

단어빈도와 단어규칙성 효과에 기초한 합성음 평가

남기춘(고려대), 최원일(공군사관학교), 김충명(고려대),
최양규(춘해대), 김종진(ETRI)

<차 례>

- | | |
|-------------------|-------------------|
| 1. 서론 | 3.1 실험 방법 |
| 2. 실험 1 | 3.1.1. 피험자 |
| 2.1. 실험 방법 | 3.1.2. 실험재료 및 설계 |
| 2.1.1. 피험자 | 3.1.3. 실험절차 |
| 2.1.2. 실험재료 및 설계 | 3.2. 실험2의 결과 및 논의 |
| 2.1.3. 실험절차 | 4. 종합논의 |
| 2.2. 실험1의 결과 및 논의 | |
| 3. 실험 2 | |

<Abstract>

The Text-to-Speech System Assessment Based on Word Frequency and Word Regularity Effects

Kichun Nam, Wonil Choi, Choong-Myung Kim, Yang-Gyu Choi, Jongjin Kim

In the present study, the intelligibility of the synthesized speech sounds was evaluated by using the psycholinguistic and fMRI techniques. In order to see the difference in recognizing words between the natural and synthesized speech sounds, word regularity and word frequency were varied. The results of Experiment1 and Experiment2 showed that the intelligibility difference of the synthesized speech comes from word regularity. In the case of the synthesized speech, the regular words were recognized slower than the irregular words, and there was smaller activation of the auditory areas in brain for the regular words than for the irregular words.

* Keywords: Text-to-speech system, Word frequency effect, Word regularity effect

1. 서론

합성음을 이용한 응용 시스템이 다양한 분야에서 사용되고 있다. 예를 들어, 전화기 자동응답시스템, 공공장소에서의 음성 안내시스템, TTS(text-to-speech system), 자동차 운행 음성 안내 시스템, e-mail 낭독 시스템 등의 분야에서 사용된다. 음성 합성 시스템을 특정 응용시스템에 적절하게 사용하기 위해서는 응용 분야의 특성 분석과 이런 응용 분야에 적합하도록 음성 합성 시스템을 개발하는 것이 필수적이다. 응용 분야에 따라 음성합성 시스템이 갖추어야 할 요건이 기술적인 측면에서 달라지지만 최종적으로 그 시스템이 적절한지는 사용자의 판단에 의해 결정된다. 이런 측면에서 합성음을 사용자 관점에서 평가하는 것은 매우 중요한 과제이다. 현재 개발되어 있는 음성 합성기는 여러 측면에서 부족하다. 예를 들어, 현재의 시스템은 소음이 있거나 반향 상황에서는 이해하기 쉬운 합성음을 생성해 내지 못하며, 감정(emotion)이나 전하려고 하는 화용적 의미(pragmatic meaning)에 따른 운율 조절(prosody control)의 어려움을 가지며, 또한 특정 화자의 말하는 스타일과 음성의 품질 등을 자유스럽게 조절하는 데에서도 어려움을 가진다. 본 연구는 현재 개발되어 있는 음성 합성기가 이처럼 여러 한계점을 지니고 있어 제한적이지만 현재의 합성기 기술로 생성된 합성음이 구체적으로 인간 언어 정보처리(human language information processing)에서는 자연음과 어떤 측면에서 유사한지 혹은 상이한 지를 조사하기 위해 실시되었다.

기존의 합성음 평가 방법은 주로 설문에 답하는 것과 같은 off-line 과제를 이용한 청취실험이었다[10]. Off-line 청취실험에서는 정보처리과정 중의 합성음 이해 과정을 다루기보다는 모든 정보처리가 종료된 후에 나타난 결과를 평가한다 [10][8][13]. 이런 off-line 청취실험과제는 일반적으로 사용하기 간편하고 기존에 개발된 여러 척도(scale)를 이용할 수 있다는 측면에서 장점을 지닌다. 그러나 off-line 합성음 평가 방법은 합성음을 듣는 동안에 일어나는 인간언어정보처리 특성을 조사할 수 없을 뿐만 아니라, 때로는 합성음 평가자의 주관에 따라 평가 결과가 달라질 수 있다는 단점을 지니고 있다. 본 연구에서는 이와 같은 기존의 off-line 합성음 평가 방법의 한계성을 극복하기 위해 언어심리학(psycholinguistics)에서 흔히 사용하는 어휘 판단 과제(lexical decision task)를 사용하여 합성음으로 생성된 단어에 대한 정보처리 특성을 밝히고자 하였다. 어휘판단과제를 사용한 평가 방법은 어휘를 이해하는 동안 일어나는 정보처리를 반영하는 반응 시간, 실수율 및 특정 변인의 효과 등을 살필 수 있어서 off-line 과제에 비해 우수하다[3].

