

감성 평가를 이용한 듣기 좋은 음성 합성음에 대한 연구*

Evaluation of Synthetic Voice which is Agreeable to the Ear Using Sensibility Ergonomics Method*

박용국**, 김재국**, 전용웅**, 조 암**

ABSTRACT

As the method of providing information is getting multimedia, the synthetic voice is used in not only CTI(Computer Telephony Integration), information service for the blind, but also applications on internet. But properties of synthetic voice, such as speech rate, pitch, timbre and so on, are not adjusted to customers' preference but providers' preference.

In order to consider customers' preference, this study proposed four subjective factors of voice through the evaluation of voice using the method of sensibility ergonomics. And the relation synthetic voice to be agreeable to the ear with emotional images was formulated as a fuzzy model.

Consequently, this study proposed the speech rate and pitch of synthetic voice which is agreeable to the ear.

Keyword: synthetic voice, subjective factors, sensibility ergonomics, speech rate, pitch

* 본 논문은 동국대학교 논문게재연구비 지원으로 이루어졌음

** 동국대학교 산업공학과

주 소 : 100-715 서울시 중구 필동 3가 26

전 화 : 02-2260-3376

E-mail : buakman@chollian.net

1. 서론

산업이 발달하고 정보화가 가속됨에 따라 인간이 생활에 필요한 정확한 정보를 빠르고 쉽게 취득하여 활용할 수 있도록 해주는 정보의 멀티 미디어화가 점차 요구되고 있다(이구형, 1998). 이러한 추세에 따라, 음성 합성, 음성 인식 등 음성을 통한 Human Interface 기술에 대한 관심이 점점 커지고 있다.

음성 합성기술은 18세기 후반기에 켐펠른(Wolfgang von Kempelen)이 처음으로 나무와 가죽으로 된 합성기를 만든 이후, 전자 음향 이론을 바탕으로 발전하였다(Peter b. Denes /Elliot N. Pinson, 1993). 디지털 컴퓨터 과학의 발전에 따라, 음성 합성 기술은 음성 발생 과정과 텍스트 과정(text process)도 모델화된 인간의 음성과 유사한 TTS(Text To Speech)까지 많은 기술적 발전이 있었다(Klatt D., 1987). TTS란 언어학, 음성학, 음향학적 지식을 이용하여 일반적인 텍스트를 음성으로 전환시켜주는 기법을 말한다(Gordon E. Pelton, 1990). 이러한 음성 합성 기술들을 통하여 정보를 제공하면 장소에 큰 제약 없이 쉽게 정보를 전달할 수 있으며 종이가 필요 없는 등의 많은 이점들이 있다.

음성 합성음은 CTI(Computer Telephony Integration)와 시각 장애자를 위한 정보제공 수단 뿐만 아니라 다양한 분야에 활용되고 있다. 음성 합성음은 인터넷이 발전하면서 음성으로 정보를 제공하는 인터넷 웹페이지 등에서 응용되고 있으며 이동전화 인터넷부분에서는 사용자가 음성 명령어로 인터넷 서핑을 하

고 원하는 정보를 음성으로 청취할 수 있는 기술 등에 응용되고 있다(J. H. Page/A. P. Breen, 1996; 최성순, 2000).

이렇게 다양한 분야에서 응용되는 음성 합성음에 대한 연구는 일반적으로 인간의 음성과 가까운 자연스러운 운율 부가를 위한 고품질 음성합성 시스템 쪽으로 연구가 이루어졌다. 이러한 연구들은 합성음의 명료도(intelligibility)와 자연성(naturalness)에 관계되어 있으며 합성된 음성의 전체적인 성질에 관한 연구는 많지 않다(권철홍외, 1998). 또한 정보 제공 매체에서 제공되는 음성 합성음들은 정보를 전달해 주는 속도(speech rate)나 음성높이(pitch), 음성색깔(timbre) 등 음성의 성질을 사용자들의 감성에 맞추기보다는 정보 제공자의 임의대로 각각 다른 성질의 합성음으로 제공되고 있다.

본 연구에서는 음성에 관련된 감성 이미지를 추출하고 감성 평가 실험을 통하여 음성의 감성적 구성 요소를 정의하였으며 일반적인 사용자들이 정보를 청취할 때 비교적 만족스러운 음성 합성음에 대하여 연구하였다. 본 연구의 목적을 요약하면 다음과 같다. 첫째, 음성의 감성적 구성 요소를 도출하고 둘째, 음성의 속성 및 '듣기 좋은 음성 합성음'과 감성 이미지와의 관계를 파악하여 셋째, '듣기 좋은 음성 합성음'의 속도와 기본주파수 제시를 목적으로 하였다.

2. 연구방법

본 연구에서 실행한 실험은 음성에 대한 감

성 평가 실험과 '듣기 좋은 음성 합성음'에 대한 감성 평가 실험, 2가지로 나눌 수 있다.

첫째로 음성에 대한 감성 평가 실험은 음성의 감성적 구성 요소를 도출해 내고 음성의 속성과 감성 어휘 즉 감성 이미지와의 관계를 파악하기 위한 실험이다. 둘째로 '듣기 좋은 음성 합성음'에 대한 감성 평가 실험은 정보를 듣기 좋은 음성 합성음과 감성 이미지와의 관계를 파악하기 위한 실험이다.

두 가지 실험의 결과를 이용하여 '듣기 좋은 음성 합성음'의 속도와 기본주파수를 도출하는 것이 본 연구의 결론이다. 다음 그림 1은 본 연구의 flow chart이다.

