



말소리와 성격 이미지 Speech sound and personality impression

이 은 영 · 유 혜 옥*
Lee, Eunyung · Yuh, Heaok

Abstract

Regardless of their intention, listeners tend to assess speakers' personalities based on the sounds of the speech they hear. Assessment criteria, however, have not been fully investigated to indicate whether there is any relationship between the acoustic cue of produced speech sounds and perceived personality impression. If properly investigated, the potential relationship between these two will provide crucial insights on the aspects of human communications and further on human-computer interaction. Since human communications have distinctive characteristics of simultaneity and complexity, this investigation would be the identification of minimum essential factors among the sounds of speech and perceived personality impression. The purpose of this study, therefore, is to identify significant associations between the speech sounds and perceived personality impression of speaker by the listeners. Twenty eight subjects participated in the experiment and eight acoustic parameters were extracted by using Praat from the recorded sounds of the speech. The subjects also completed the Neo-five Factor Inventory test so that their personality traits could be measured. The results of the experiment show that four major factors(duration average, pitch difference value, pitch average and intensity average) play crucial roles in defining the significant relationship.

Keywords: voice image, acoustic features, speech sound, personality, impression, emotion recognition

1. 서론

말소리의 음향학적 특성과 사람의 이미지 간에 연관이 있는지? 있다면 어떤 요소와 어떻게 상관관계를 이루는지? 이런 의문을 갖는 일은 쉬우나 직접 규명해내기는 어렵다. 왜냐하면 실제 커뮤니케이션 상황에서는 말소리의 음향적 특성만 분리해서 인지할 수 없는데다, 말소리가 실어 나르는 의미내용이나 언어 외적인 요소까지 포괄적으로 인식하기 때문이다. '인간 대 인간(human-human interaction)'에는 이런 복합적인 요소들의 영향에도 불구하고 각 개인에 대한 이미지와 성격을 본능적으로 판단한다. 최근에는 컴퓨터가 인간과 유사한 방법의 대화시스

템을 구동할 수 있도록 하는 연구가 활발해지고 있으므로, 인간의 이런 판정능력을 로봇에게 입력하여 인간을 흉내 내게 할 수 있다면 '인간 대 컴퓨터 간(human-computer interaction)'에 의사소통이 훨씬 더 자연스러워질 것이고, 나아가서 사용자-적응 발화(user-adaptive utterance) 유형과 같은 대화유형으로 재생산될 수도 있을 것이다. 화자의 메시지에 실려 오는 물리적 에너지는 청자에게 화자의 여러 정보를 가져다주는 정보매체로 화자의 성격에 대한 정보까지도 얻게 된다.

실험음성학 연구는 정보매체로서의 말소리 특성을 최대한 객관화하고 수치화함으로써 음운기술, 음성인식합성, 언어치료, 커뮤니케이션 등 다양한 분야에서 많은 발전을 이루고 있

* 경북대학교, hoakyuh@knu.ac.kr, 교신저자

Received 1 November 2017; Revised 28 November 2017; Accepted 6 December 2017

다. 그런데 보통 사람들은 음성학자와는 달리 다소 주관적이기는 하지만 특별한 의도나 의식 없이도 나름대로 각자 소리를 분석하고 인식한다. 특히 청자가 화자의 목소리를 듣고 성격이나 이미지를 인식하는 부분은 어떤 기계도 대신할 수 없는 대단한 능력이다. 보통 사람이 그토록 쉽게 감지하는 그 인식 구조에 접근하려면 인식 요소 중 가장 기본이 되는 최소한의 소리특성에서부터 출발하여 그 요소를 하나씩 확장시키거나면서 구조화를 시도해볼 필요가 있다고 본다. 따라서 처음 만난 사람의 말소리를 들으면서 은연중에 성격유형을 예측할 수 있는 동시적이고 복합적인 인식 능력을 구조화하려는 큰 틀의 연구를 위한 출발점으로써 본 연구를 시작한다. 주된 목적은 사용자의 발화유형을 만들어 내는 여러 요인들 중 어떤 물리적인 요소를 매개변수화 할 수 있는지를 찾아내어 화자의 성격 유형과의 연관성을 밝히는 일이다. 즉 말소리의 어떤 음향특성이 화자의 성격유형을 가능하는 단서로 규정될 수 있는지, 더 나아가서 최소 기본 매개변수만으로 성격유형을 구분할 수 있다면 그 최소화는 어디까지 가능한지가 본 연구의 초점이다.

2. 선행연구

최근 동향을 보면 컴퓨터에 성격을 부여하여 인간과 컴퓨터 간에 보다 더 자연스러운 상호작용을 할 수 있도록 하는 자동 대화시스템에 관한 연구가 활발히 이루어지고 있다[1]. 예를 들면 컴퓨터에게 지배적인 성격과 복종적인 성격을 각각 부여하고 사용자와의 상호작용을 관찰하거나[2], 외향적 성격과 내향적 성격을 부여했을 때 사용자의 반응을 관찰하기도 하였다. 최근 TTS(Text-to-Speech) 분야의 관심사도 발화의 명료성뿐만 아니라 자연스러움을 향상시키는 일이다. 명료성에 대한 기술개발은 97%정도에 이르지만 자연스러움 즉 사람같이 성격을 지니는 말소리 단서에 대한 연구와 기술은 아직 부족하다[3]. 합성 목소리에 인성까지 담아낼 수 있는 자연스런 발화를 만들기 위해 핵심 단서가 될 수 있는 운율적 요소를 찾아내는 연구가 진행되고 있다.

인공지능 로봇이나 GPS와 같은 멀티미디어 정보검색시스템의 수요가 점차 확장되면서 점점 더 자연스런 의사소통 시스템을 원하는 인간의 욕구를 피할 수 없으므로 말소리와 성격 간의 관계를 인지하는데 관여하는 단서들을 식별해 내기 위한 연구가 계속 되고 있다. 예컨대 피치와 발화속도가 다양한 감정표현과 서로 상관관계를 갖고 있다는 점에 주목한 연구[4]나, 다양한 발화유형에 근거해서 인종, 나이, 성격, 사회적 지위 그리고 성별을 알아보고자 한 연구[5]들이 있다. 또한 말이 외에 목소리, 표정, 제스처와 같이 의사소통에서 정보전달의 수단이 되는 여러 가지 단서(paralinguistic cues)에 대한 연구도 있지만, 한편으로는 이에 따른 정보를 모두 상세히 기술하는 것이 과연 효과적이냐에 대한 시각도 있는데, 이는 다양한 의사소통 방법 가운데 필요와 중요도에 따라 선택 가능한 것으로 보았기 때문이다[6].