인간의 언어정보처리 특성을 연구하는 언어심리학 연구에서는 음성, 음운, 어휘, 문장 및 담화 등의 언어학적 단위를 연구의 대상으로 한다. 본 연구에서는 이런 여러 종류의 언어정보 단위 중에서 단어재인(word recognition) 과정에 초점을 두었다. 단어재인은 물리적으로 외부에 제시된 자극을 해독(decoding)하여 내뇌에

저장되어 있는 심성어휘집(mental lexicon)에서 물리적 자극에 해당되는 어휘를 찾아 그 뜻을 이해하는 인간정보처리 과정이다. 단순히 자연음 어휘와 합성음 어휘를 듣고 어휘를 판단하는 속도나 실수율을 조사하는 것만으로는 합성음 단어재인의 특성을 충분히 규명하기에는 미흡한 측면이 있어 단어빈도(word frequency)와 단어규칙성(word regularity) 효과를 중심으로 연구를 수행하였다. 단어빈도효과는 단어가 재인될 때 자주 사용되는 단어일수록 더 빨리 재인되는 현상을 의미한다. 기존의 언어심리학 연구에 따르면, 단어빈도효과는 대뇌(brain)에 저장되어 있는 심성어휘집(mental lexicon)에서 해당되는 어휘에 접근(lexical access)하는 데 영향을 미치는 것으로 알려져 있다[5][9][12]. 또한, 단어규칙성효과(regularity effect)는 문자와 음운간의 규칙을 따르지 않는 불규칙어휘(irregular word)를 이해하는 시간이 규칙어휘(regular word)를 이해할 때 소요되는 시간보다 길어지는 현상을 의미한다 [2][7][15]. 단어규칙성효과도 단어빈도효과처럼 해당 어휘를 심성어휘집에서 찾는 데 관여하는 변인으로 알려져 있다. 국내에서 개발된 음성 합성시스템이 대부분 규칙 기반(rule-based)이기보다는 말뭉치(corpus)를 이용하는 경향이 강해서 단어빈도와 단어규칙성 변인이 민감하게 작용할 것으로 생각된다.

본 연구는 두 종류의 실험 연구로 구성되어 있다. 실험 1은 언어심리학적 연구 방법을 사용하여 음의 종류, 단어빈도, 단어 규칙성이 음성 단어재인에 어떤 영향을 미치는지를 어휘판단과제를 이용해 조사하였다. 실험 2는, 실험 1에서 사용된 변인에 따라 자연음과 합성음의 청각 단어재인 시의 대뇌 활성화 영역이 어떻게 차이 나는지를 조사하였다.

2. 실험 1

실험 1은 두 가지 목적을 가지고 있다. 첫째는 단어를 재인할 때, 합성음의 음질의 수준에 따라 청취 어휘정보처리에서 차이가 있는가를 알아보는 것이다. 둘째는 합성음과 자연음을 들을 때 나타나는 단어재인 과정이 동일한지를 알아보고자 하였다. 실험 1에서 사용된 과제는 기존의 합성기 평가 연구에서는 사용되지 않은 청각 어휘판단과제(auditory lexical decision task)이다. 청각 어휘판단과제에서는 제시된 음성자극이 단어인지 아닌지를 판단하는 시간을 측정한다. 어휘판단 시간은 음성자극이 제시된 순간부터 반응키를 통해 반응할 때까지의 시간 간격이다. 만약 음의 종류에 따라 음질이 달라지고 이것이 단어재인에 영향을 미친다면, 시각적 단어재인이나 청각적 단어재인 연구에서 나타나던 단어빈도효과 [2][4][6][11][14], 단어빈도와 단어규칙성간의 상호작용효과가 합성음과 자연음에서 다른 양상으로 나타날 것이라 예측할 수 있다.

2.1. 실험 방법

2.1.1. 피험자

서울 시내에 소재한 대학에 재학 중이고, 청력에 문제가 없는 대학생 21명이 실험에 참여하였다.