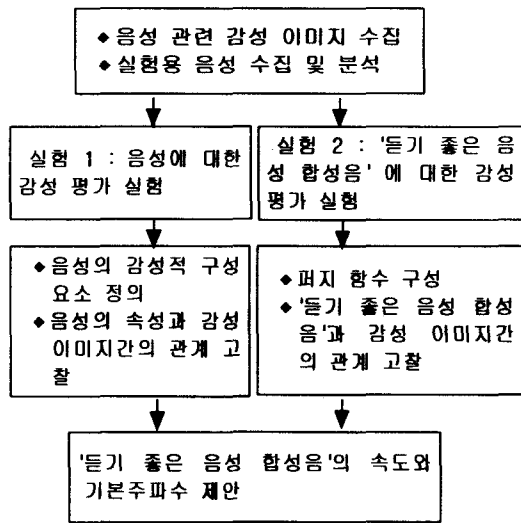


그림 1. 연구의 flow chart

2.1 음성에 대한 감성 평가 실험

2.1.1 음성에 관련된 감성 어휘 추출

음성의 감성 평가를 위한 음성과 관련된 감

성 어휘를 추출하기 위하여 음(音) 및 음악(音樂)과 관련된 29개의 형용사 쌍(손진훈, 1998)과 한국어 형용사 사전(1991)에서 음성과 관련 있는 형용사 340개를 일차적으로 선별하였다. 선별된 형용사에 대해 2명의 패널들이 판단하여 일상 생활에서 사용 빈도가 낮거나 의미가 매우 흡사한 형용사들을 제외시키고 형용사들을 반대말로 쌍을 만들어 2차적으로 형용사 쌍을 추출하였다.

2차적으로 선별된 형용사들을 다시 다른 5명의 패널들이 평가하여 최종적으로 총 76개, 38쌍의 음성과 관련되는 감성 어휘를 선별하였다[표 1].

표 1. 음성에 관련된 감성 어휘 쌍

굵은	-	가느다란	안정적인	-	불안정적인
강한	-	약한	자연적인	-	인공적인
공손한	-	불손한	친숙한	-	낯선
기쁜	-	슬픈	평범한	-	개성적인
좋은	-	나쁜	남성적인	-	여성적인
무딘	-	날카로운	부드러운	-	딱딱한
낮은	-	높은	따뜻한	-	차가운
다정한	-	무뚝뚝한	또렷한	-	흐릿한
맑은	-	흐린	정적인	-	역동적인
무거운	-	가벼운	침착한	-	흥분된
정확한	-	부정확한	세련된	-	촌스러운
유쾌한	-	불쾌한	똑똑한	-	멍청한
친절한	-	불친절한	웅장한	-	여린
명랑한	-	우울한	어린	-	늙은
느린	-	빠른	시원한	-	답답한
선량한	-	불량한	나직한	-	앙칼진
순한	-	사나운	느긋한	-	급한
조용한	-	시끄러운	편한	-	불편한
커다란	-	작은	대담한	-	소심한

2.1.2 실험용 음성 수집

실험에 사용한 음성은 여성음성 10종류, 남성음성 10종류를 사용하였으며 음성내용은 "이따가 다시 걸어 주시겠어요?"로 설정하였

다. 음성의 수집은 다음 2가지 방법을 사용하였다.

첫 번째 방법으로 TTS software 인 '거원 음성 마법사(Ver 1.0)'를 이용하여 합성음의 속도, 높이, 효과 조절을 통해 다른 성질을 지닌 남녀 음성 각각 6종류씩 합성하였다. 두 번째 방법으로 한국 표준과학 연구원의 G7 감성공학 2단계(1995-1998)까지 수행된 연구의 결과로 만들어진 음성 DB에서 각각 다른 감정으로 발성한 여자 음성 4종류, 남자 음성 4종류를 수집하였다.

수집된 음성은 음향 분석 소프트웨어인 'praat'와 'cool edit 2000'을 이용하여 음성의 크기(loudness)평균을 동일하게 하였으며 음성의 다른 파라미터인 속도와 기본주파수 등도 분석하였다.

2.1.3 감성 평가 실험

남성 음성과 여성음성은 음색, 음의 높이 등 음성의 전반적인 성질에서 많은 차이점이 있다. 본 연구에서는 이러한 많은 차이점을 고려하여 남녀 음성을 각각 독립적으로 실험하여 연구하였다. 피실험자는 먼저 청력에 이상이 있는 사람은 제외 시켰으며 여자 음성 감성 평가 실험에서 20대 중심의 남자 36명, 여자 26명 총 62명을 대상으로 하였고, 남자 음성 감성 평가 실험에서는 20대 중심의 남자 46명, 여자 18명으로 총 64명으로 구성되었다.

수집된 20종류의 평가는 추출된 총 38쌍의 음성에 대한 감성 형용사를 카테고리 한 7점 척도 SD(Semantic Differential Method)법에 의하여 이루어졌다.

음성 합성음은 일반적으로 전화나 컴퓨터 스피커를 통하여 청취자들에게 전달되고 있다. 본 연구에서는 실험 시 음성을 들려주기 위하여 일반적인 전달 매체인 컴퓨터(pentium III 680)와 컴퓨터 전용 스피커를 사용하였다. 그리고 음성의 크기를 청취 위치별로 정확히 측정하기 위하여 음향 측정기(Range : 45 ~ 130 dB, Accuracy: +/-0.7dB)를 사용하였으며 음성의 크기는 음성 시작부터 끝까지의 평균으로 측정하였다.

실험장소는 강의실을 이용하였고 소음을 최대한 줄인 후 실험하였으며, 실험 시 피실험자가 충분히 인지하고 평가 할 수 있도록 음성을 반복하여 들려주었다. 다음 그림 2는 실험을 그림으로 표현한 것이다.

