퍼지 및 지능시스템학회 논문지 2001, Vol. 11, No. 6, pp. 510-515

음성으로부터 감성인식 요소 분석

Analyzing the element of emotion recognition from speech

심귀보 · 박창현

Kwee-Bo Sim and Chang-Hyun Park

중앙대학교 전자전기공학부

School of Electrical and Electronic Engineering, Chung-Ang University

요 약

일반적으로 음성신호로부터 사람의 감정을 인식할 수 있는 요소는 (1)대화의 내용에 사용한 단어, (2)톤(Tone), (3)음성신호의 피치(Pitch), (4)포만트 주파수(Formant Frequency), 그리고 (5)말의 빠르기(Speech Speed) (6)음질(Voice Quality) 등이다. 사람의 경우는 주파수 같은 분석요소 보다는 톤과 단어, 빠르기, 음질로 감정을 받아들이게 되는 것이 자연스러운 방법이므로 당연히 후자의 요소들이 감정을 분류하는데 중요한 인자로 쓰일 수 있다. 그리고, 종래는 주로 후자의 요소들을 이용하였는데, 기계로써 구현하기 위해서는 포만트 주파수를 사용할 수 있게 되는 것이 도움이 된다. 그러므로, 본 연구는 음성 신호로부터 피치와 포만트, 그리고 말의 빠르기 등을 이용하여 감성 인식시스템을 구현하는 것을 목표로 연구를 진행하고 있으며, 그 1단계 연구로서 본 논문에서는 화가 나서 내뱉는 말을 기반으로 하여 화난 감정의 독특한 특성을 찾아내었다.

Abstract

Generally, there are (1)Words for conversation (2)Tone (3)Pitch (4)Formant frequency (5)Speech speed, etc as the element for emotional recognition from speech signal. For human being, it is natural that the tone, voice quality, speed, words are easier elements rather than frequency to perceive other's feeling. Therefore, the former things are important elements for classifying feelings. And, previous methods have mainly used the former things, but using formant is good for implementing as machine. Thus, our final goal of this research is to implement an emotional recognition system based on pitch, formant, speech speed, etc from speech signal. In this paper, as first stage, we found specific features of feeling 'angry' from his words when a man got angry.

Key Words : 톤(Tone), 피치(Pitch), 포만트 주파수(Formant Frequency), 음질(Tone quality), 인두강((Pharyngeal cavity)

1. 서 론

인간은 일반적으로 시각, 청각, 촉각 등을 다양한 방법을 통하여 상호간에 정보를 교환한다. 감정/감성의 전달 또한 같은 방식으로 전달된다고 생각하는데 Chan et al[1]의 감성인식에 대한 연구 결과에 의하면, 감성의 6가지 기본 요소인 행복, 슬픔, 분노, 증오, 놀람, 두려움을 음성모델과 시각 모델로 분류하여 놓고 음성모델만으로 알아본 인식률은 75%, 시각모델만으로 수행된 인식률은 70%라는 결과를 각각 얻었다. 그리고 음성과 시각 모델을 함께 표현하여 얻은 인식률은 97%에 이르렀다고 한다. Chan의 연구에 의하면 음성을 통한 인식이 시각에 의한 인식보다 조금 더 효과적이라는 것을 알 수 있고, 시각과 청각이 함께 할 때 훨씬 더 높은 인식률을 얻을 수 있음을 알 수 있다. 물론 여러 감각이 합해질 때 더욱 더 높은 인식을 할 수 있지만 여러 감각을 통하여 전달되는 감성 정보를 처리하기 위해서는 그 만큼 많은 비용과 시간을 필요로 한다. 따라서 본 논문에서는 단일 감각 기관에 대해서 인식률이 가장 높은 청각 즉, 음성으로부터의 감성 정보를 인식하는 것을 목적으로 그 기초 연구인 감성 요소를 찾아내는 것을 목적으로 한다.