2.1.2. 실험재료 및 설계

실험에 이용된 자극은 빈도에 따라 고빈도와 저빈도[1], 음운 변화에 근거한 규칙과 불규칙으로 나뉘어져 각 조건당 30개씩, 모두 120개의 어휘가 사용되었다. 그리고 이 어휘들은 음 종류 조건에 따라 크게 3가지 종류로 만들어졌다. 첫 번째 것은, DAT 녹음기로 중년의 여성이 녹음한 어휘, 두 번째 것은, 대량의 코퍼스를 통해 확인된, 고품질의 합성기를 이용하여 생성된 어휘, 그리고 마지막으로 저품질의 합성기를 이용하여 생성된 어휘였다. 어휘들은 모두 2음절의 한국어였고, “물가”(고빈도 불규칙어휘), “방침”(고빈도 규칙어휘), “물증”(저빈도 불규칙어휘), “변심”(저빈도 규칙어휘)과 같은 자극들이 사용되었다.

이 실험은 3 x 2 x 2요인 피험자 내 설계(within-subject design)를 사용하였다. 첫 번째 독립변인은 3개의 수준을 갖는 음 종류 조건이고, 두 번째 독립 변인은 2개의 수준을 갖는 빈도이며, 세 번째 독립 변인은 역시 2개의 수준을 갖는 규칙성 조건이다. 종속 변인은 각 조건별 평균 반응 시간으로 1/1000초(millisecond) 단위로 측정되었다.

2.1.3. 실험절차

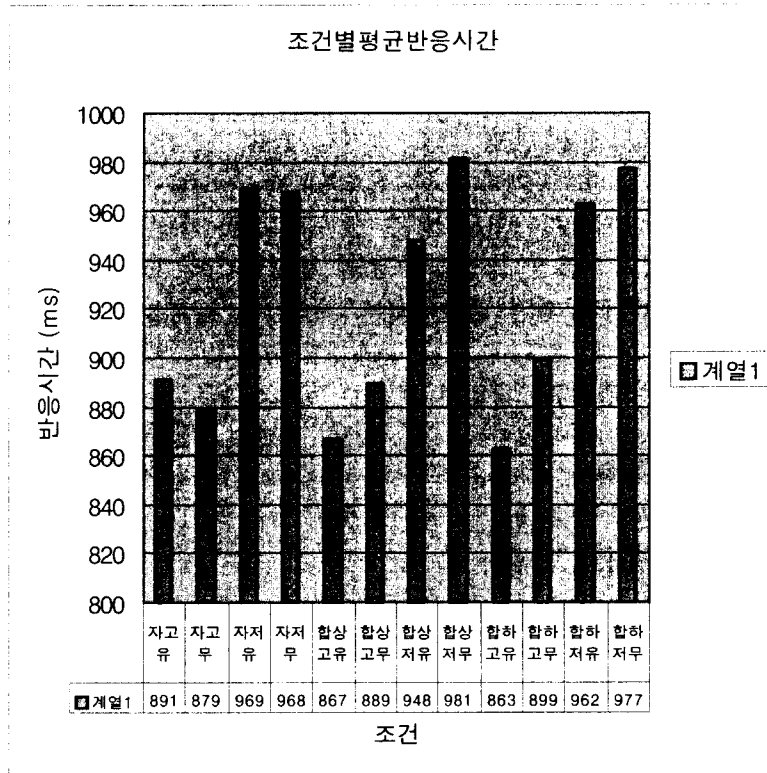
이 실험에 사용된 과제는 어휘판단과제이다. 어휘판단과제란 제시되는 자극이 단어인지 아닌지 최대한 빠르게 판단하는 것이다. 모니터에 “****”과 같은 부호가 800ms동안 제시되고, 500ms의 시간 간격이 있는 후에 피험자에게 소리자극이 제시된다. 그러면 피험자는 그 자극을 듣고 단어인지 아닌지 판단해서 적절한 반응 키를 눌러야 한다. 이러한 일련의 과정이 하나의 시행이고, 이러한 시행들이 240번(120개의 단어와 120개의 비단어 조건에 대한 시행) 반복되었다.

2.2. 결과 및 논의

각 실험조건의 대표 값(central tendency)을 추출하기 위해 피험자 반응시간의 중앙값(median)을 통계치(statistic)로 사용하였고, 각 조건의 중앙값에 대한 평균 반

응시시간을 그림 1에 나타내었다. 평균보다 중앙값을 사용한 이유는 평균은 양극단치(extreme value)에 의해 영향을 지대하게 받는데 비해서 중앙값은 극단치의 영향을 적게 받으면서도 모든 자료의 특성을 반영하고 있기 때문이다.

