황준*, 한상효*, 김봉현*, 가민경*, 조동욱*
*충북도립대학교 전자정보계열
kyoyhj@naver.com

A Study of Emotional Variation Tendency by Movie Genre Based on Speech Signal Analysis

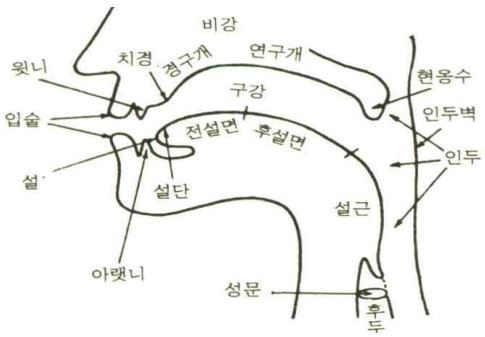
Hwang-Jun Yoo*, Sang-Hyo Han*, Bong-Hyun Kim*, Min-Kyoung Ka*, Dong-Uk Cho*
*School of Electronics and Information, Chungbuk Provincial University

요 약

인간의 능력 중 가장 주목할 만한 것은 언어를 습득하고 그것을 이용하여 서로 의사소통을 할 수 있다는 것이다. 모든 언어에는 그 언어만이 가지는 특수성뿐만 아니라 공통적으로 존재하는 보편적인 특성이 있다. 이것 외에도 언어 위에 입혀지는 사람의 목소리는 의사소통을 하는데 있어 상대의 심리를 파악하는 중요한 단서가 된다. 특히, 언어는 습득되어져야 활용되고 그 습득되어지는 환경에 영향을 받으며 이러한 환경에 따라 사람의 목소리, 억양 등이 변화하게 되는 것이다. 따라서 본 논문에서는 음성신호 분석 기법을 적용하여 장르별 영화시청에 따른 시각적, 청각적 요인이 목소리에 미치는 영향을 분석하는 연구를 수행하였다. 이를 위해 장르별 영화를 시청한 후 성대 진동 및 음성에너지의 크기 변화를 측정하여 감정변화를 분석하는 실험을 수행하였다.

1. 서론

일상에서 우리가 살아가면서 감정의 변화에 따라 얼굴 표정이 달라지듯 목소리도 감정에 따라 표현이 달라진다. 용의자를 심문하는 검사의 목소리로는 날카롭고 확신에 찬 강한 목소리가 적합하고, 판사의 냉정하고 위엄 있는 목소리는 다소 긴장감과 두려움을 느끼게 만든다. 반면 텔레마케터의 높고 상냥하며 애교 섞인 목소리는 가볍고 즐거운 느낌을 선사한다. 이처럼 사람의 목소리는 기분이나 상황, 혹은 직업에 따라 다양한 느낌을 주며 듣는 사람들로 하여금 다양한 분위기나 감정의 변화를 유발할 수 있다. 특히 고객과 대면하지 못하고 전화로만 대화를 해야 하는 텔레마케터에게 있어 목소리는 유일한 접촉수단이다. 때문에 언제나 밝고 높은 톤의 목소리와 경쾌하고 짧은 목소리로 상대방에게 호감을 주어야 한다. 음성이란 인간의 음성 기관을 통하여 나오는 소리이다. 발음이 되는 음성 기관은 호흡기, 후두, 성문상관이다. 호흡기에서는 음을 생산하는 데 필요한 공기를 공급하며, 후두에서는 말하는 데 사용되는 음의 에너지를 만든다. 그리고 성문상관에서는 말에 사용되는 대부분의 소음이 만들어진다. 호흡에는 들숨과 날숨이 있는데, 횡격막이 올라가고 늑골이 내려감으로써 그 결과 허파에 담겨 있던 공기의 대부분은 배출되는 날숨에 실려 밖으로 나오게 된다. 이 날숨에 의해 공기가 밖으로 배출될 때 발성이 된다[1]. 일반적인 음성 연기는 숨을 내쉬고 난 후에도 감정에 따른 음이 발해지기도 한다. 아래의 [그림 1]은 인체 발성기관을 나타낸 단면도이다.