최근에 기계 지능 특히 애완용(pet) 로봇 분야에서 감성 인식에 대한 필요성이 크게 대두되고 있다. 애완용 로봇은 그 특성상 인간과의 상호작용이 필수적이다. 예를 들어, 인간이 로봇에 장난을 하면 웃거나, "장난하지 마!" 등의 인간과 유사한 감정표현을 할 때, 애완용으로써 인간과의 친밀감을 만들 수 있는 것이다. 이러한 기능을 수행하기 위해서는 상대가 누구인지 판단 할 수 있는 얼굴 인식과, 표정과 함께 감정표현의 주요 분출도구인 음성인식 그리고, 로봇 또한 역동적인 감정표현을 위한 기계적인 움직임의 조화가 필요하다. 이러한 기능을 하는 로봇은 현재 일본에서 활발하게 제시되고 있다. 휴먼로봇이란 타이틀로 붐을 일으키고 있는 이 로봇들은 인간이 손을 대거나 소리를 내는 것에 반응하는 기능을 갖추고 있다. 그러나, 이 로봇은 장난감으로써의 기능에

접수일자 : 2001년 11월 1일
완료일자 : 2001년 12월 1일
감사의 글 : 본 연구는 산업자원부의 2000년도 차세대 신기술개발사업인 『수퍼지능칩 및 응용기술개발』 과제의 제5세부과제인 「Autonomous Family Machine(AFM) 요소기술개발(N09-A08-4301-05)」의 위탁연구로 이루어졌으며, 산업자원부의 연구비지원에 감사 드립니다.

초점이 맞춰져 있을 뿐이다. 앞으로의 로봇의 발전방향은 단지 장난감으로써가 아니라 인간의 감정에 맞춰서 기분을 풀어줄 수 있는 행동을 하거나, 들떠있는 경우는 차분하게 유지시켜주는 행동으로 업무를 수행하는 데 있어 보조자로써의 역할을 수행할 수 있어야 한다. 이런 목적에서 감정/감성인식은 매우 중요하다. 그리고, 앞서 언급되었듯이 여러 인식 중에서 음성의 인식을 통한 감정/감성인식이 우선 연구되어야 한다.

필요성을 강조하기 위해서 다음과 같은 간단한 예를 들면, 간단한 예로써 화난 상태에서 "가라" 고 하는 것과 그냥 아무런 감정상태 없이 "가라" 고 하는 것에는 청자의 행동방향에 차이가 생길 수밖에 없다. 이렇듯 대인관계에서 감정을 인식하느냐 못하느냐로 상대의 진정한 의도파악 여부가 결정된다. 그러므로, 감정의 인식이 말, 문자와 같이 중요한 의사소통 수단이다. 그런 의미에서 감성 인식의 연구 또한 필연적이다. 또한, 상대방이 "잘했다."라고 말했을 때, 대화의 전후 문맥상 우리는 그것이 칭찬인지, 비꼬는 말인지 알 수도 있지만, 굳이 문맥을 파악하지 않더라도 억양으로부터 그 둘의 차이를 쉽게 알아낸다. 이것은 단순한 음성인식과 감성인식의 차이를 간단히 보여주는 예라고 할 수 있다. 즉, "잘했다" 라는 문장을 인식하는 것이 음성인식이라면, 그 말의 억양으로부터 화자(話者)의 감정까지 알아내는 것이 감성인식인 것이다. 위의 예에서는 감정이 억양만으로 파악 가능한 것으로 묘사했지만, 그림 1과 같이 사람들의 감정이 변할 때는 신체의 물리적이고 생리적인 여러 가지 변화가 바로 음성신호의 변화로 연결되어 주파수에 변화를 주게되므로 이러한 점에 초점을 맞추어 연구를 진행하였다.

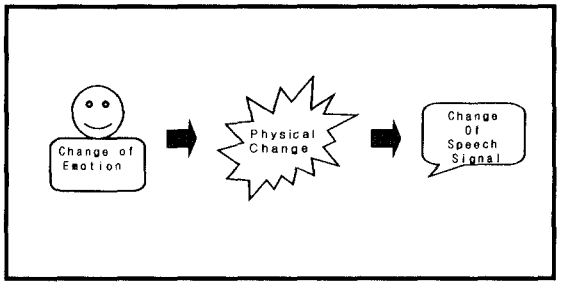


그림 1. 감정의 변화에 따른 음성신호의 발생과정
Fig. 1 Process of development of speech signal caused by change of emotion

본 논문에서는 감성 인식기를 만들기 이전에 음성 신호로부터 감성 특징을 나타내는 요소를 찾아내는 것을 목표로 하고 있다. 음성을 기반으로 감정을 인식하는 방법으로는 (1)대화의 내용에 사용한 단어, (2)톤(Tone), (3)음성신호의 피치(Pitch), (4)포만트 주파수(Formant Frequency), 그리고 (5)말의 빠르기(Speech Speed) (6)음질(Voice Quality) 등이 있지만 그 중에서 현재 가장 많이 접근하고 있는 방법은 피치(Pitch)에 의한 방법이다.[4]