그림 1에 나타난 조건별 반응시간을 가지고 삼원 피험자 내 변량분석(ANOVA)을 실시하였다. 먼저 음 종류에 대한 주효과(main effect)를 살펴보면 음 종류에 따른 주효과가 나타나지 않은 것을 볼 수 있다[F(2,20)=.311, p=.734]. 그림 1에서도 볼 수 있듯이 자연음, 고품질 합성음, 저품질 합성음간에 차이가 없다. 이는 적어도 어휘를 판단하는 어휘 확인 단계에서는, 자연음과 합성음에서 차이가 나지 않는다는 것을 알 수 있다. 빈도에 대한 주효과와 규칙성에 따른 주효과는 모두 통계적으로 유의미한 결과를 나타내었다[F(1,20)=216.445, p<.05; F(1,20)=8.721, p<.05]. 즉 빈도가 높은 단어일수록 반응시간이 빨랐고, 음 변화가 있는 규칙 단어일수록 반응시간이 빠르게 나타났다. 그리고 음 종류와 규칙성간의 상호작용효과(interaction effect) 역시 통계적으로 유의미하게 나타났다[F(2,40)=6.121, p<.05]. 이 상호작용효과를 해석하기 위해서 각 음 종류에 따른 규칙성효과를 t-test를 통해 검증하였다. 그 결과 자연음 조건에서는 규칙성에 따른 반응 시간의 차이가 유의미하지 않았지만[t(41)=.680, p>.05], 고품질의 합성음과 저품질의 합성음의 경우에는 규칙성에 따른 반응시간의 차이가 통계적으로 유의미하였다[t(41)=-3.257, p<.05, t(41)=-3.172, p<.05]. 그림 1에서도 볼 수 있듯이 결국 자연음 조건에서는 규칙성효과가 나타나지 않은 반면에, 합성음 조건에서는 규칙성효과가 나타났다. 음 변화가 있는 불규칙조건에서 더 빠른 반응 시간을 나타냈다. 그리고 합성기의 품질에 따른 차이는 보이지 않았다. 마지막으로 음 종류, 단어 빈도, 단어 규칙성에 따른 상호작용효과는 유의미하지 않았다. 즉 위에서 보고되었던 음 종류와 단어 규칙성에 따른 상호작용효과가 단어 빈도에 따라 변화하지 않았다. 이런 결과는 자연음을 대상으로 이루어진 연구 결과[12]와는 상이한 것인데, 이런 현상은 합성음 어휘에 대한 단어재인 과정이 다르기 때문에 나타난 것으로 해석할 수 있다. 합성음의 품질과 단어 규칙성과 이원 상호작용이 있었던 것과 이 상호작용이 단어 빈도와 상호작용하지 않은 실험 결과는 결국 합성음 어휘 정보처리가 자연음 어휘의 정보처리와는 섬세한 어휘정보처리에서 합성음과 자연음에 차이가 있다는 것을 암시한다고 생각할 수 있겠다.



<그림 1> 실험 1의 조건별 반응시간

자고유: 자연음, 고빈도, 불규칙

자고무: 자연음, 고빈도, 규칙

자저유: 자연음, 저빈도, 불규칙

자저무: 자연음, 저빈도, 규칙

합상고유: 합성음/상품질, 고빈도, 불규칙

합상고무: 합성음/상품질, 고빈도, 규칙

합상저유: 합성음/상품질, 저빈도, 불규칙

합상저무: 합성음/상품질, 저빈도, 규칙

합하고유: 합성음/하품질, 고빈도, 불규칙

합하고무: 합성음/하품질, 고빈도, 규칙

합하저유: 합성음/하품질, 저빈도, 불규칙

합하저무: 합성음/하품질, 저빈도, 규칙

3. 실험 2

실험 2는 합성음을 이해하려고 노력할 때와 자연음을 이해하려고 노력할 때에 활성화되는 대뇌 영역을 조사하여 자연음과 합성음의 차이를 밝혀 보기 위해 실시되었다. 대뇌 영역의 활성화를 조사하기 위해 fMRI(functional magnetic resonance imaging technique)기법사용하여 실험을 시행하였다. 만일에 자연음과 합성음간에 질적인 차이가 있다면 어휘 정보처리를 다루는 뇌 영역의 활성화에서 차이가 있을 것으로 예측된다. 반면에 자연음과 합성음이 근본적인 정보처리는 유사하지만 세부적인 어휘정보처리 부분에서의 차이라면 두 자극에 대해 유사한 영역이 활성화되지만 활성화의 정도에서 차이가 있을 것으로 예측된다.