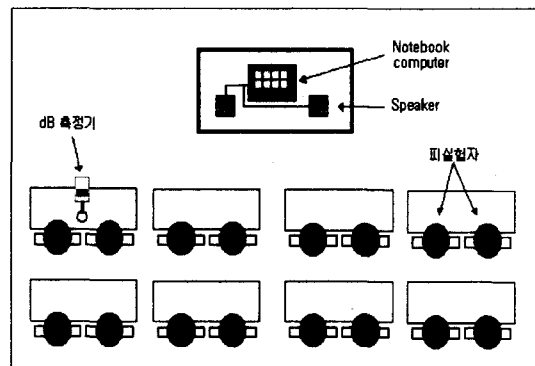


그림 2. 실험 장면

2.2 "듣기 좋은 음성 합성음" 에 대한 감성 평가 실험

2.2.1 '듣기 좋음' 감성의 세분화

본 실험에서는 '듣기 좋음'이라는 감성을 연

구하기 위하여 정보를 들을 때 인식하는 정도에 따라 '강한 인식으로 들을 때 듣기 좋음'과 비교적 '약한 인식으로 들을 때 듣기 좋음'으로 나누어 연구하였다.

강한 인식이란 청취자가 주관적인 판단 하에 중요하다고 생각되는 정보를 들을 때 인식 정도를 말한다. 중요한 정보란 매우 주관적인 것이나 일반적으로 통신매체에서 제공하는 정보 중에서 사용자가 원하는 정보 및 뉴스, 안내 등의 정보를 말한다. 약한 인식이란 청취자에게 크게 중요하지 않은 가벼운 정보를 들을 때를 말한다. 가벼운 정보는 청취자가 주관적인 판단 하에 완전히 이해하지 않아도 되는 정보를 말하며 일반적으로 광고 혹은 엔터테인먼트(Entertainment)등에 쓰이는 정보를 말한다.

2.2.2 '듣기 좋은 음성 합성음'에 대한 감성 평가 실험

'듣기 좋은 음성 합성음'과 감성 이미지와의 관계를 알아보기 위하여 음성 감성 평가 실험에서 실시한 피실험자들을 대상으로 평가 실험을 실시하였다. 실험 전 먼저 '강한 인식', '약한 인식'에 대한 설명을 자세히 하여 피실험자들이 충분히 이해하도록 하였다.

'듣기 좋은 음성 합성음'에 관한 평가 실험은 음성 감성 평가 실험처럼 실제 음성을 듣고 평가하는 것이 아니라 피실험자들이 생각하는 음성의 느낌을 평가하는 것이기 때문에 비슷한 감성 이미지의 미묘한 차이를 구별하기 어렵다. 따라서 본 실험에서는 38쌍의 감성어휘 중 음성 감성 평가 실험 분석 결과로

상관계수가 0.7이상 되는 어휘를 통합하여 총 28쌍의 감성 이미지들로 최적화된 평정 척도를 구성하였다.

또한 평정 척도 외에 각각에 감성 이미지에 '중요도'를 추가하여 체크하게 하였다. '중요도'란 감성 이미지가 듣기 좋은 음성 합성음에 대해 느끼는 감성에서 차지하는 중요한 정도를 5점 척도로 나타낸 것이다. 피실험자들은 '강한 인식으로 듣기 좋은 음성 합성음', '약한 인식으로 듣기 좋은 음성 합성음'에 대하여 각 감성 이미지별로 평정 점수와 중요도 점수를 체크하도록 하였다.

3. 음성 감성 평가 실험 결과 및 고찰

3.1 음성의 감성적 구성 요소

3.1.1 요인 분석

음성 감성 평가 실험 결과에서 감성 이미지를 변수로 각 변수들간의 의미공간을 파악하고 주요 변수 군을 추출하기 위하여 요인 분석(Factor Analysis)을 하였다. 요인추출방법으로는 Kaiser normalization과 함께 주성분 분석을 사용하였고 verymax방법으로 요인 회전시켰다.

분석 결과 전체 요인 중 고유치가 1이상인 4개의 지배적인 요인이 추출되어 총 변량의 67.31%를 설명하였다. 다음 표 2는 요인분석의 결과를 요약한 것이다.

표 2. 요인 분석의 결과

	요 인			
	1	2	3	4
굵은 - 가느다란	-.720	-.121	-.103	.432
우거운 - 가벼운	-.795	-.224	-.090	.185
평범한 - 개성적인	-.161	.233	.681	-.164
높은 - 낮은	.082	.804	.313	-.033
나직한 - 앙칼진	.857	-.160	-.026	.046
낮은 - 높은	.874	.042	.038	-.052
친숙한 - 낯설은	.029	.295	.777	-.192
느긋한 - 급한	.857	-.160	.027	.005
느린 - 빠른	.856	.019	.174	.017
시원한 - 답답한	.569	.245	.562	.048
부드러운 - 딱딱한	-.104	.365	.389	-.285
똑똑한 - 멍청한	.123	.211	.673	.069
우인 - 날카로운	.845	-.120	.031	.014
다정함 - 무뎠직한	.332	.694	.294	-.208
정확한 - 부정확한	.042	.241	.692	.164
선량한 - 불량한	-.206	.783	.163	-.187
공손한 - 불손한	-.340	.705	.195	-.089
안정적인 - 불안정적인	-.379	.422	.568	.095
친절한 - 불친절한	-.121	.846	.246	-.086
유쾌한 - 불쾌한	.135	.806	.301	-.047
편한 - 불편한	-.150	.603	.518	-.037
순한 - 사나운	-.441	.664	.150	-.260
대담한 - 소심한	.124	.079	.223	.590
기쁜 - 슬픈	.505	.644	.135	.144
강한 - 약한	-.064	-.204	-.047	.733
어린 - 늙은	-.567	-.099	-.146	.422
웅장한 - 여린	-.315	-.120	-.035	.710
남성적인 - 여성적인	-.528	-.118	-.039	.432
명랑한 - 우울한	.566	.652	.169	.064
자연적인 - 인공적인	.128	.055	.806	-.204
커다란 - 작은	.168	-.089	-.188	.617
정적인 - 역동적인	.725	-.033	.052	.142
조용한 - 시끄러운	.630	-.152	-.210	.207
따뜻한 - 차가운	-.287	.625	.383	-.158
세련된 - 촌스러운	.073	.213	.680	-.059
맑은 - 흐린	.522	.590	.324	.004
도정한 - 흐릿한	.191	.233	.642	.156
침착한 - 흥분된	-.751	.284	.190	.010