피치[10]는 사람이 귀로들을 때의 음의 높낮이를 말하거나 준 주기적(Quasiperiodic)인 파형을 나타내는 유성음의 1주기를 뜻한다. 보통, 인간의 언어에 사회성이 있듯이, 그 언어에 실리는 감정 또한 사회성을 갖고 있기 때문에 자신의 감정을 타인에게 정확히 알리기 위해선 보편적으로 인정되는 음의 높낮이를 보여야 한다. 이러한 관점에서 보면 거꾸로 음의 높낮이로부터 감정을 알아낼 수 있다는 의미를 내포하고 있다. 그렇기 때문에 피치에 의한 감정인식방법에 많이 접근하고 있는 것이다. 정리하면, 사회화 과정에서 감정표현도 학습되어 감정의 표현이 인간의 원초적인 성질에 의존하지 않고 상호간에 인정되는 수준에서 유사한 패턴을 보이고 있고, 가장 분석이 쉬운 화난 감정의 경우 표현 방식의 분류와 그 범주 내에서 보이는 일반적인 성향을 보인다[2].

그리고 포만트 주파수란 부분 음 중에서 어느 특정 배음들이 강화되는 위치의 주파수를 말하고, 그 부근의 부분까지 포함해서 포만트라고 한다. 그러므로 주파수의 변화가 에너지 분포의 변화로 연결되고 또한 감정의 변화가 생기면 신체적/생리적인 변화가 발생하고 포만트의 변화가 연쇄적으로 일어나기 때문에 이들의 분석도 필수적이다.

본 논문에서 화나는 감정의 특징을 찾기 위해 10명의 연기자에게 몇 가지 대사를 연기하도록 하여 녹음을 하였고, 그 외에도 TV드라마에서 연기자들의 대화를 녹음하였다. 녹음된 파일의 형식은 16bit, mono, 22kHz이고, 이 웨이브 파일을 분석하기 위한 도구로는 Praat 3.9.20(made by Paul Boersma and David weenink)를 사용하였다. 또한 포만트 주파수는 5,500Hz까지, 0.025s의 간격으로 분석하였고, 피치는 75Hz~500Hz의 범위에 대부분 분포하기 때문에 그 범위로 제한을 하였다. 이 도구들을 이용하여 피치와 포만트에 초점을 맞추어 각 하위범주의 특징을 찾아내었다.

2. 화난 감정(Angry)의 피치

일반적으로 화난 감정의 경우는 흥분하여 말이 빨라지고 굉장히 격하다. 표 1은 흥분한 경우와 그렇지 않은 경우의 음성 특징을 비교한 표이다. 흥분한 경우에는 피치의 변화가 크므로 넓은 영역에 분포하는 반면 평서형의 경우는 비슷한 높이로 말하므로 그 영역이 좁다. Intensity또한 흥분한 경우가 평서형보다 자주 나타나므로 더욱 높다. duration의 경우는 뒤에서도 비교를 하겠지만, 화내는 경우에도 화자의 의도에 따라서 그 빠르기가 다르므로 흥분했다고 분명히 더 빨라진다고 말할 수는 없다.

표 1. 흥분한 경우의 음성 특징
Table 1. Voice feature in case of excitement

		Excitement	Ordinary
Acoustic	Pitch	Wide range	Narrow
	Intensity	High	Medium
	Duration	?	

그리고, 화난 감정 등과 같이 흥분한 경우에는 "왜 말을 안 듣냐", "그래서 어떻게 할 건데", "그렇게 하지 말랬지" 등과 같이 대체로 6~10음절로 이루어져 있는 경향을 보인다. 뿐만 아니라 이 정도의 길이의 문장은 다음의 그림 2, 3와 비슷한 억양을 보인다는 것을 실험을 통하여 발견했다. 이들 그림들은 세 가지 대사("왜 말을 안 듣냐", "그래서 어떻게 할 건데", "그렇게 하지 말랬지")를 화를 내면서 말한 것인데, 그림에서 보이는 바와 같

이 한 음절 음절을 말할 때 강세가 주어졌다가 다시 약해지는 패턴이 반복되는 모양을 보인다. 위의 문장들은 10명의 연기자에게 밀폐된 공간에서 서로의 억양에 영향을 미치지 못하도록 하고 연기하였고, 그들의 음성을 분석해본 결과 모두 유사한 Pitch Contour를 보여주었다.