한편 음성 특성을 분석하거나 발화내용을 해석하기보다 오

히려 시각적으로 의사소통을 할 때 성격 파악이 더 쉬울 수도 있다. Zukerman, Amidon, Bishop 와 Pomerantz의 연구결과[7]는 얼굴 모습과 말소리의 상호관계 판단은 경우에 따라 중요도가 달라질 수 있다고 했다. Burns and Beier의 연구[8]에서도 발화에서 얻게 되는 정보는 목소리뿐만 아니라 표정에서도 비롯된다고 하였다. 불같은 성격, 낙천적 성격, 우울한 성격들은 얼굴표정에서 훨씬 더 많은 정보를 얻는 반면 불안정한 성격의 경우는 얼굴표정보다 음색과 더 연관성이 있는 것으로 보았다. 비록 메시지 전달이 음향수단보다 시각수단에 의존하는 경향이 높다고 하더라도 수신자의 메시지 해석방법은 경우에 따라 다르게 작동한다[9]. 즉 발신자가 의도적으로 전달 정보를 조작할 수도 있고 수신자가 자의적으로 해석할 수도 있으므로 이러한 발신자-수신자 간의 상황요인이 소통수단을 선택하는데 중요한 요소가 되기도 한다[10].

3. 연구방법

살펴본 바와 같이 응용분야에서는 모든 음성특성의 상세한 기술보다는 발신자-수신자 간의 상황요인에 따라 필요성과 중요도가 선택적으로 작용한다. 연구결과가 실용화로 이어지려면 다양한 시각에서 기술이 필요하기에 전례 없는 새로운 시도를 해보고자 한다. 즉 화자의 말소리를 듣고 바로 성격유형을 예측하는 매커니즘처럼 복잡함 속에서 단순함을 읽어내려는 역방향의 연구방법이다. 우선 화자의 말소리 특성을 모두 분석한 후 성격유형과 연결 요소들을 파악하고 그 중 기여도가 낮은 순으로 요소를 하나씩 배제하면서 마지막에 가장 핵심 요소를 남기는 방식이다.

그러나 복합적이고 세부적인 의사전달 행위를 통해 발신자의 이미지를 도출하려는 연구는 무척 어려운 도전이다. 본 연구는 이와 같은 출발점에서 있는 연구이므로 일단 복잡한 전체를 단순화시켜 보려는 목적에 초점을 맞추고 말소리만으로 화자의 성격을 예측할 수 있는지에 대한 연구를 추진하였다. 또한 심리학에서 분류하는 기본성격의 유형과 말소리의 음향적 특성 상호 간에 연관성을 찾는 작업 역시 처음 시도하는 연구주제라서 명확한 연구결과를 얻으려면 현실적으로 대형 공동프로젝트를 추진해야만 제대로 된 연구가 가능하다. 그러므로 연구 타당성 확인을 위한 예비연구로써 현실적인 여건 하에 가능한 범위 안에서 계획되었고 후속 연구에 대한 가능성을 확인하는 것이 중요한 목적임을 밝혀둔다.

본 연구를 위해 세 가지 실험을 실시했다. 첫 번째 실험에서는 사람들(피실험그룹 A)이 말소리를 듣고 떠올리는 이미지를 조사하고 분류하는 실험이었고(실험 1), 두 번째는 사람들(피실험그룹 B)의 말소리를 녹음하고 음향 특성을 측정 분석하였고(실험 2), 세 번째는 목소리를 제공한 피실험그룹 B 전원이 각자 실시한 성격유형검사 결과지를 수합하여 분석하였다(실험 3).

3.1. 성격유형조사

실험 1은 불문학과 2학년생 35명(남 6명, 여 29명)을 대상(피실험그룹 A)으로 개별과제를 발표하는 한 학기 동안 발표자의 소리 이미지를 모든 수강학생들에게 자유롭게 생각나는 대로 적게 하였다. 본 실험은 수신자가 느끼는 발신자의 이미지 즉 타인의 시각으로 성격을 분류하였다. 학생들이 이미지를 나타내기 위해 자발적으로 사용한 형용사들을 모아 정리 분류하였더니 아래 <표 1>과 같이 다섯 가지 성격유형(Five Image Types)이 나왔다.

이 분류(FIT)에 대한 객관성 확보를 위해 심리학에서 쓰는 성격유형 분류를 이용하기로 하였다. 그러나 심리검사는 질문지를 통해 자기 생각을 전달하고 심리학자가 해석한 것이므로 타인의 관점에서 본 FIT와는 달리 심리검사는 본인의 시각이 담겨있다. 따라서 말소리(실험 2)와 성격(실험 3) 간의 연관관계에 대해 좀 더 객관화된 검증할 수 있는 자료가 될 것이다.

표 1. 성격의 다섯 가지 유형 분류

Table 1. Five Image Types for Personality Classification

유형 Type	호감도 Average good feeling	이미지 Image
A	호감/비호감 Favorabl./Unfavor.	유쾌한, 활기있는, 사교적인, 외향적인 cheerful, lively, sociable, outgoing
B	호감/비호감 F / U	조심성 있는, 진실된, 진지한, 성실한, 조용한, 안정감 있는, 믿음직한, 여유있는, 친절함, 부드러운, 온화함, 편안한 trustful, relaxed, gentle, soft, mild, comfortable
C	호감/비호감 F / U	자신만만한, 인내심 있는, 확고한, 의지력 있는, confident, persevering, determined, strong willed, 자신감 있는, 단호한, 고집스런 self-assured, resolute, persistent
D	호감/비호감 F / U	내성적인, 자신감 없는, 소극적인, 조심한, 비판적인, Reserved, unconfident, passive, timid, pessimistic, 확신없는, 요령없는, 어설픈, 부자연스러운, 생기없는 doubtful, tactless, clumsy, awkward, lifeless,
E	호감/비호감 F / U	경솔한, 신경질적인, 불안한, 예민한, frivolous, nervous, insecure, sensitive, 화를 잘 내는, 냉소적인 easily offended, snarky

한편 McCrae 연구팀이 제안한 성격특성검사 Neo-Five Factor Inventory (Neo-FFI)[11] 역시 BFI를 바탕으로 보완된 성격테스트방법으로 동일한 5 가지 유형으로 성격을 분류하였고, 이후에도 세부적인 수정을 하였다[12].