표 2의 분석 결과, 각 요인별 소속 감성 이미지들을 고찰해 보면 요인 1은 '낮은 - 높은', '느린 - 빠른'과 같이 음성의 높낮이와 속도에 관련된 감성 이미지를 포함하였고, 요인 2는 '친절한 - 불친절한', '기쁜 - 슬픈'과 같이 화자(話者)의 정서나 감정에 관련된 감성 이미지들을 포함하고 있었다. 요인 3은 '정확한 - 부정확한', '자연적인 - 인공적인'과 같이 발음 및 억양 등의 정확성 및 자연성에 관한 어휘들을 포함하였고, 요인 4는 '강한 - 약한', '커다란 - 작은'과 같이 음성의 강도 및 크기에 관련된 어휘들을 포함하는 것으로 나타났다.

3.1.2 음성의 감성적 구성 요소에 관한 고찰

본 연구에서는 요인분석 결과로 나온 4가지 요인에서 소속 감성 이미지들을 고찰하여 각 요인에 이름을 붙이고 이를 이용하여 음성의 감성적 구성 요소를 제시하였다(그림 3).

Factor 1은 음성의 전체적인 속도와 높이 및 음질에 관련되어 음성의 전반적인 특징과 보이스 폰트(voice font)의 특성을 나타내는 요소로 설명할 수 있으며 이름을 '개인 특성 요소'로 정의하였다.

Factor 2는 음성에 포함되어 있는 화자의 정서 및 감정에 대한 요소로 억양이나 강세의 변화 등으로 나타내어질 수 있는 요소이며 '정서 표현 요소'로 정의하였다.

Factor 3은 명료도 및 자연성에 관련되었으며 지금까지 음성 합성 분야에서 중점적으로 연구되고 있는 요소로 '정확성 요소'라고 정의하였다.

Factor 4는 음의 크기, 즉 음량 및 강도 변화 등에 관련된 요소이며 '크기, 강약 요소'로 정의하였다.

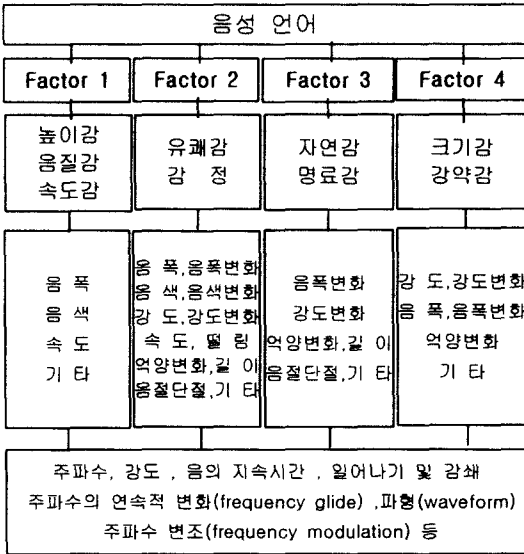


그림 3. 음성언어의 감성적 구성요소

3.3 피실험자의 속성과 음성의 속성이 감성 이미지에 미치는 영향

3.3.1 실험에 사용한 음성의 속성

음성의 속성을 모두 물리량으로 표현하는 것은 발음하는 내용에 따라 차이가 나고 또한 수많은 파라미터들이 서로 복잡하게 관계되어 있어 거의 불가능하다. 따라서 본 연구에서는 실험에 사용한 음성의 특징을 구별하기 위하여 음성 속성을 연구 대상인 속도와 높이(기본주파수)를 포함하여 5가지 아이টে姆으로 크게 합하여 나누고, 각각의 수준을 정하여 통계 분석 하였다.

① 속 도 : 분당 청취되는 음절수로 정보제공 속도를 말하며 단위는 syllable/min을 사용.

② 기본주파수 : 전체적인 음성의 높이(pitch)를 파악하기 위하여 음성의 높이 척도로 일반적으로 많이 사용하는 기본주파수의 평균을 사용하였다. 단위는 Hz를 사용하였으며 발연 내용에 많은 영향을 받는다.

③ 음성색 : 스펙트럼의 형태 등에서 변하는 음성을 구분 짓는 전체적인 색을 말하며 실험에서 사용한 음성색은 3종류로 남성적인-여성적인 이미지 점수에 따라 수준을 1, 2, 3으로 정함.

④ 합성종류 : 음성 합성에서 자연성에 비할 수 있으며 억양, 발음 처리방법 등으로 나누어짐. 실험에 사용한 음성은 크게 2종류, 음성 합성음과 사람의 음성이며 '자연적인 - 인공적인' 이미지 점수에 따라 1, 2로 수준을 정함.