물론, 개개인 별로 극대, 극소점에서 약간의 차이가 있었다. 한편 그림 2에서 『 a : 왜, b : 말, c : 안, d : 냐 』의 음절에 대해서 '왜'는 Rising Pitch를, '말'은 '왜'에서 바로 Falling Pitch의 파형을 보여주었고, '안'의 경우는 c 점 바로 앞의 극대값까지, '을'이 rising Pitch로서 올라간 뒤 falling Pitch의 파형을, 그리고 '냐'의 경우도 c, d 사이에서 '듣'의 발음이 극소점을 중심으로 내려갔다가 올라가는 형태를, 그리고 나서 '냐'는 대체로 d의 모양처럼 내려가고 끝에서 짜증 섞인 억양으로써 "아아"를 내렸다 올렸다 하는데 10명중 한명 정도가 끝에서 그런 형태를 보이지 않았고 나머지는 극대점과 극소점의 폭에서 차이를 보일 뿐 각 음절에 대한 극소점 극대점의 위치는 비슷한 양상을 보였다. 결론적으로 말해서, 화내는 대사의 피치를 음절 및 사람별로 그 주파수를 비교한 결과 남자와 여자의 톤(Tone)이 다르듯이 남자들 사이에서의 톤의 차이가 있을 뿐 억양에는 유사점이 보인다는 것이다. 그리고, 그림 3은 화난 감정에서 유사한 피치형태를 보이는 예로써 제시되었다.

한편 피치의 비교를 위해서 그림 4에 평서형으로 말했을 때의 Pitch contour를 나타낸다. 그림 4를 보면 화가 난 경우와는 달리 강세와 약세의 반복이라기보다는 띄어쓰기 되는 첫음절마다 조금 강세를 주고 점진적으로 약해지는 형태를 나타내고, 억양의 변화가 작음을 알 수 있다.

그리고 피치에 대한 극대점과 극소점에 대한 평균과 편차를 계산해보면 다음 표 2와 같다.

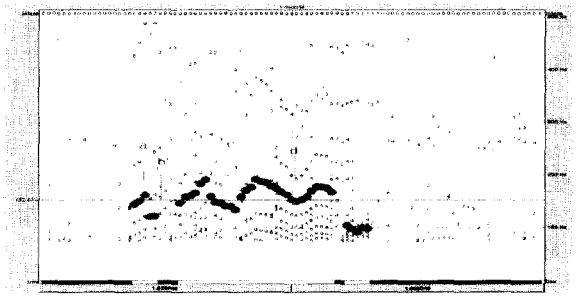


그림 2. "왜 말을 안 듣냐" 에 대한 피치(화남)
Fig 2. "왜 말을 안 듣냐" Pitch contour(Angry)

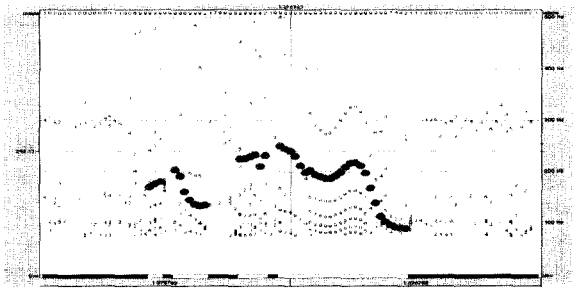


그림 3. "그래서 어떻게 할 건데" 피치(화냄)
Fig 3. "그래서 어떻게 할 건데" Pitch contour(Angry)

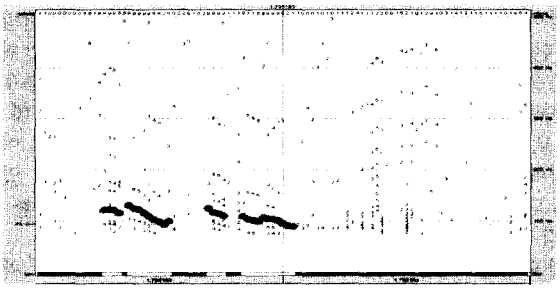


그림 4."그래서 어떻게 할 건데" 피치 (평서형)
Fig 4."그래서 어떻게 할 건데" Pitch contour (normal)

표 2. Pitch의 극대·극소점에 대한 남성의 평균(화냄)
Table 2. Pitch mean of max · min for man.(Angry)

	극대	극소	극대	극소	극대
	a	b	c앞	c뒤	d앞
1	161	134	190	140	190
2	137	125	130	123	130

9	113	111	120	92	147
10	153	141	179	150	200
Mean	146.2	128.2	152.8	120.8	158.6
S.D	32.57	11.26	30.52	25.15	34.33

(※ 남자 연기자 10명의 평균, 단위:Hz)