3.2. 기존 성격유형검사 BFI 에서 Neo-FFI 까지

심리학에서 일반적으로 사용하는 성격분류 방법에 관한 연구에서는, 각기 다른 인간의 생각과 행동에 대한 심리적인 차이를 규명하기 위해 오랜 연구와 개선과정을 통해 기본적인 성격 요인을 Big Five Inventory(BFI)로 압축시켜 놓았다[13]. 성격 평가에 가장 널리 사용되어온 BFI의 다섯 가지 성격 유형과 각 해당 하위 특성은 다음과 같다.

- Openness to Experience : Artistic, Imaginative, Curious, Insightful, Original, Wide interests
- Conscientiousness : Efficient, Organized, Planful,

- Reliable, Responsible, Thorough
- Extraversion : Active, Assertive, Energetic, Outgoing, Talkative
- Agreeableness : Appreciative, Forgiving, Generous, Kind, Sympathetic, Trusting
- Neuroticism : Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying

BFI의 장점을 들자면 성격을 평가하는데 일반적으로 사용되어 왔고, 복합적으로 나타나는 개인의 성격을 다섯 가지 특성 중 어디에 해당하는지 분류하고 그 일치 정도를 점수로 보여준다는 점이다. 사실 성격을 한마디로 단정 짓기는 쉽지 않다. 왜냐하면 “성격은 심리적 메커니즘과 더불어 개인의 사고, 감정, 행동의 특정 패턴이 함께 고려되는 잠재적인 구조”이기 때문이다[14]. 또 G.W. Allport의 연구[15]에 따르면 성격이란 “개인의 환경에 대한 고유한 적응을 규정하는 정신적 물리적 조직으로서 개인 안에 존재하는 역동적 체제”라고 하였다.

표 2. 새로운 다섯 가지 목록

Table 2. Neo-Five Factor Inventory

Openness to experience (vs. closedness)	Conscientiousness (vs. lack of direction)	Extraversion (vs. introversion)	Agreeableness(vs. antagonism)	Neuroticism (vs. emotional-stability)
Fantasy	Competence	Warmth	Trust	Anxiety
Aesthetics	Order	Gregariousness	Straight forwardness	Hostility
Feelings	Dutifulness	Assertiveness	Altruism	Depression
Actions	Achievement striving	Activity	Compliance	Self-Consciousness
Ideas	Self-discipline	Excitement seeking	Modesty	Impulsiveness
Values	Deliberation	Positive-emotion	Tendermindedness	Vulnerability to Stress

위 FIT 도표에서 보듯이 자체적으로 진행한 연구에서는 이미지 유형에 대해 명명을 하지는 않으나 FIT 5유형과 기존의 심리학자들이 사용하는 BFI에서 NEO-FFI에 이르는 5유형이 일치하였고 그 하위에 분류된 성격특성도 호감/비호감에 따라 용어 선택에 차이는 있으나 대체로 일치하였다. 따라서 본 연구에서는 객관성이 검증된 심리학에서 사용하는 다섯 가지 성격유형인 O, C, E, A, N, 즉 경험에 대한 개방성(openness), 성실성(conscientiousness), 외향성(extraversion), 동조성(agreeableness), 정서적 불안정성(neuroticism)을 말소리 이미지 분류에 그대로 가져와 사용할 수 있는 합당한 근거를 얻게 되었다.

4. 실험방법 및 분석결과

4.1. 실험 방법과 절차

자료체 구성을 위해 본 학과 수강생 32명(남 4명, 여 28명)을 피실험자로 삼았다(피실험그룹 B). 학과 특성 상 남성이 극소수여서 추가인원을 충분히 확보할 수 없는데다, 남녀 기본주파수 값이 달라 남성의 포함/미포함 여부에 따라 측정값 평균이 왜

곡될 우려가 있으므로 부득이 여성 28명의 분석 자료만을 대상으로 기술하였음을 밝혀둔다.

녹음자료는 두 가지 점에 유의하여 선택하였는데, 한 가지는 운율특성 측정에 적합하도록 액센트 단위(Accent Phrase) 시작 부분에 반드시 모음이 오도록 구성하였고, 다른 한 가지는 성격에 따라 감정표현의 정도가 얼마나 달라지는 가를 보기위해 동일 발화체를 평온(Calm), 기쁨(Joy), 슬픔(Sad), 놀람(Surprise), 네 가지 감정을 담아 말하도록 요구하였다. 의사전달에서 기능적 비중으로 봤을 때 분절음은 의미전달을, 운율요소는 감정전달을 더 많이 담당하고 있기 때문에 성격유형을 파악하기에는 분절음 특성 보다 운율요소가 더 효과적이고 평온, 기쁨, 슬픔, 놀람은 감정의 기본요소이기 때문이다. 감정이 담긴 자연스러운 말소리를 얻기 위해 “아영이가 오늘 오후에 인천으로 이사간대”라는 말에 각각의 감정을 불러일으키는 감탄사를 앞에 넣어서 아래와 같이 말하게 하였다.

- 평온(Calm) : 근데, ...
- 기쁨(Joy) : 우와! ...
- 슬픔(Sad) : 야, 어쩐냐! ...
- 놀람(Surprise) : 뭐라고? ...

기기는 CSL 모델 4300B를 사용하여 음성실험실에서 녹음하고 파이프라인을 Praat에서 분석하였다. 학생들이 직접 실험실에 가서 가능한 한 동일조건에서 녹음하도록 주의를 기울였고, CSL을 이용하여 입과 마이크 간 간격은 15cm 유지, 소리 입력 레벨도 동일하게 유지하였다.