⑤ 운율변조 : 음성을 청취할 때 같은 목소리, 전체적인 속도와 기본주파수가 같아도 화자의 발연 의도나 정서, 감정에 따라서 크게 다른 느낌을 받을 수 있음. 이러한 효과는 부분적인 운율을 변조시키면서 일어날 수 있으며 본 연구에서는 말하는 감정변화에 따라 1~4로 수준을 정함.

일반적으로 음성 합성음을 듣는 수단인 컴퓨터나 개인용 단말기 등은 음의 크기 조절이 쉽게 되며 사용자는 자신의 환경에 맞는 소리로 보통 조절하여 듣는다. 이러한 이유로 본 연구에서는 음성의 가장 중요한 성질 중의 하나인 음의 크기를 한정하여 연구하였으며 실험용 음성들의 평균크기를 동일하게 만들어 실험하였다. 실험 음성은 인간의 일반적인 대화 크기인 70dB이상으로 청취하도록 하였으

며 74dB~76dB 크기로 청취한 피실험자가 가장 많은 것으로 나타났다.

다음 표 3, 표 4는 실험에 사용한 음성의 속성 즉 아이템과 카테고리를 나열한 표이다.

표 3. 실험에 사용한 여성 음성의 속성

음성	속도	기본주파수	음성색	합성종류	운율변조
1	311	207	2	1	1
2	237	208	2	1	1
3	446	228	2	1	1
4	311	99	2	1	1
5	311	427	2	1	1
6	311	132	1	1	1
7	420	258	3	2	1
8	390	305	3	2	2
9	489	311	3	2	3
10	311	225	3	2	4

표 4. 실험에 사용한 남성 음성의 속성

음성	속도	기본주파수	음성색	합성종류	운율변조
11	298	128	2	1	1
12	230	126	2	1	1
13	413	133	2	1	1
14	295	91	2	1	1
15	293	245	2	1	1
16	291	275	1	1	1
17	392	153	3	2	1
18	472	229	3	2	2
19	425	135	3	2	3
20	461	251	3	2	4

* 속도 : syllable/min

기본주파수 : 평균 pitch (Hz)

음성색 : 여성스러움에 따라 (남성적 : 1, 보통 : 2, 여성적 : 3)

합성종류 : 자연적인 정도에 따라 (부자연적 : 1, 자연적 : 2)

운율변조 : 감정의 정도에 따라 (보통 : 1, 기쁨 : 2, 화남 : 3, 슬픔 : 4)

* 음성의 크기 : 72~81dB

* 여성 음성 평가 실험, 남성 음성 평가 실험은 독립적으로 연구

3.3.2 피실험자의 속성이 감성 이미지에 미치는 영향

본 연구에서 나눈 피실험자의 속성은 나이, 성별, 청취위치별 소리크기 3가지이다. 피실험자의 속성이 감성 이미지에 미치는 영향을 고찰하기 위하여 속성과 감성 이미지 점수와의 편상관 분석을 하였다.

분석 결과 여성음성 감성 평가 실험, 남성 음성 감성 평가 실험 모두 편상관 계수는 낮아 피실험자의 속성이 감성 이미지에 크게 영향을 주지 않는 것으로 나타났다. 이러한 결과는 피실험자의 나이를 20대로 한정시키고 청취위치에 따른 소리 크기는 범위가 10dB를 넘지 않아 나이 및 청취위치가 결과에 끼치는 영향이 매우 작은 것으로 분석되었다. 피실험자의 속성중 성별과 감성 이미지와의 편상관 관계에서는 유의한 이미지가 여성음성 2개, 남성 음성 3개로 나타났으나 마찬가지로 상관 계수는 매우 낮아 피실험자의 성별이 결과에 미치는 영향이 크지 않은 것으로 분석되었다. 다음 표 5은 피실험자의 속성과 감성 이미지와의 편상관 분석에서 P-value가 0.05이하인 감성 이미지들이다.

표 5. 피실험자의 속성과 감성 이미지와의 편상관계수

여성 음성						남성 음성					
나 이		성 별		소 리 크 기		나 이		성 별		소 리 크 기	
이미지	상관계수	이미지	상관계수	이미지	상관계수	이미지	상관계수	이미지	상관계수	이미지	상관계수
평범한	-.0958	웅장한	-.1315	부드러운	.0856	굵 은	.0866	양활진	.0883	굵 은	-.1085
좋은	-0.872	커다란	-.1306	강 한	-.1043	웅장한	.0847	시원한	.1054	다정한	.0914
친숙한	-.0901			또렷한	-.1373			정확한	.0977	강 한	-.1774
공손한	-.1122									커다란	-.1367
안정적	-.1152										
인	-.0970										
친절한	-.0810										
편 한	-.0955										
자연적											
인											

* 이미지 이름은 감성 이미지쌍 중 하나만 표기

* 통제 변수 : 음성 속성(속도, 기본주파수, 음성색, 합성종류, 운율변조), 피실험자 속성(나이, 성별, 소리크기)

3.3.3 음성의 속성이 감성 이미지에 미치는 영향

실험에 사용한 음성의 속성과 감성 이미지와의 관계를 파악하기 위하여 편상관 분석을 하였다. 다음 표 6은 결과의 예로 여성 음성에서 음성의 감성적 구성요소 Factor1인 개인 특성 요소 소속 감성 이미지들과 음성의 속성과의 편상관 계수를 나타낸 것이다.