위의 표를 보면 알 수 있는 바와 같이 편차가 약 10Hz에서 30Hz 정도로 나타났다. 이 정도의 편차는 인식을 하는데 있어서 범주화시키는데 큰 무리가 없다. 그리고, 위에서 예로 든 것들은 인간의 수많은 화난 감정 중에서도 일부에 지나지 않기 때문에 그 감정 전부에 대해서 일반화되었다고 할 수 없고, 여기서 보인 패턴뿐 아니라 많은 패턴의 데이터베이스화를 통해서 만이 좀더 높은 인식률의 감성인식을 할 수 있게 될 것이다. 그리고, 화난 감정에 대한 말의 빠르기를 알아보기 위해 연기자 10명에게 다음의 대사,

- 왜 말을 안 듣냐
- 꼴두 보기 싫어
- 그게 아니야
- 그렇게 하지 말랬지

에 대해서 화났을 때와 평서형으로 말할 때의 빠르기를 비교하였다. 다음의 표 3은 각 대사에 대해서 평균 길이를 기록한 것이다. 표를 보면 알 수 있는 바와 같이, 본 실험에 의하면 화가 났다고 반드시 말이 빨라지는 것이 아니라는 것을 알 수 있었다. "왜 말을 안 듣냐" 와 "꼴두 보기 싫어" 같은 경우는 동일하게 6음절로 이루어져 있음에도 불구하고 음절이 2음절 더 긴 "그렇게 하지 말랬지" 나 1음절 짧은 "그게 아니야" 보다 그 길이의 차가 더 커졌다.

표 3. 화난감정과 무감정의 빠르기 비교
Table 3. speed comparison between Angry and Normal state

	화났을 경우	평서형
왜 말을 안 듣냐	1.08s	1.04s
꼴 두 보기 싫어	0.76s	0.88s
그게 아니야	0.92s	0.71s
그렇게 하지 말랬지	1.09s	1.10s

3. 화난 감정에서의 포만트 특징

포만트는 주파수가 낮은 쪽에서부터 F1, F2, F3, …로 이름을 붙인다. 영어 모음의 예를 들어 포만트를 살펴보면 그림 5(A)의 모음(前舌 母音)은 F1과 F2의 차이가 크고, 그림 5(B)의 모음(後舌 母音)은 그 차이가 작다. 따라서 여기서 전설·후설 모음의 음향학적 차이를 알 수 있다. 신체적 변화와 연관지어 살펴보면, 전설모음의 경우 혀의 가장 높은 부분과 입 천정과의 간격이 벌어질수록 F1의 주파수는 높아지고 F2는 낮아진다. 후설모음은 인두의 간격이 가장 좁아진 부분이 성문으로부터 멀어질수록 F1의 주파수는 낮아지며, 입술이 점점 오무라질수록 F2의 진폭이 줄어드는 것을 볼 수 있다. F1은 입안의 뒤쪽 및 목구멍에서 나는 공명에 기인하는데, 이것은 인두강의 공간에 기인함을 의미한다. 이 인두강은 혀의 높이에 따라 달라지며, 혀의 높이가 높을수록 인두강은 넓어지고 F1이 낮아진다. 그리고 F1은 모음의 고·저 자질과 관계가 깊어서 F1이 낮을수록 고모음이다. 한편 F2는 혀의 가장 높은 부분을 기준으로 하여 입안의 앞쪽 부분의 공명에 기인하므로 공명실의 길이에 좌우됨을 알 수 있다. 즉, 입안의 앞쪽이 넓을수록 F2는 낮아진다. 그러므로 F2는 모음 전·후 자질과 관련이 있어서 F2가 낮을수록 후설 모음이라고 할 수 있다[9].

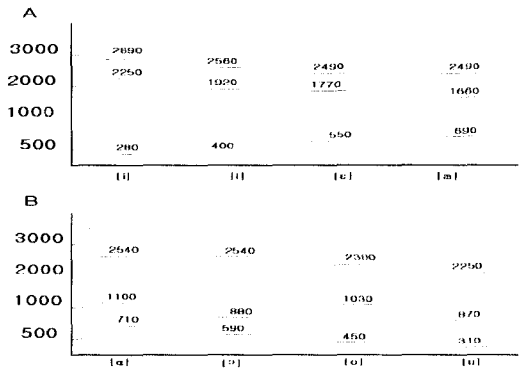


그림 5. 영어 모음의 제1, 제2, 제3 포만트 주파수
Fig 5. F1, F2, F3 of vowels

표 4. 화(anger)표현 분류(TV드라마로부터 발췌)
Table 4. Anger classification (from TV drama)

Category	Examples
의문형	"이제 눈에 뵈는 게 없냐" "일루 못와" "내가 그렇게 ...한 사람인줄 알어" "니가 뭘 안다고 함부로 떠들어" "당신 어딜 그렇게 싸돌아 다니는 거야"
설명형	"형은 형 방식대로 살어 난 내 방식대로 살테니까."
외치는 (Shout)형	"야이 양아치야!" "자꾸 부르지마!" "빨리 해!" "너 이리 못와" "너나 잘해!"