[그림 1] 발성기관 단면도

2. 목소리와 감정

목소리란 인간이라면 누구든지 가지고 있는 것 말 그대로 목에서 나는 소리이다. 감정이란 전에는 심리학에서 감각과 감정을 구별하지 않았으나, J.워드와 W.분트는 감각은 객관적이며, 감정은 주관적인

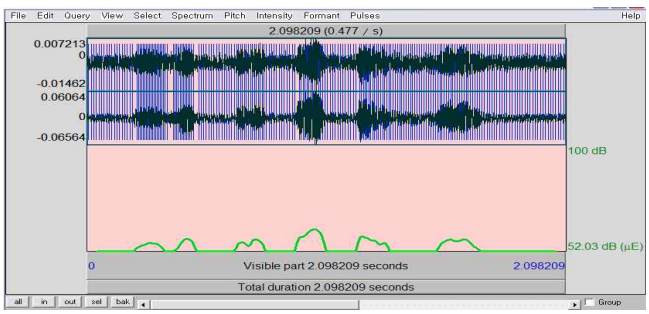
것이라 구별하였다. 감정은 인식작용이나 충동의지와 다른 것이지만 엄밀히 구분할 수는 없다. 감정과 의지가 하나가 된 정의(情意)를 독일어에서는 'Gemt(心情)'라 하고, 감정과 지각(知覺)이 합쳐진 상모적 지각(相貌的知覺)이라는 현상도 있다. 따라서 최근에는 엄밀한 의미에서 지(知)·정(情)·의(意)로 의식(意識)을 구분하는 견해는 부정되고 있다. 감정이 생기는 원인으로는 생리적 원인, 심리적인 원인, 사회적인 원인, 문화적인원인 네 가지로 나눌 수 있는데, 먼저 생리적 원인으로는 몸을 의지할 곳이 없어지면 공포심을 느끼거나, 몸의 어떤 부분을 자극하면 쾌감이 생기고, 몸을 세게 치면 고통의 감정을 느끼는 등에 원인을 말한다. 다음으로 심리적 원인은 어떠한 행동에 요구수준이 도달하면 성공감을 느끼고, 도달하지 못하면 실패감을 느끼는 등에 쾌·불쾌, 행복감·불행감이 주된 감정을 말한다. 사회적 원인으로는 승리와 패배의 감정, 당해 낼 수 없는 상대를 대할 때의 열등감과 이와 반대 경우의 우월감이 존재하는 것을 말한다[2].

3. 억양

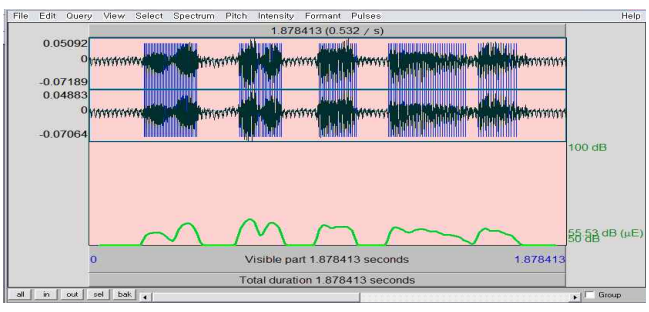
억양은 크게 두 부분으로 나뉜다. 하나는 위의 '틀렸어' 안에서 나타나는 소리의 높낮이고, 다른 하나는 그 말 끝에 나타나는 소리의 상승(上昇)과 하강(下降) 부분이다. 앞의 것을 흔히 고저조(高低調, pitch level)라 하고, 문장 끝의 상승과 하강부분을 문말조(文末調, terminal contour)라 부른다. 문말조는 구두점(句讀點)에 의해서 얼마간 짐작이 되지만 그것도 같은 구두점이 다른 문말조로 나타나는 일도 있어 충분하지 못하며, 고저조는 아예 글을 쓸 때는 표시되지 않으므로 억양은 전체적으로 특수한 경우가 아니면 어떤 시각적인 부호로 표시되는 일이 드물다. 억양은 또 문장에서 나타나므로 단어가 단독으로 있을 때의 모습을 보여주는 사전(辭典)에서도 표시할 수 없다. 이 때문에 억양은 문장을 이루는 어떠한 다른 요소보다도 소홀히 다루어지는 경향이 있다. 고저조란 문장 안에 나타나는 일정한 체계가 있어 이를 잘못 지키면 말의 뜻이 달라지거나 적어도 부자연스럽게 들린다[3]. 국어 억양의 한 규칙으로 '수식어+피수식어'의 구성에서 수식어에 고조가 놓인다는 것을 들 수 있다. 국어의 억양 중 의미의 분화에 관여하는 가장 전형적인 예는 '누구·언제·어디·무엇' 등으로 시작되는 의문문에서 볼 수 있다. 억양은 이상에서처럼 문장의 의미결정에 관여하거나 자연스러운 문장을 만드는 일에 기여하는 변별적이고 상대적인 소리이다. 화가 났을 때라든가 비명을 지를 때 소리를 더 높여 말하는 따위는 억양의 범주에 들지 않는다. 소리를 크게 지른다고 해서 단어들 간의 상대적인 높이는 달라지지 않기 때문이다[4].