분석 결과, 개인 특성 요소 소속 이미지에서는 음성의 속도와 기본주파수에서 편상관계수가 높게 나타났으며 정서 표현 요소 소속 이미지들은 운율변조 부분 편상관 계수가 비교적 높게 나타났다. 그리고 정확성 요소 소속 이미지들은 합성종류와 운율변조에서 높은 상관관계가 나타났다. 크기, 강약 요소 소속

이미지들은 모든 속성에서 대부분의 상관 관계가 적게 나왔는데 이러한 결과는 본 연구에서 제외된 음성의 속성 중 크기와 관련이 있기 때문이라고 분석되었다. 또한 음성의 속도와 대표적으로 관련된 감성 이미지 '느린 - 빠른'과 음성의 속성과의 상관계수를 고찰한 결과 사람이 주관적으로 느끼는 속도감은 음성의 속도뿐만 아니라 다른 속성과도 밀접한 관계를 가지며 특히 음성의 높이와 많은 관련이 있다는 것을 알 수 있다. 마찬가지로 주관적으로 느끼는 높이감도 음성의 높이 뿐만 아니라 속도 등 다른 속성과도 관련되어 있는 것으로 분석되었다.

표 6. 음성의 속성과 감성 이미지와의 편상관 계수의 예

이미지	속도 편상관계수	기본 주파수 편상관계수	음성색 편상관계수	합성종류 편상관계수	운율변조 편상관계수
굵은 - 가느다란	.1061	.5811	.2743	-.1378	-.2115
무거운 - 가벼운	.3143	.5842	.3290	-.2130	-.3515
나직한 - 앙칼진	.4793	.6050	.2274	-.3454	.0504
낮은 - 높은	.3885	.6254	.2836	-.2562	-.1449
느긋한 - 급한	.6107	.5513	.1713	-.3103	.1406
느린 - 빠른	.6896	.5762	.1426	-.1810	-.2389
시원한 - 답답한	-.2597	-.2265	-.1173	-.2200	.4675
무딘 - 날카로운	.4984	.6099	.2095	-.3089	.0148
어린 - 늙은	-.0466	-.4040	-.2231	.0818	.0034
남성적인 - 여성적인	.0262	.5523	.3339	.0069	-.0861
정적인 - 역동적인	.3669	.3858	.1116	-.1022	-.0654
조용한 - 시끄러운	.4036	.4903	.1333	-.3333	.1499
침착한 - 흥분된	.4819	.5115	.0966	-.3612	.4058

* 여성음성, 개인 특성 요소 소속 감성 이미지

실험에 사용된 음성의 속성이 감성 이미지 들 점수에 기여하는 정도를 파악하기 위하여 중회귀 분석을 하였다. 분석 시 꼭 필요한 속 성을 선택하기 위해 단계별 선택법(stepwise selection method)을 사용하였다. 회귀 분석

결과로 얻어진 회귀식은 뒤의 5장에서 '듣기 좋은 음성 합성음'에 대한 실험 결과와 연결 하여 '듣기 좋은 음성 합성음'의 속도와 기본 주파수를 제시하는 데에 이용하였다. 표 7은 회귀 분석의 예이다.

표 7. '느린 - 빠른' 종속변수 회귀분석

음성종류		비표준화계수		표준화계수	R ²
		B	표준오차	베타	
여성음성	상수	-4.234	.396		0.716
	속도	0.01656	0.001	0.650	
	기본 주파수	0.007493	0	0.390	
	음성색1	0.429	0.207	-0.125	
	음성색2	1.106	0.152	0.292	
	운율변조2	1.000	0.167	0.158	
남성음성	운율변조4	-0.788	0.197	0.067	0.650
	상수	-8.303	0.601		
	속도	0.02174	0.001	0.916	
	기본 주파수	0.0172	0.001	-0.032	
	음성색2	3.237	0.222	0.849	
	합성종류1	-2.280	0.271	0.565	
	운율변조1	2.108	0.238	0.507	
	운율변조3	-0.202	0.237	-0.586	
운율변조4	0.123	0.203	0.019		

4. '듣기 좋은 음성 합성음' 감성 평가 실험 결과 및 고찰

다음 그림 4, 그림 5는 정보의 2종류에 따른 '듣기 좋은 음성 합성음'에 대한 감성 이미지 평정점수의 평균과 중요도 점수의 평균을 나타낸 그림이다.

4.1 '듣기 좋은 음성 합성음' 감성 평가 실험 결과

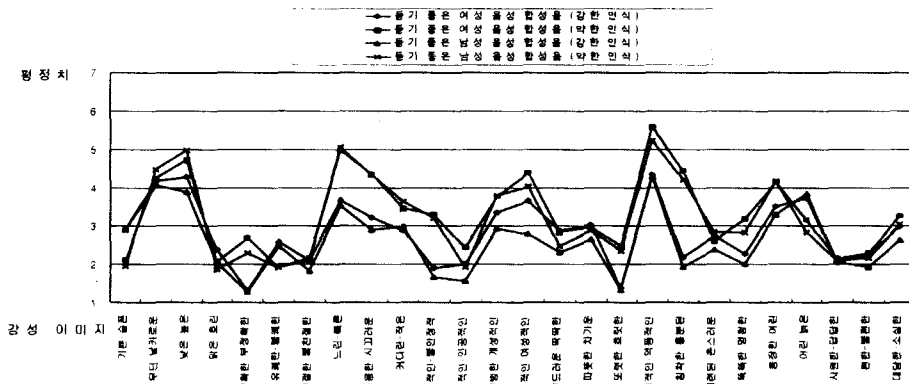


그림 4. 실험 결과 (평정 점수 평균)