화나는 감정의 특징을 찾기 위한 방법으로 먼저 감정을 하위범주로 분류하고 범주의 최소단위로부터 감정의 특성을 찾아내었다. 문장의 구성에도 의문형, 명령형, 평서형의 구분이 되어 있듯이 감정의 표현 또한 문장으로 구성되어 있기 때문에 화나는 감정 한가지를 표현하는 경우에도 표 4과 같이 3개의 범주로 구분할 수 있다. 그리고 각각의 형에 따라서 각기 다른 특징을 보인다.

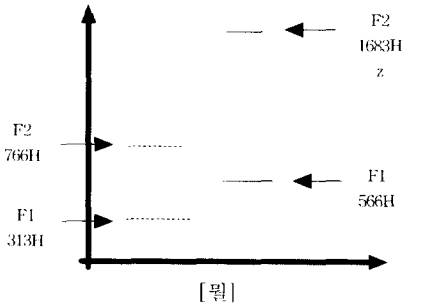
3.1 의문형의 특징

의문형의 형태로 화를 내는 경우를 살펴볼 때, 그 특징은 감정 없는 대사에서와 마찬가지로 말끝을 올리는 것이지만, 감정 없는 대사와의 차이는 말끝에서의 피치의 상승 편차이다. 그런데, 여기서 주의할 점은 대사의 내용이 의문형이라고 해서 화난 감정의 분류에서 반드시 의문형이 되는 것은 아니다. 그 이유는 내용상으로 의문형의 문장이더라도 외치는 형으로 표현될 수 있기 때문이다. 피험자로부터 얻은 웨이브 파일을 분석한 결과 문장의 마지막 극소 피치와 끝 피치의 차를 살펴보면 표 5과 같다.

표 5. 화난 감정의 피치 특징
Tabel 5. Pitch feature of anger

	대사	차
화난 감정	이제 눈에 뵈는게 없냐	135Hz
	내가 그렇게...한 사람인줄 알어	128Hz
	니가 뭘 안다고 함부로 떠들어	116Hz
감정 없음	이거 얼마에요	51Hz
	이거 어때요	11Hz
	뭐라고 해야되지	46Hz
	그럼 어떻게 되는 거야	33Hz

위의 표에서 보는 바와 같이 화난 감정에서는 화자의 격앙된 느낌을 표현하고, 상대를 위협하기 위해서 문장 끝을 감정 없이 말할 때 보다 훨씬 과장되게 말하는 것을 피치의 차로 알 수 있다. 이것 한가지만으로 감정을 알 수 있는 것은 아니지만 여러 가지 요소 중의 한가지로써 중요한 역할을 한다. 그리고, 감정이 섞인 경우에는 무의식적으로 마디마다 강세를 섞는 경우가 많은데 그중 한 음절을 살펴보았을 때 그림 6와 같은 결과를 얻을 수 있었다. 앞에서 언급한 바와 같이 인두강의 공간 변화가 F1에 영향을 미치기 때문에 위 그림 6에서 사용된 '뭘'이란 음절에 감정을 넣게되면 인두강이 좁아지면서 F1이 높아지게 되는 것이다. 그리고, 입안의 앞쪽 공간에 따라서 F2가 영향을 받게되는데 이 경우는 입안의 앞쪽이 좁아지기 때문에 F2가 높아졌다.



(실선) : 화남, …(점선) : 무감정
그림 6. 음절 '뭘'에서의 감정유무에 따른 F1과 F2의 비교
Fig 6.F1, F2 comparison between anger and normal for '뭘'

그리고, 또 한가지 예로써, 그림 7을 보면 똑같이 '너'라는 음절에 대해서 별 감정 없이 조용히 말한 가장 좌측의 부분과 화내는 감정으로 크게 외친 가운데 부분을 비교해보면, 앞에서 말한 바와 같이 F1과 F2에서 변화를 보임을 알 수 있다(그림에서 F1은 점들 중 가장 낮은 주파수대역을 말하고, F2는 두 번째로 낮은 주파수대역을 말한다).