1 차 분석에서는 아래 <표 3>과 같이 각 화자별로 네 가지 발화체에 대해 1) 실현 길이, 2) 최고 피치, 3) 최저 피치, 4) 최고/최저 피치 차이 값, 5) 피치 평균 값, 6) 최고 세기 값, 7) 세기 평균 값, 8) 휴지(Pause) 개수와 실현 길이, 9) 억양곡선을 측정하고, 각 측정변수에 대한 전체 화자 평균값을 도출하였다. 단, 네 가지 감정 실현 유도용으로 덧붙인 감탄사 부분은 측정에서 제외 시

켰다. 그러나 피실험자 간 차이가 기대치보다는 뚜렷하게 나타나지 않았는데, 그 이유는 나이, 성별, 거주 지역, 사회 계층, 문화 계층, 지식수준 등에서 동일계열 집단이기 때문으로 보았다.

1 차 분석에서 측정된 휴지(Pause) 개수와 억양곡선은 2 차 분석에서는 배제하였다. 문장이 짧아 휴지 개수가 2-3 개로 거의 비슷하게 나타났고 억양곡선은 피치 변화를 가시적으로 파악하고 실험을 디자인 하는데 참고자료로만 활용하였다.

2 차 분석에서는 8 가지 측정변수만 사용하기로 하고, 길이 외에 세기와 피치 변화 폭이 슬픔 ≤ 평온 < 기쁨 < 놀람의 순으로 나타났기 때문에 감정의 표현이 가장 강한 놀람(Surprise)을 평온(Calm)의 경우와 비교함으로써 화자의 성격유형에 따라 어떤 변화가 일어나는지를 살펴보았다. 분석 항목은 1) 발화체 전체 길이 평균값(Duration_ave), 2) S와 C 간 최고/최저 길이 평균 값 차이(Duration_dv), 3) 각 발화체 내의 최고/최저 피치 차이 평균 값(Pitch DV_ave), 4) S와 C간 최고/최저 피치 평균 값 차이(Pitch DV_dv), 5) 발화체 전체 피치 평균값(Pitch Mean_ave), 6) S와 C간 피치 평균 값 차이(Pitch Mean_dv), 7) 세기 평균 값(Intensity_ave), 8) S와 C간 최고/최저 세기 평균 값 차이(Intensity_dv), 총 8 개로 측정을 진행하였다.

한편 피실험자로 녹음에 참가한 동일 학생 28 명에 대한 성격유형검사를 실시하였다. 학생들이 각자 인터넷을 통해 고용노동부 워크넷(Worknet <http://www.work.go.kr/>)에 개별적으로 검사 신청을 하고 <직업선호도검사 L형> 검사지를 선택하여 검사하고 받은 성격검사결과지를 제출토록 했다. 심리상담센터를 찾지 않고 고용노동부 홈페이지를 이용한 이유는 장소, 시간, 비용에 대한 제약 없이 편하게 이용할 수 있는데다, 사용한 성격분류유형이 Neo-FFI의 5 유형과 동일했기 때문이다.

위에서 제시한 8 변수는 동일유형 화자별로 묶어서 해당 측정값을 평균내고, 8 변수 간 서로 다른 측정단위와 수치범위에 상관없이 직접 비교 가능한 Z-score로 변환한 후, 8 변수와 성격유형 간에 어떤 분포를 보이는지와 분포결과가 유의미한지를 그래프를 통해 알아보았다.

표 3. <1차 분석>의 예 : 화자, 발화체, 측정변수

Table 3. Example of Analysis: Speakers, Sentences, and Measurement Parameters

5) 김한지	평온	기쁨	슬픔	놀람	
Duration (sec)	2.457	2.418	2.753	2.532	
Pitch (Hz)	Maxi	248.52	338.197	347.225	404.568
	Mini	82.898	79.208	92.087	76.151
	Difference Value	165.622	258.989	255.138	328.417
Intensity (dB)	Mean	191.607	223.354	222.266	292.004
	Maxi	76.554	79.434	79.317	77.101
	Mean	72.133	74.629	74.157	70.848
Pause 1 (sec)	0.112	0.125	0.096	0.098	
Pause 2 (sec)	0.088	0.107	0.093	0.076	
Pause 3 (sec)	/	0.058	/	0.07	
Intonation Contour					
texte	아영이가 오늘 오후에/ 인천으로/ 이사간대	아영이가 오늘/ 오후에/ 인천으로/ 이사간대	아영이가 오늘 오후에/ 인천으로/ 이사간대	아영이가 오늘/ 오후에/ 인천으로/ 이사간대	

4.2. 분석결과

성격유형 분석결과 개인 별 성격분포는 개방성(O) 5명, 성실성(C) 5명, 외향성(E) 1명, 동조성(A) 14명, 정서적 불안정성(N) 3명으로 드러났다.

성격유형분류는 직업선호도검사 검사지 내에 Neo-FFI 유형과 동일한 성격유형의 검사가 포함되어 있어서 유형별 점수를 그대로 가져와 위 <그림 1>과 같은 그래프를 얻었다. <그림 1>에서 보듯 다른 유형에서도 어느 정도 높은 값을 보이기도 하였으나 일단 이번 연구에서는 가장 강세를 보인 성격유형을 대상으로 분류하였다. 피실험군 가운데 53.5%가 동조성으로 나타났고 개방성 17.8%, 성실성 17.8%, 정서불안정성 10.7%, 외향성 3.5%로 드러났다.