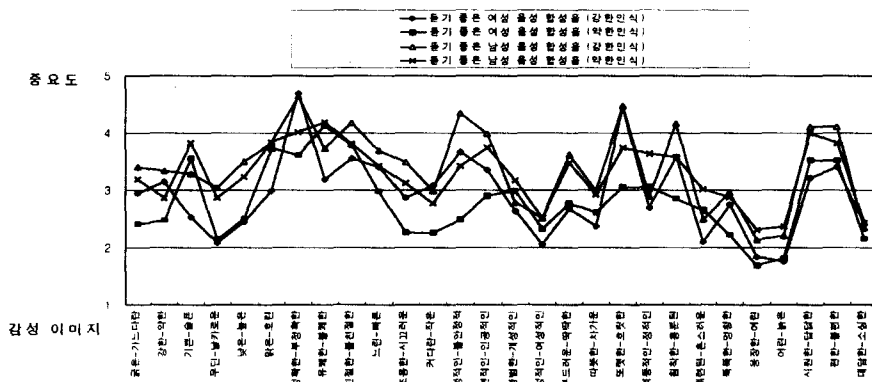


그림 5. 실험 결과 (중요도 점수 평균)

4.2 "듣기 좋은 음성 합성음"에 대한 감성 이미지 평정 점수의 퍼지화

불확실한 정보를 제어하는데 유용한 기법인 퍼지이론을 응용하여 실험 결과에서 나온 이미지 점수를 퍼지화 시켰다(H. J. Zimmermann, 1991).

퍼지 함수는 객관성을 높이고 편차를 고려하기 위하여 각 이미지 별로 평균, 표준편차와 신뢰구간을 이용한 삼각 퍼지함수로 구현하였으며 감성 이미지별 중요도를 가중치로 응용하였다. 다음 식 1~4는 본 연구에서 구현한 퍼지 소속 함수이다.

$$t \leq \mu_i - w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}} \quad \dots\dots\dots (식 1)$$

$$f_i(t) = 0$$

$$\mu_i - w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}} < t < \mu_i$$

$$f_i(t) = \frac{t - \mu_i}{w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}}} + 1 \quad \dots\dots\dots (식 2)$$

$$\mu_i < t < \mu_i + w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}}$$

$$f_i(t) = \frac{\mu_i - t}{w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}}} + 1 \quad \dots\dots\dots (식 3)$$

$$t \geq \mu_i + w_i \cdot Z_{\alpha/2} \cdot \frac{\delta_i}{\sqrt{n}} \quad \dots\dots\dots (식 4)$$

$$f_i(t) = 0$$

- * μ_i = i 이미지의 평균 점수
- * δ_i = i 이미지의 표준 편차
- * $Z_{\alpha/2}$ = 유의 수준 α 에 대한 정규분포값
- * n = 표본 수, * w_i = i 이미지의 가중치

다음 그림 6은 예로 강한 인식으로 듣기 좋은 여성 음성 합성음의 '느린 - 빠른' 감성 이미지 점수를 퍼지 함수로 구현한 것이다.

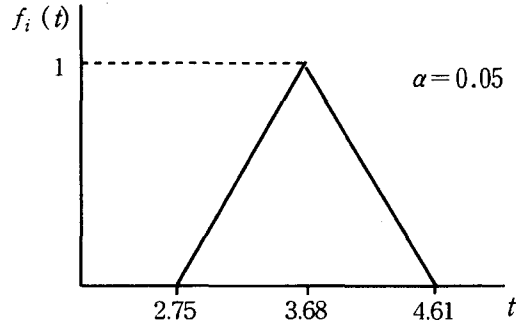


그림 6. '느린 - 빠른' 이미지 점수의 퍼지 소속 함수 (여성 음성 합성음, 강한 인식)

5. '듣기 좋은 음성 합성음'의 속도와 기본주파수

'듣기 좋은 음성 합성음'의 속도와 기본주파수를 제시하기 위하여 본 연구에서는 앞 절에서 제시한 음성의 감성적 구성 요소 중 전체적인 속도와 기본주파수가 다른 요소에 비해 많이 관계되어 있는 Factor 1인 개인 특성 요소의 소속 감성 이미지들과 음성의 속성과의 중회귀식을 '듣기 좋은 음성 합성음'의 퍼지 함수들과 연결하여 계산하였다. 또한 계산의 편의를 위해 임의의 음성 합성음 속성으로 시뮬레이션 하여 각 이미지 퍼지 함수 값의 합이 가장 높은 속성을 선택하게 하는 프로그램

을 작성하였다. 다음 그림 7은 작성한 시뮬레이션 프로그램이다.

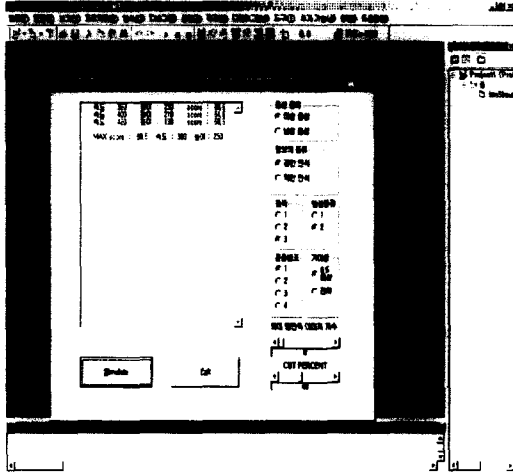


그림 7. 시뮬레이션 프로그램

이 프로그램 알고리즘은 퍼지 함수값들이 가장 높게 나오게 하는 음성의 속도와 기본주파수를 구하는 것이 기본 목적이다. 즉 속도와 기본주파수를 제외한 나머지 속성은 조건으로 하여 임의의 음성의 속성으로 회귀식에 대입한 결과로 나타난 개인 특성 요소의 소속 감성 이미지 퍼지 함수 값들의 합이 가장 높게 나오게 하는 음성 합성음의 속도와 기본주파수를 구하는 프로그램이다. 그 밖의 프로그램의 옵션은 다음과 같다.