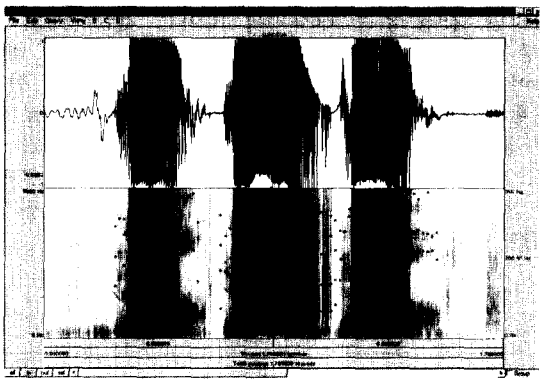


그림 7. '너'음절의 감정에 따른 비교
Fig 7. Comparison of each other emotion for '너'

이러한 특징은 위의 한 음절에만 해당되는 것이 아니고 이외의 다른 음절에 대해서도 성립하는 것을 반복된 실험을 통해 알 수 있다.

3.2 설명형의 특징

설명형의 경우는 표 4의 간단한 예를 통해서 나왔듯이 훈계하는 형식으로 의문형이나 외치는 형과는 달리 긴 문장으로 이루어져 있다. 긴 문장으로 이루어져 있기 때문에 짧은 경우처럼 갑자기 큰 소리를 내는 경우는 드물다. 이런 타입의 경우는 완전히 주기적이진 않지만 준주기적으로 피치의 변화가 주어진다. 그 이유는 화를 내면서 설명하는 경우 상대에게 명확하게 자신의 감정을 표시하기 위해서 한 마디마다 강조를 하기 때문이다. 즉, 예를 들면 드라마 상에서 나온 대화 중 『당신한텐 엄마가 아무 내색도 안한 모양인데 …』의 피치모양을 살펴보면 그림 8의 아래쪽 부분과 같다. 그림에서 원으로 표시된 부분이 위에서 설명한 마디마다의 강조된 부분으로써 강조를 위한 피치의 변화가 일정하다는 것을 알 수 있다.

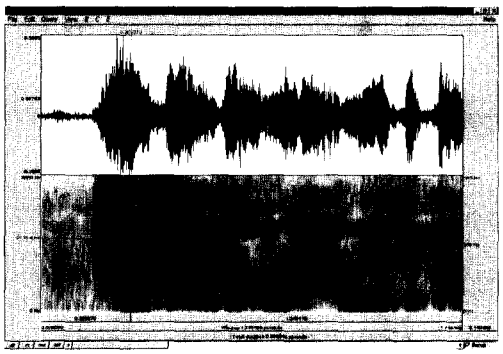


그림 8. 『당신한텐 엄마가 아무..』의 신호분석
Fig 8. Signal analysis for 『당신한텐 엄마가 아무..』

그림 9도 동일한 설명을 보충해주는 그림으로써 다른 드라마로부터 녹취한 대사로써 강조되는 부분을 원으로 표시해 놓았다.

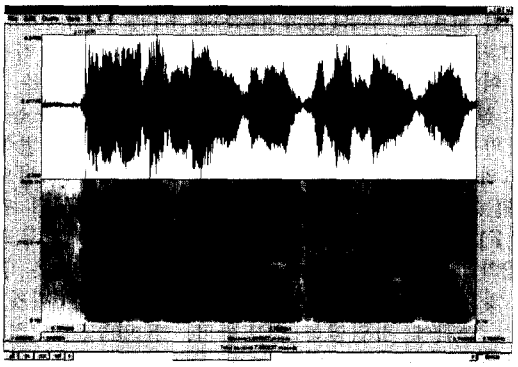


그림 9. 『엄만 엄마의 인생이..』 신호분석
Fig 9. Signal analysis for 『엄만 엄마의 인생이..』

3.3 외치는 형(Shout type)의 특징

외치는 형은 3.1절에서 설명한 의문형과 어미 가 거의 유사하다. 즉, 의문형의 경우도 기본적으로 화내는 중의 격렬한 의사표현의 한가지 방법이기 때문에 그 신호의 특성은 외치는 형과 같다. 단지, 의문형의 고유특징인 어미에서의 피치의 변화가 다를 뿐이다. 포만트 주파수에서의 특성은 같다. 이 경우에도 피치모양으로 봤을 때, 강조되는 부분(그림 8, 그림 9참조)에서는 신체적 변화가 평정이 유지되고 있을 때 보다 크게 발생하기 때문에 그 때의 포만트 변화가 두드러지게 나타난다. 이것에 대한 예는 그림 7에서 설명하였다. 즉, 본 논문에서 외치는 형과 의문형의 구분만은 어미에서의 피치 변화 성향이 의문형인 것과 그렇지 않은 것을 외치는 형으로 정했다.