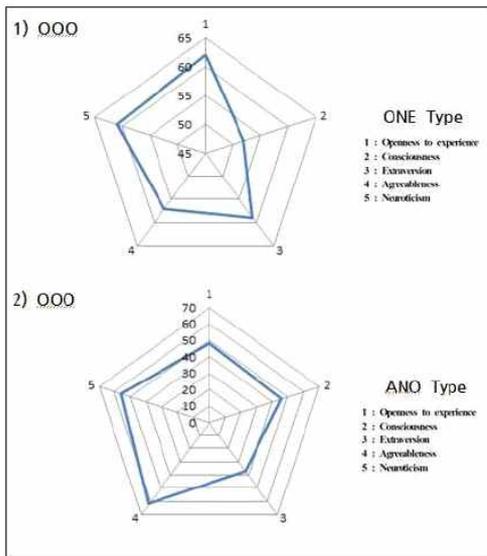


그림 1. 성격유형 분석결과 개인 별 성격분포 예
Figure 1. Individual Personality Distribution by Image Type Analysis

4.2.1. 성격유형과 8음향변수 간 상관관계

검사결과에 따라 화자들을 성격 5유형 - 개방성, 성실성, 외향성, 동조성, 정서적 불안정성 - 으로 분류하고 동일 유형에 속하는 화자들의 측정값 평균을 8개 항목별로 구한다음 Z-score로 변환하였다. <표 4>는 성격유형 별로 8 음향변수 - 발화체 전체 길이 평균값, 놀람과 평온 간 최고/최저 길이 평균값 차이, 각 발화체 내의 최고/최저 피치 차이 평균값, 놀람과 평온 간 최고/최저 피

치 평균값 차이, 발화체 전체 피치 평균값, 놀람과 평온 간 피치 평균값 차이, 세기 평균값, 놀람과 평온 간 최고/최저 세기 평균값 차이-의 Z-score 변환 값을 나타낸다. <그림 2>를 살펴보면 길이, 피치, 세기의 변화폭이 변수 간에 다소 차이는 있으나 전체 스케일은 N > E > O > A > C 순으로 점점 작아지고 있다. 다시 말해 음향변수의 변화가 크면 불안정하거나 외향적인 성격이고, 변화가 작으면 성실하거나 동조적인 성격일 가능성이 높다. 감정표현 측면에서도 음향변수의 변화가 크면 전자에 적으면 후자에 해당될 확률이 크다.

4.2.2. 성격유형과 음향변수 변화에 따른 상관관계

이번에는 8변수의 중요도에 따라 잉여적인 특성부터 하나씩 제외하면서 단순화 가능성을 단계적으로 검토해보았다. 맨 먼저 놀람과 평온 간 최고/최저 세기 평균값 차이(Intensity_dv)를 제외한 7변수(<그림 3>), 그 다음은 놀람과 평온 간 최고/최저 피치 평균값 차이(PitchDV_dv)를 추가로 제외한 6변수(<그림 4>), 놀람과 평온 간 피치 평균값 차이(PitchMean_dv)를 추가로 제외한 5변수(<그림 5>), 놀람과 평온 간 최고/최저 길이 평균값 차이(Duration_dv)를 추가로 제외한 4변수를 검토하였고(<그림 6>), 끝으로 발화체 내의 최고/최저 피치 차이 평균값(PitchDV_ave)을 제외한 3변수만을 살펴보았다(<그림 7>). 왜냐하면 소리의 기본 세 가지 음향변수인 장단, 고저, 강약에 따라 화자의 목소리 기본 특성을 나타내주는 매개변수, 1) 발화체 전체 길이 평균값, 2) 발화체 전체 피치 평균값, 3) 세기 평균값, 세 가지만으로도 성격유형과의 상관관계를 예측할 수 있는지 살펴보고 그 중에서 가장 최적화된 상관관계를 보여주는 매개변수를 찾아보기 위해서였다. 다섯 가지 그림을 비교해 보니 4변수(<그림 6>) 즉 발화체 전체 길이 평균값, 발화체 내의 최고/최저 피치 차이 평균값, 발화체 전체 피치 평균값, 세기 평균값으로 성격유형 간의 차이가 가장 잘 드러났고, 세 가지 소리 기본 파라미터인 길이, 피치, 세기만으로도 비교적 차이가 잘 드러났다(<그림 7>).

표 4. 성격유형과 8변수 평균값 z-score
Table 4. Personality Types vs. z-score of 8 Parameters

Type	Z_Duration_ave	Z_Duration_dv	Z_PitchDV_ave	Z_PitchDV_dv	Z_PitchMean_ave	Z_PitchMean_dv	Z_Intensity_ave	Z_Intensity_dv
O Type	0.792	-0.416	0.381	-0.016	0.120	-0.119	0.363	0.310
C Type	-0.055	-0.029	-0.242	0.323	-0.468	-0.531	-0.411	-0.532
E Type	-0.218	0.609	-0.299	0.059	1.193	-0.100	1.200	-0.356
A Type	-0.205	0.114	0.033	-0.206	-0.237	0.033	-0.210	-0.190
N Type	-0.194	0.005	-0.287	0.433	1.292	0.960	0.662	1.378

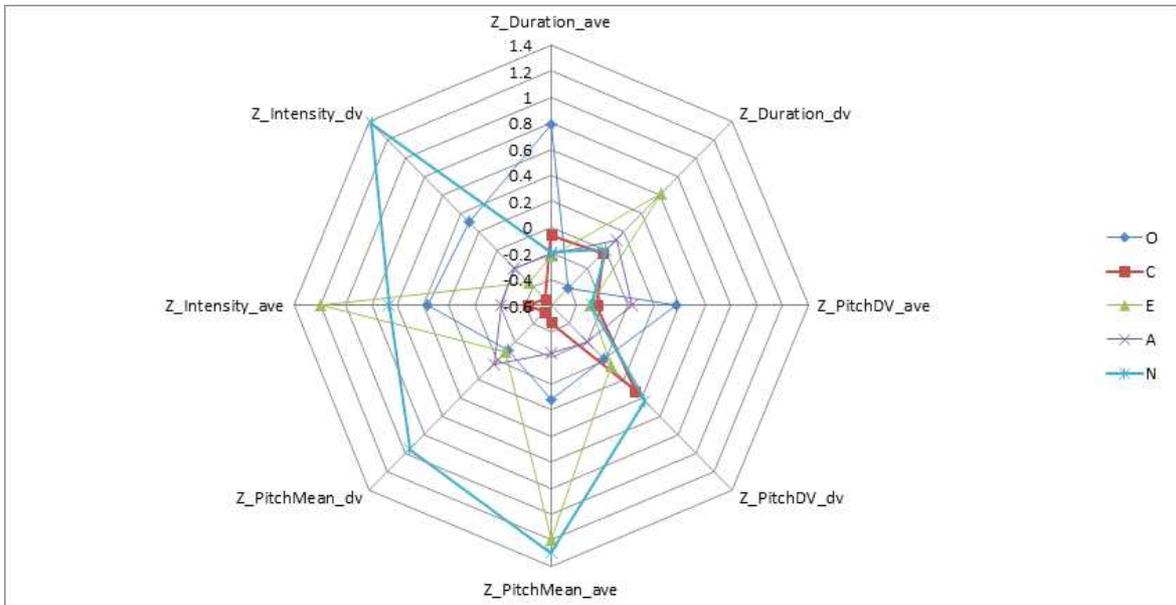


그림 2. 성격유형과 8변수 평균값 z-score
Figure 2. Personality Types vs. z-score of 8 Parameters