- 여성음성, 남성음성 선택 가능
- 음성색, 합성종류, 운율변조 속성 선택 가능
- 만족 회귀식 개수 선택 가능
- 각 회귀식 결과의 퍼지 함수 소속 값 제한 가능
- 조건(기여율)에 따른 회귀식 선택 가능

6. 결론 및 추후연구

본 연구에서는 음성에 대한 감성 평가를 실시한 후 요인 분석으로 음성에 관련된 감성 이미지를 요인별로 나누고 음성의 감성적 구성 요소를 개인 특성 요소, 정서 표현 요소, 정확성 요소, 크기 및 강약 요소 4가지로 분석하였다. 또한 '듣기 좋은 음성 합성음'과 감성 이미지들과의 관계를 연구하고 감성 이미지 점수를 퍼지 함수로 구현하였다.

음성의 속성과 감성 이미지들과의 회귀식들을 '듣기 좋은 음성 합성음'의 개인 특성 요소 소속 이미지 퍼지 함수와 연결하여 '듣기 좋은 음성 합성음'에 근접한 속도와 기본주파수를 구하였다. 본 연구에서 분류한 5가지 속성을 임의로 회귀식에 넣어 퍼지 함수 값의 합이 가장 큰 속성을 선택하였다. 표 9는 본 연구에서 제시한 '듣기 좋은 음성 합성음'의 속도와 기본주파수를 나타낸 것이며 다른 속성은 여성 음성 합성음, 남성 음성 합성음 모두 음성색은 3, 합성 종류는 2, 운율 변조는 1 일 때 퍼지 함수 값이 가장 높게 나왔다. 다른 조건은 여성 음성 합성음일 경우, 회귀식 기여율이 0.5이상 되는 소속 이미지 퍼지 함수 값이 모두 0.4이상일 때이며, 남성 음성 합성음인 경우 여성 음성 합성음에 비하여 회귀식에 의하여 예측된 퍼지 함수값들이 적게 나왔으며 표 8에서 제시한 경우는 전체 10개 회귀식 중 6개의 퍼지 함수 값이 20 이상으로 나온 경우이다.

표 8. '듣기 좋은 음성 합성음'의 속도와 기본주파수

	인식의 종류	속 도	기본 주파수
여성 음성 합성음	강한 인식	375~385 syllable/min	225~235 Hz
	약한 인식	430~440 syllable/min	250~260 Hz
남성 음성 합성음	강한 인식	395~405 syllable/min	125~135 Hz
	약한 인식	400~410 syllable/min	165~175 Hz

* 음성 : "이따가 다시 걸어 주시겠어요?"

* 크기(dB) : 70~80dB

* 옵션 - 음성색 : 3, 자연성 : 2, 운율 변조 : 1

추후 연구 과제로서 본 연구에서 피실험자를 20대로 한정하여 연구하지 못했던 연령대 별 음성 합성음에 대한 감성에 관련된 연구와 본 연구에서 제시한 결과로 합성한 음성을 실제적으로 검증하는 연구도 필요하다. 그리고 음성 합성음의 속성 중 속도와 기본주파수 뿐만 아니라 크기 등의 많은 음성 합성음의 속성 파라미터들이 연구되어 '듣기 좋은 음성 합성음'의 표준을 만드는 연구도 요구되어진다.

참고 문헌

- 권철홍, 최영익, 이금주, 심갑중, "명료도에서 사람 목소리로-TTS에 관하여", 한국 음향학회 학술대회 논문집, 17권, 1호, 1998.
- 박준하, 한국어 형용사 사전, 계명문화사, 1991.
- 손진훈, "청각 감성 측정 기술 및 DB개발", 충남 대학교, 1998.
- 이구형, "사회 및 산업환경의 변화와 감성과 학", 한국 감성과학회지, 1권, 1호, pp. 13-17, 1998.
- 최성순, 이윤근, "음성 기술 응용 서비스", 마이크로 소프트웨어, 9월호, 소프트뱅크 미디어, pp. 242-280, 2000.
- Gordon E. Pelton, Voice Processing, McGraw-Hill, pp. 67-82, 1990.
- H. J. Zimmermann, Fuzzy set theory--and its applications, Kluwer Academic Publishers, 1991.
- J. H. Page, A. P. Breen, "The Laureate text-to-speech system", BT Journal, Vol 14, No 1, 1996.
- Klatt D., "Review of text to speech conversion for English", J Acoust Soc Am, 82, No3, pp. 737-739, 1987.
- Peter B. Denes, Elliot N. Pinson, The Speech Chain: The Physics and Biology

of Spoken Language, 2nd edition,
W.H.Freeman, 1993.

Rossing. T. D., The Science of sound,
Addison Wesley, 1990.

저자 소개

◆ 박용국

2000년 동국대학교 산업공학과 학사
2002년 동국대학교 산업공학과 석사

◆ 김재국

2000년 동국대학교 산업공학과 석사
2002년 동국대학교 산업공학과 박사과정
수료

◆ 전용웅

1998년 동국대학교 산업공학과 학사
2000년 동국대학교 산업공학과 석사
현 동국대학교 산업공학과 박사과정

◆ 조 알

日本 早稻田大學 工業經營學科 공학석사
동대학원 인간과학 박사
전 동국대학교 정보산업대학 학장
동국대학교 정보산업대학 산업시스템공학
부 교수

논문접수일 (Date Received): 2001/8/28

논문게재승인일(Date Accepted): 2002/4/10