4. 실험 결과 및 고찰

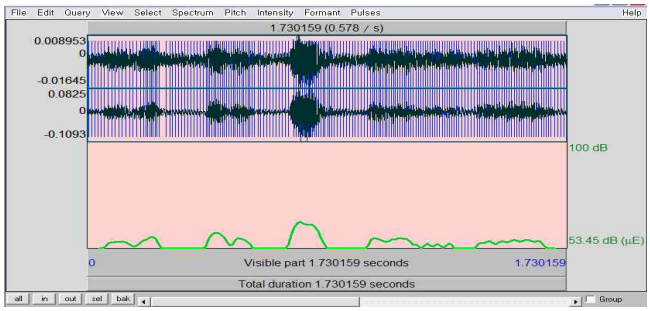
본 논문에서는 발음이 정확한 20대 남성으로 피실험자 집단을 구성하였으며 실험의 변칙이 발생하지 하도록 약물 복용 및 각 신체의 질병으로 인해 과거 이력이 없는 피실험자들로 형성하여 실험을 진행하였다. 피실험자 수는 10명이며 실험 환경으로 잡음이 섞이지 않도록 10평 가량의 밀폐된 공간에서 SONY사의 ICD-SX750 레코드를 이용하여 녹취를 실행하였다. 목소리와 감정의 연관성을 알아보기 위한 실험 전 각 개체군들이 안정을 취할 수 있도록 5분간 휴식기를 갖고 "보다 아름다운 우리의 세상" 문구를 사용해 녹음을 진행하였다. 실험은 기본적인 목소리를 녹음한 후 각각 20분씩 즐거운 감정을 느낄 수 있는 영화와 슬픈 감정을 느낄 수 있는 영화를 각각 시청하고 음성 자료를 수집하였다. 수집된 실험 전과 후의 음성을 분석하는 방법으로는 IBM-PC상에서 공학 음성 분석 프로그램인 Praat를 사용하여 [그림 2]에서 [그림 7]과 같은 방법으로 실험 전과 후의 음성자료 데이터를 추출하여 음성데이터 값을 분석하였다.