4. 결론 및 향후과제

본 논문에서는'화(angry)'의 좀 더 용이한 분석을 위해서 3가지 타입으로 분류를 하였고, 감정의 변화에 따라 신체적으로 생기는 변화가 주파수에 영향을 주는 것을 드라마와 직접 연기하여 녹음한 대화내용을 분석하여 확인하였다. 그리고, 감정을 분석하는 요소로써 피치나 포만트의 분석도구를 사용하였다. 본 논문에서는 'angry'의 경우의 특징을 주로 살펴보았는데, 강조 점에서의 포만트의 특징은 'angry'에서 뿐만이 아니라 기쁜 감정에서도 나타날 수 있다. 이런 중복되는 기본적인 특징들 때문에 다른 요소들과 함께 감정을 인식해야 할 것이다. 그리고, 차후에는 이들을 좀 더 체계화시켜서 감성인식 시스템을 구축할 수 있도록 할 것이다.

참 고 문 헌

[1] C. Becchetti, and L. P. Ricotti, "Speech Recognition Theory and C++ Implementation," John Wiley & Sons, New York, 1998.

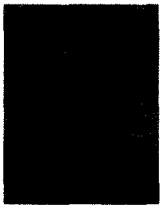
[2] F. Dellaert, T. Polzin and A. Waibel, "Recognizing Emotion in Speech," in Proc. International Conf. on

Spoken Language Processing, Philadelphia, USA, pp. 1970-1973, October 3-6,1996.
[3] L.S. Chan, H. Tao, T.S. Huang, T. Miyasato, and R. Nakatsu, "Emotion Recognition From Audiovisual Information", IEEE Second Workshop on Multimedia Signal Processing,pp.83-88,1998.
[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsi, S. Kollias, W. Fellenz, and J. G. Taylor ,"Emotion recognition in Human Computer Interaction", *IEEE Signal Processing Magazine*, pp. 33-80, January, 2001
[5] T.L. Nwe, F.S. Wei, L.C.De Silvia,"Speech Based Emotion Classification" ,IEEE, 2001.
[6] T.W.Parsons,"*Voice and Speech Processing*," McGraw-Hill, New York, 1986.
[7] T.Yamada, H.Hashimoto, and N. Tosa, "Pattern Recognition of Emotion with Neural Network," Proceedings of the 1995 IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation, vol. 1 , pp. 183-187, 1995.
[8] Y. Iwano et al. "Extraction of Speaker's Feeling using Facial Image and speech," in Proc. IEEE International Workshop on Robot and Human Comm. RO-Man '95, Tokyo, Japan, pp. 101-106, July 5-7, 1995.
[9] 이규식, 석동일 "*청각학*", 대구대학교출판부, pp. 49-60, 1996.
[10] 박경범, "*선형예측분석법에 의한 음성의 압축과 재생*", 도서출판 하늘소, pp. 53-60, 1994

저 자 소 개

심귀보(Kwee-Bo Sim)
1984년 : 중앙대학교 전자공학과 공학사
1986년 : 동 대학원 전자공학과 공학석사
1990년 : The University of Tokyo 전자공학과 공학박사
1997년 ~ 현재 : 한국퍼지 및 지능시스템학회 편집이사 및 논문지 편집위원장
1997년~현재 : 한국퍼지및지능시스템학회 편집이사
2000년 ~ 현재 : 제어자동화시스템공학회 논문지 편집위원 및 직선평위원
2000년 ~ 현재 : 대한전기학회 제어 및 시스템 부문회 편집위원 및 학술이사
1991년 ~ 현재 : 중앙대학교 전자전기공학부 교수

관심분야 : 인공생명, 진화연산, 지능로봇시스템, 뉴로-퍼지 및 소프트 컴퓨팅, 자율분산시스템, 로봇비전, 진화하드웨어, 인공면역계 등



박창현(Chang-Hyun Park)
2001년 : 중앙대학교 전자전기공학부 공학사
2001년 ~ 현재 : 동 대학원 전자전기공학부 석사과정

관심분야 : 진화연산, 신경회로망, 감성정보처리 등