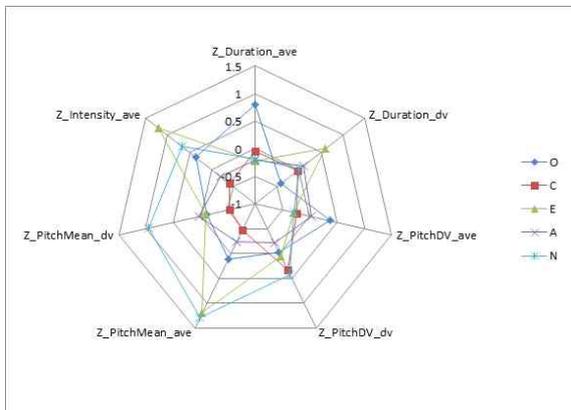


그림 3. 성격유형과 7변수(Intensity_dv제외)
Figure 3. Personality Types and 7 parameters (except Intensity dv)

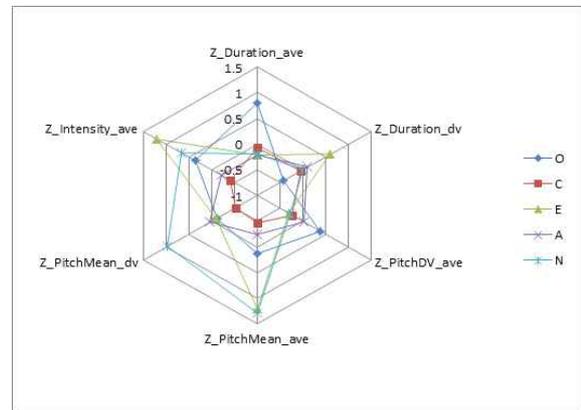


그림 4. 성격유형과 6변수(PitchDV_dv제외)
Figure 4. Personality Types and 6 parameters (except PitchDV dv)

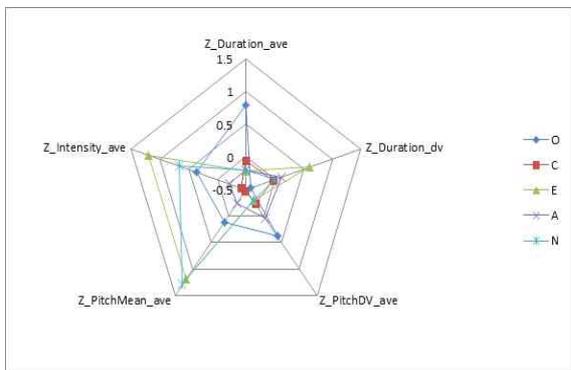


그림 5. 성격유형과 5변수(PitchMean_dv제외)
Figure 5. Personality Types and 5 parameters(except PitchMean dv)

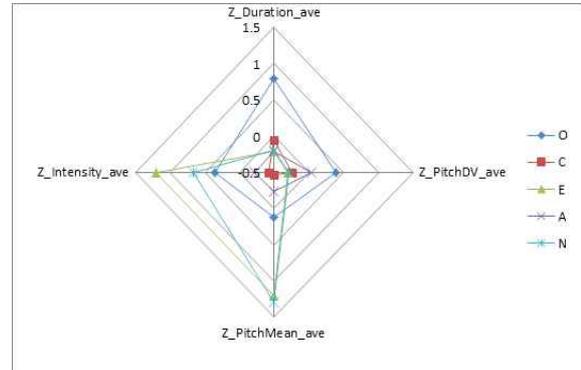


그림 6. 성격유형과 4변수(Duration_dv제외)
Figure 6. Personality Types and 4 parameters(except Duration dv)

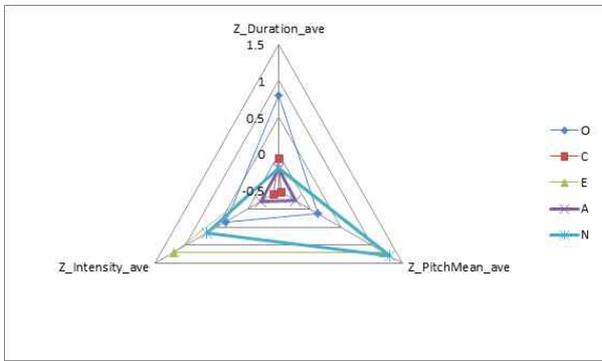


그림 7. 성격유형과 3변수(PitchDV_ave제외)

Figure 7. Personality Types and 3 parameters(except PitchDV_ave)

4.2.3. 성격유형과 4음향변수 간 상관관계

<표 5>는 성격유형별 4변수의 평균값 z-score와 <그림 8> 그래프를 통해 4변수에 따른 변화를 좀 더 자세히 살펴보고 성격유형과의 상관관계를 비교하였다.

발화 속도를 나타내는 길이 값을 보면 O > C > N > A > E 순으로 개방적이고 성실한 성격은 비교적 차분하고 또렷하게 말하고, 동조적이고 외향적 성격은 빨리 말한다고 볼 수 있다.

역양변화를 보여주는 변화폭에서는 O > A > C > N > E 순으로 개방적이고 동조적인 성격이 성실하고, 불안정적이고, 외향적인 성격보다 역양 변화폭이 크다.