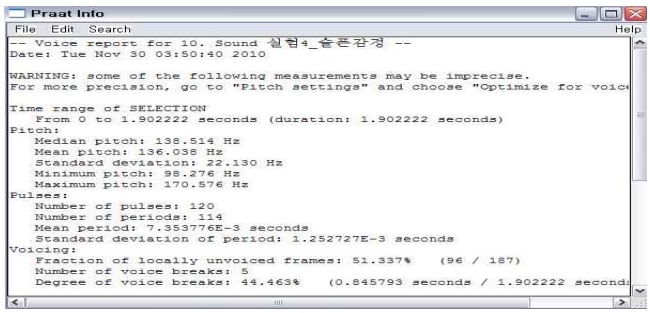
[그림 2] Praat를 이용한 일반목소리 음성파형



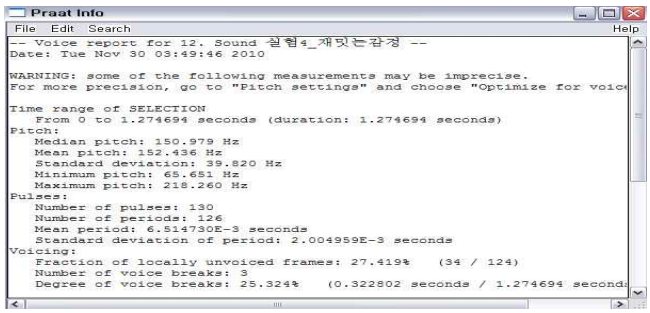
[그림 3] Praat를 이용한 즐거운 목소리 음성파형



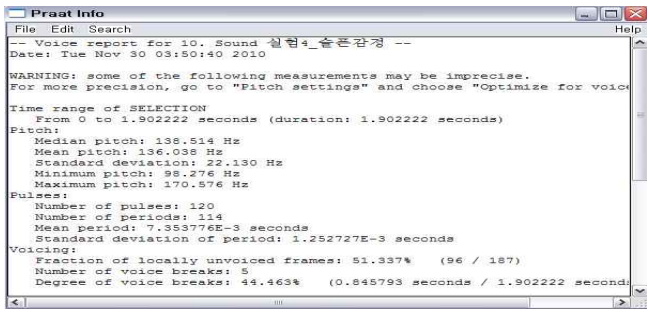
[그림 4] Praat를 이용한 슬픈 목소리 음성파형



[그림 5] 일반목소리 Intensity 데이터 값 추출과정



[그림 6] 즐거운 감정 Intensity 데이터 값 추출과정



[그림 7] 슬픈 감정 Intensity 데이터 값 추출과정

본 논문에서는 감정에 따른 음성의 변화를 음성공학 요소 Pitch값과 Intensity 값으로 선정하여 분석하였고, 목소리와 감정과의 연관성을 도출하여 실험을 행하였다. [표 1]은 피실험자들에 대한 일반적인 목소리와 감정변화 실험 전과 후의 Pitch값을 추출한 결과이며 [표 2]는 일반적인 목소리와 감정변화 실험 전과 후의 Intensity값을 추출한 결과이다.

[표 1] 실험 전, 후 Pitch 값

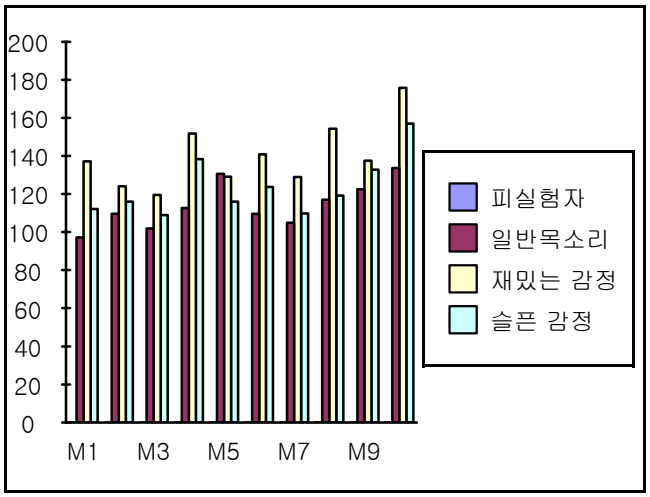
피실험자		일반목소리	재밌는 감정	슬픈 감정
M01	Pitch	97.263 Hz	137.222 Hz	112.164 Hz
M02	Pitch	109.582 Hz	124.087 Hz	115.953 Hz
M03	Pitch	101.911 Hz	119.600 Hz	108.857 Hz
M04	Pitch	112.580 Hz	151.742 Hz	138.404 Hz
M05	Pitch	130.554 Hz	129.169 Hz	115.936 Hz
M06	Pitch	109.687 Hz	140.784 Hz	123.757 Hz
M07	Pitch	104.922 Hz	128.913 Hz	109.779 Hz
M08	Pitch	117.075 Hz	154.262Hz	119.198 Hz
M09	Pitch	122.607 Hz	137.477 Hz	132.814 Hz
M10	Pitch	133.693 Hz	175.836 Hz	156.938 Hz

[표 2] 실험 전, 후 Intensity 값

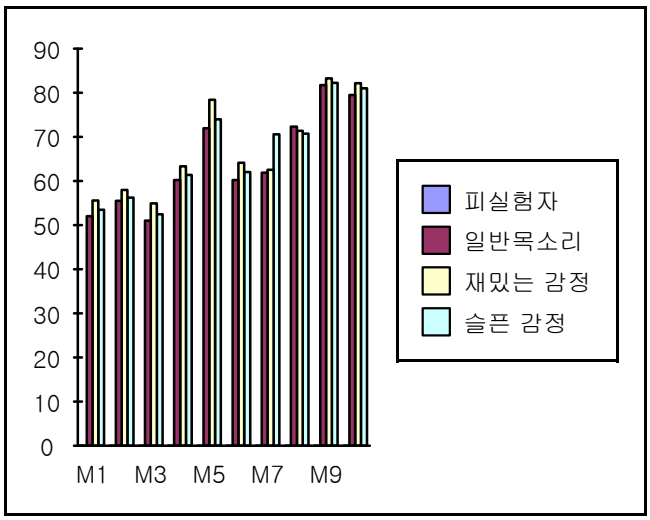
피실험자		일반목소리	재밌는 감정	슬픈 감정
M01	Intensity	52.03dB	55.53dB	53.45dB
M02	Intensity	55.51dB	57.96dB	56.21dB
M03	Intensity	50.96dB	54.88dB	52.43dB
M04	Intensity	60.17dB	63.33dB	61.35dB
M05	Intensity	71.95dB	78.38dB	73.94dB
M06	Intensity	60.22dB	64.12dB	62.02dB
M07	Intensity	61.86dB	62.53dB	70.56dB
M08	Intensity	72.31dB	71.37dB	70.73dB
M09	Intensity	81.68dB	83.21dB	82.21dB
M10	Intensity	79.45 dB	82.13dB	81.01dB