3.1. 방법

3.1.1 피험자

피험자로는 신경학적 이상이 없고 오른손잡이인 20대 남자 2명과 여자 2명이었다.

3.1.2. 실험 재료 및 설계

실험 1에서 사용하였던 실험 재료를 사용하였다. 실험에서 조작한 변인은 음의 종류(자연음 혹은 합성음), 단어 빈도(저빈도 혹은 고빈도), 단어규칙성(규칙어휘 혹은 불규칙 단어)으로 2x2x2 설계였다. 실험 1에서는 합성기의 품질에 따라 두 종류의 합성음을 사용하였지만, 실험 2에서는 합성음의 품질이 더 좋은 합성음만을 사용하였다. 그 이유는 실험의 1의 결과에서 합성기의 품질에 따른 각 변인의 효과가 유의미한 차이를 나타내지 않았기 때문이다.

3.1.3. 실험절차

MRI 시스템은 한국과학기술원 fMRI센터에 있는 ISOL 3.0T forte를 사용하였다. 30초간의 활성화기(activation phase)와 휴식기(resting phase)로 구성된 블록 디자인을 사용하였다. 자연음과 합성음은 활성화기에 제시되었다. 휴식기에는 어떤 활동도 하지 않았다. 활성화기 동안에 헤드폰으로 제시된 단어를 듣는 동안 기능적 자기공명 영상을 획득하였다. 피험자는 활성화기에 제시되는 단어를 듣고 그 단어의 의미가 무엇인지를 이해하라고 지시를 받았으며, 이외의 다른 특별한 과제를 수행하지는 않았다.

3.1.4. 영상 획득 및 데이터 분석

피험자가 헤드폰을 통해서 자극 단어들을 듣는 동안 EPI-BOLD(Echo Planner Imaging-Blood Oxygen Level Dependent)기법을 사용하여 영상을 획득하였다. 영상은 매 3초마다 전교련(anterior commissure; AC)과 후교련(posterior commissure; PC)을 연장한 AC - PC line을 기준으로 뇌의 아래에서부터 위로 횡단면 20절편(slice)이 획득되었다. 본 실험의 총 체적(volume)영상은 244개를 얻었으며, 첫 4개의 영상은 신호 보정을 위한 모조 영상(dummy image)이었다.

데이터 분석은 SPM 99(Statistical Parametric Mapping 99)를 이용하였다. 데이터 처리의 과정은, 총 영상 중 첫 번째 영상을 기준으로 움직임이나 기타 호흡, 맥박 등으로 인한 움직임을 보정하였고, 각 피험자의 뇌를 표준화하기 위해 SPM 99와 함께 제공되는 template image의 공간에서 정의된 영상을 이용하였으며, 이후 소음 제거 및 Gaussian field를 만들기 위해 영상 다듬기를 수행하였다. 데이터 분석은 실험 시 자극을 주었던 패턴을 대상으로 일반 선형 회귀분석을 수행하였고, 유의 수준 $p < 0.00001$ 을 기준으로 t-검정을 수행하였다. 각 개별 피험자의 영상을 얻은 후, 공통된 활성화 영역을 찾기 위해 모든 피험자의 영상을 통계적으로 합산하였다. 최종활성화 영상화 지도를 표준화된 3D T1 영상에 중첩하여 기능적 자기공명 영상을 얻었다.

3.2. 실험 2의 결과 및 논의

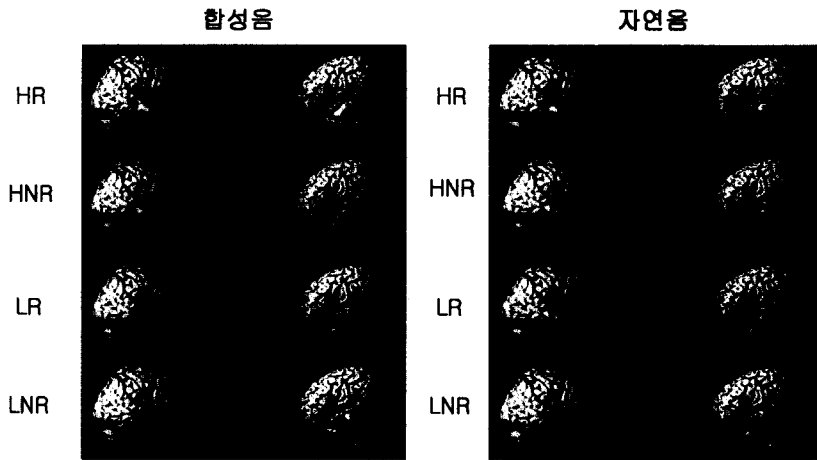
자연음과 합성음을 이해하는 동안 활성화되는 대뇌 영역 사진이 그림 2에 제시되어 있다. 또한 정확한 대뇌의 영역을 알기 위해서 Talairach Daemon Software를 이용하였다. 자연음과 합성음에 따른 대뇌 활성화 영역은 아래의 표1과 표2를 통해 제시하였다.