발화자의 목소리 높이, 기본주파수에 해당하는 피치평균은 N > E > O > A > C 순으로 불안정하고, 외향적인 성격이 높게 말하고 개방성, 동조성, 성실성 순으로 목소리가 낮게 나타났다. 화자의 목소리 크기를 보여주는 세기 평균은 E > N > O > A > C 순으로 외향적이고 불안정한 성격은 크게 말하고, 개방적이고, 동조적이며, 성실한 성격은 목소리가 낮아졌다. 위의 <그림 3>에서 그래프를 살펴보면 성격유형에 따라 4 음향변수가 뚜렷이 다르게 실현된 것을 볼 수 있다.

O type으로 분류된 5명의 피 실험자들은 성격 유형 상 새로운 것에 대한 호기심이 많고 개방적이며 감정을 잘 표현하는 그룹으로 볼 수 있다. 이들의 발화에서는 시간길이가 평균값보다 길게 나타났고 피치 변화폭과 세기는 평균치에 가까웠으며 피치평균은 낮았다. C type으로 분류된 5명의 피 실험자들은 성격

표 5. 성격유형과 4변수 평균값 z-score
Table 5. Personality Types vs. z-score of 4 Parameters

Type	Z_Duration_ave	Z_PitchDV_ave	Z_PitchMean_ave	Z_Intensity_ave
O	0.791	0.380	-0.119	0.362
C	-0.055	-0.242	-0.468	-0.411
E	-0.218	-0.299	1.193	1.200
A	-0.205	0.033	-0.237	-0.210
N	-0.194	-0.287	1.292	0.662

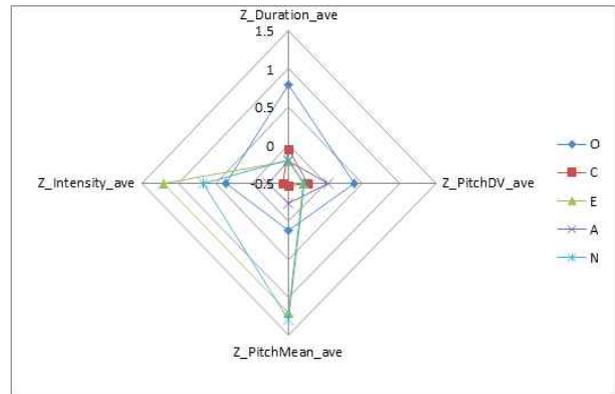


그림 8. 성격유형과 4변수 평균값 z-score

Figure 8. Personality Types vs. z-score of 4 Parameters

유형이 책임감 있고 성취지향적인 특성을 보이는 집단으로 발화의 시간길이와 피치 변화폭이 다른 두 변수보다는 높게 나타났지만 다른 성격유형에 비해 전반적으로 약하게 나타났다. E type으로 분류된 피 실험자는 1명이어서 객관성이 다소 떨어질 수는 있겠으나 성격유형은 외향적으로 사교적이고 활동적인 특성을 보였다. 시간길이를 짧게 피치나 세기는 강하게 실현되었다. A type으로 분류된 14명의 피 실험자들은 동조적이며 솔직하고, 겸손한 특성을 가진 집단으로 전반적으로는 수치가 낮지만 피치 변화폭이 비교적 큰 편이고 시간길이, 피치, 세기는 비슷하다. N type은 3명으로 성격상 민감하고 예민하여 정서적으로 불안정한 성향을 나타내는 반면 자의식이 강하고 적대적인 특성을 보이기도 한다. 시간길이는 짧은데 피치변화 폭은

표 6. 말소리와 성격 이미지

Table 6. Speech Sound vs. Personality Impression

음향 변수	말소리	성격 형 (크기 순: 좌→우)					이미지
Z_Duration_ave	발화 속도	O	C	N	A	E	천천히 말을 하면 성실하고 다양한 경험에 마음을 여는 개방적인 이미지를 주는 반면 빠르게 말하면 불안하거나 동조를 잘 하거나 외향적인 경향을 보임
Z_PitchDV_ave	역양 변화	O	A	C	N	E	역양의 고저변화가 크면 개방적이고 동조를 잘 하는 인상이고, 단조로우면 불안정하거나 외향적인 경향을 보임
Z_PitchMean_ave	음성 높이	N	E	O	A	C	목소리가 높으면 불안정하거나 외향적이며 낮으면 동조를 잘 하거나 성실하고 차분한 경향을 보임
Z_Intensity_ave	음성 크기	E	N	O	A	C	목소리가 크면 외향적이거나 불안정하며 작으면 동조를 잘 하거나 성실하고 차분한 경향을 보임

작다. 위 <표 6>에서는 실험분석을 통해 드러난 음향변수와 성격유형의 관계를 정리하였다. 사람들이 일상에서 인식하는 말소리와 이미지의 관계와 어느 정도는 일치 도를 보였고, 음성 특성 가운데 길이, 세기, 피치와 피치 고저 차이에서 성격유형 별로 뚜렷한 차이가 있었으므로 큰 틀에서는 수용할 만한 의미 있는 결과를 얻었다. 소리의 음향특성 측정값은 절대 값이지만 인식과정에서는 상대 값으로 지각되기 때문에 상대적인 순서로 분류하는 것이 더 의미가 있다고 보았다.

5. 논의 및 종합

지금까지 행한 실험 1, 2, 3의 종합결과가 가장 최적화 되려면 실험 2, 3의 일치도가 실험 1에서도 검증 확인되는 것이다. 그러나 세 학기에 걸쳐 진행된 본 연구에서는 아쉽게도 실험 1의 피실험자는 실험 2, 3에서 재차 동원이 불가능한 상황이었다. 처음에는 화자의 ‘말소리 이미지’를 기술하려는 목적이었지만, 연구 결과에 대한 객관성 확보를 위해 ‘말소리 특성과 이미지 간의 관계’를 도출하는 실험 2, 3이 진행되었기 때문이다. 선행연구를 참고할 수 없는 처음 시도하는 연구였기 때문에 실험디자인을 수정하는 과정에서 아쉬운 점이 드러났고 이를 보완하였으나 완전한 대안이 되지는 못한 것 같다. 그러나 실험 1의 이미지유형조사는 수신자가 느끼는 발신자의 인상을 정리한 결과이므로 타인의 시각이 담겨있는 반면 심리검사를 통한 성격유형조사인 실험 3은 본인 스스로 기술한 자료를 토대로 분석한 것이어서 타인에게 비친 이미지가 아닌 본인의 시각이 담겨있다. 따라서 실험 1과 실험 3의 이미지유형이 상당한 일치를 보이는 것은 말소리 특성(실험 2)과 심리검사 성격유형(실험 3)간의 관계를 말소리와 성격 이미지의 관계로 변환 가능하게 해주는 중요한 핵심 사안이 된다. 한 가지 더 집고 넘어갈 사항은 실험 1의 설문에서는 발표자를 보면서 시각과 청각이 동시에 작동한 결과이므로 말소리만 듣고 평가할 때에 결과가 다르게 나타날 수도 있다. 이 때 개인 별 결과는 다르게 나올 수 있었으나 본 실험에서는 개인 평가용으로 사용하지 않고 전체 안에서 이미지유형을 도출한 것임을 밝혀둔다.