실험 결과 소리의 강도를 나타내는 Pitch값에서는 실험 전과 후의 데이터 값이 모두 증가하는 결과를 보여주었으며, 음의 높이를 나타내는 Intensity 값에 대한 데이터 값은 마찬가지로 모든 실험자가 증가하는 것을 보였으나, M07이 나머지 실험자들과는 다르게 '재밌는 감정' 보다 '슬픈 감정' 일 때 더 높은 데이터 값을 나타내었다. 이를 제외한 모든 실험자

가 같은 데이터 증가율을 보였다. 이를 통해서 목소리와 감정의 연관성이 있음을 확인 할 수 있었다. 아래 [그림 8]과 [그림 9]의 음성 분석 결과 그래프를 보면 실험 전, 후의 음성 분석 요소 Pitch값과 Intensity 값의 변화를 쉽게 분별할 수 있다.



[그림 8] Pitch 값 분석 결과 그래프



[그림 9] Intensity 값 분석 결과 그래프

위의 [그림 8]과 [그림 9]에서 볼 수 있듯이 실험 전과 실험 후의 Pitch 값과 Intensity 값이 모든 실험자가 증가하는 것을 볼 수 있다. 또한 이를 통해 결과값에 대한 고찰을 수행할 수 있었다.

5. 결론

현대 사회에서 우리가 생활함에 있어서 살아가는데 중요한 역할을 하는 것이 여가 활동이라고 생각한다. 여가 활동중에는 많은 것들이 있지만 우리가 살아가는데 가장 접하기 쉽고 또 많은 사람들이 할 수 있는 것들 중에 찾은 것이 영화 관람과 음악 감상이다. 본 논문에서는 우리가 가장 접하기 쉬운 것을 선택하여 인체의 변화를 보기 위해서 실험을 하고 논문을 쓰게 되었다. 실험을 떠나서 요즘 세대에 사람들은 일에 찌들어 살아있고, 삶을 즐기기 보다는 돈이라는 물질적인 것만을 위해서 자신의 건강과 자신만의 시간을 버리는 사람들이 다반사 인 것 같아 매우 안타깝게 생각한다.

본 논문에서는 음악 감상을 통해서 실험하기 전과 시청각적으로 자극을 주는 실험을 한 후의 음성 변화를 비교 분석하기 위하여 실험을 수행하였다. 정확한 실험을 위해서 지극히 평범한 실험자 10명을 대상으로 결과를 지켜봤으며, 그 결과 자극 적인 영화를 보았을 때 Pitch 값과 Intensity 값 모두가 증가하는 경향을 보였다. Intensity 값은 90%의 실험 대상자가 재미있는 감정을 느꼈을 때가 슬픈 감정을 느꼈을 때 보다 더 높은 값을 보였다. 단 한 명의 피실험자는 다른 변화율을 보였는데, 이는 다른 피실험자들과 같은 증가율을 보였지만, 이 피실험자는 슬픈 감정을 느꼈을 때가 더 높은 Intensity 값을 보였던 것이지 떨어지는 값을 보였던 것은 아니었기 때문에 다른 시각적인 분석 차원에서 피실험자 모두가 증가했다고 볼 수 있다.

본 논문의 실험 결과를 통해 결과값이 100%의 신뢰도를 나타내지 못했지만 실험 결과에서도 알 수 있듯이 평균 90.0%의 피실험자가 동일한 음성분석 결과를 나타냈다. 즉, 모든 피실험자가 똑같은 반응을 보이는 것은 아니었지만, 동일한 패턴의 증가율을 나타냈으며 추후 지금보다 더 많은 실험 대상자를 가지고 실험을 수행한다면 더욱 신뢰성 있고 정확한 결과를 가질 수 있을 것으로 생각된다.

또한 차후에 더 확실하게 시각적인 면이나 청각적인 면에서 음성신호 분석 요소의 적용을 통한 변화율을 비교, 분석하여 데이터를 평가할 수 있다면 더욱 신뢰도 높은 결과 값을 얻을 수 있을 것으로 생각된다.

참고문헌

[1] 도미니크 모이사, 감성의 지정학, 랜덤하우스, 2011.
[2] Graium music, 감정, EJB북스, 2002.
[3] 최현배, 우리말본, 正音社, 1937.
[4] 박창해, 국어의 얹침 음운에 관한 연구, 『東方學志, 1963.