정리하자면, 본 연구는 말소리 특성과 성격유형 간 관계에서 안정적으로 작동되는 기본 틀을 파악하기 위해 복잡하고 상세한 음향특성을 단순화하려는 것이 목적이었다. 현실적으로 부득이한 여건 때문에 피실험자 확보에 어려움이 있어 분석용 녹음자료가 충분치 않았고, 특정 집단의 여성화자만을 대상으로 삼은 점 등이 분석결과에 대한 해석을 제한적으로 수용할 수밖에 없었지만 처음에 기대했던 바대로 후속 연구의 가능성은 확인할 수 있었다. 복합적인 음성특성 가운데서 기본 매개변수를 도출한 본 연구가 단순함에서 복잡함으로 기술을 확장하는 컴퓨터 프로그래밍 기법처럼 점진적인 연구로 확장되어 보다 섬세한 기술이 가능해지기를 희망하며 대용량 코퍼스 실험에 앞서 미리 실현 가능성을 타진해 보았다는 데 주된 의미를 두고자 한다.

6. 결론

목소리의 어떤 물리적 요인이 화자의 성격에 관한 정보를 전달해주고 있는지 예측 가능한 단서를 도출해내기 위해 이미지 분류(실험 1), 말소리 녹음 분석(실험 2), 성격유형검사(실험 3)를 함께 진행하였고, 분석의 초점도 예측 가능성과 그 단서에 맞추었다. 한정된 수의 피실험자와 자료체를 대상으로 분석한 결과지만 발화 속도(Duration), 억양 변화(Pitch Difference Value), 음성 높이(Pitch Mean), 음성 크기(Intensity), 네 가지 단서에 따라 비교적 뚜렷하게 5성격유형 - 경험에 대한 개방성(O), 성실성(C), 동조성(A), 불안정성(N), 외향성(E)-으로 구분되었다. 또한 그 상관관계가 인간의 본능적인 이미지 판단과도 유사하다는 결론에 이르렀다. 여러 가지 복합적인 요소들의 영향에도 불구하고 4가지 음향변수로 각 개인에 대한 이미지와 성격을 판단할 수 있다면 역방향으로 음성합성에서 음향변수를 조작하여 말소리에 이미지를 부여할 수 있다. 따라서 본 논문은 음향변수의 수치 조작을 통해 성격유형을 변화시킬 수 있는 가능성을 보여주었다는 점에서도 의미가 있다고 본다. 앞으로 피 실험 집단을 확대해서 연구를 진행한다면 보다 신뢰도 높은 정교한 결과를 얻을 수 있을 것으로 판단되어 이를 후속 연구과제로 삼고자 한다.

참고문헌

- [1] Lee, E.-J., & Sundar, S. (2009). Human-computer interaction. *The handbook of communication science* (2nd edition) (pp. 507-517). Thousand Oaks, CA: Sage Publications.
- [2] Moon, Y., & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, 23(6), 657.
- [3] Nass, C., & Lee, K. (2001). Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction. *Journal of Experimental Psychology*, 7(3), 172 - 173.
- [4] Brown, B., Strong, W., & Rencher, A. (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America*, 55, 313-318.
- [5] Scherer, K., & Giles, H. (1979). *Social markers in speech*. Cambridge: Cambridge University Press.
- [6] Ekman, P., Friesen, W., O’Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgements of personality and affect. *Journal of Personality and Social Psychology*, 38, 270-277.
- [7] Zuckerman, M., Amidon, M., Bishop, S., & Pomerantz, S. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Personality*, 43, 347-357.
- [8] Burns, K., & Beier, E. (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of*

Communications, 23, 118-130.

- [9] Berman, H., Schulman, A., & Marwit, S. (1976). Comparison of multidimensional decoding of affect from audio, video and audiovideo recordings. *Sociometry*, 39, 83-89.
- [10] Krauss, R., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgements of another's affect. *Journal of Personality and Social Psychology*, 40, 312-320.
- [11] McCrae, R., & Costa Jr., P. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research*, 24, 107-124.
- [12] McCrae, R., & Costa Jr., P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36, 587-596.
- [13] Goldberg, L. (1981). Language and individual differences : The search for universals in personality lexicons. *Review of Personality and Social Psychology*, 2, 141-165.
- [14] Mohammadi, G., Vinciarelli, A., Mortillaro, M. (2010). The voice of personality: mapping nonverbal vocal behavior into trait attributions. *Proceedings of ACM Multimedia Workshop on Social Signal Processing* (p. 19).
- [15] Allport, G. (1961). *Pattern and growth in personality*. Oxford, England: Holt, Rinehart & Winston.

• **이은영 (Lee, Eunyung)**

경북대학교 불어불문학과
대구광역시 북구 대학로 80
Tel: 053-950-5165 Fax: 053-950-6159
Email: eylee@knu.ac.kr
관심분야: 음성학, 음운론
현재 경북대학교 불어불문학과 교수

• **유혜옥 (Yuh, Heaok)** 교신저자

경북대학교 불어불문학과
대구광역시 북구 대학로 80
Tel: 010-6306-4392
Email: hoakyuh@knu.ac.kr
관심분야: 음성학, 음운론
현재 경북대학교 불어불문학과 강사