<표 1> 합성음 청취 시 각 조건에 따른 활성화 영역 및 좌표 값

조건	반구	활성화영역	좌표값(x,y,z)	Z score
합성음 HR	L	Middle Temporal Gyrus	-58, -6, -10	9.15
	L	Middle Temporal Gyrus	-52, 2, -18	8.97
	L	Middle Temporal Gyrus	-60, -12, -4	8.47
	R	Middle Temporal Gyrus	62, -16, -10	8.08
	R	Middle Temporal Gyrus	60, -2, -12	7.65
	R	Superior Temporal Gyrus	62, -22, 4	7.64
	L	Precentral Gyrus	-50, 0, 38	4.61
합성음 HNR	L	Middle Temporal Gyrus	-60, -6, -12	8.35
	L	Superior Temporal Gyrus	-46, -20, -6	6.7
	L	Superior Temporal Gyrus	-48, -26, 0	6.51
	R	Middle Temporal Gyrus	58, -20, -6	6.8
	R	Superior Temporal Gyrus	62, -24, 0	6.28
	R	Middle Temporal Gyrus	68, -30, -4	5.89
합성음 LR	R	Middle Temporal Gyrus	60, -2, -14	5.61
	R	Middle Temporal Gyrus	56, -14, -10	5.23
	R	Middle Temporal Gyrus	56, 6, -20	4.65
	L	Middle Temporal Gyrus	-60, -8, -6	5.6
	L	Middle Temporal Gyrus	-54, -2, -14	5.52
합성음 LNR	L	Middle Temporal Gyrus	-58, -6, -8	6.95
	L	Middle Temporal Gyrus	-54, -18, -4	6.86
	L	Superior Temporal Gyrus	-64, -26, 0	5.76
	R	Middle Temporal Gyrus	60, -8, -6	6.06
	R	Middle Temporal Gyrus	60, 2, -16	5.85
	R	Middle Temporal Gyrus	68, -18, -4	5
	R	Superior Temporal Gyrus	64, -26, 6	5.34
	R	Superior Frontal Gyrus	6, 46, 44	4.96
	R	Precentral Gyrus	42, -10, 30	4.69
	L	Left Cerebellum	-16, -64, -34	4.59
	R	Middle Temporal Gyrus	66, -32, 0	4.42

<표 2> 자연음 청취 시 각 조건에 따른 활성화 영역 및 좌표 값

조건	반구	활성화영역	좌표값(x,y,z)	Z score
자연음 HR	L	Middle Temporal Gyrus	-60, -8, -6	8.34
	L	Superior Temporal Gyrus	-54, -20, -2	7.46
	L	Middle Temporal Gyrus	-52, 0, -18	7.12
	R	Middle Temporal Gyrus	60, -10, -6	8.03
	R	Middle Temporal Gyrus	58, -2, -14	7.41
	R	Superior Temporal Gyrus	64, -24, 4	6.41
	L	Superior Frontal Gyrus	-10, 40, 46	5
	L	Middle Frontal Gyrus	-32, 38, -12	4.5
	L	Medial Frontal Gyrus	-4, 50, 34	4.33
자연음 HNR	L	Middle Temporal Gyrus	-60, -6, -8	6.21
	L	Superior Temporal Gyrus	-50, -26, 0	5.04
	L	Sub-Gyral	-44, -32, 4	4.79
	R	Middle Temporal Gyrus	58, -10, -6	4.78
	R	Middle Temporal Gyrus	60, -2, -14	4.68
	R	Middle Temporal Gyrus	68, -18, -10	4.41
자연음 LR	R	Middle Temporal Gyrus	56, -12, -6	7.77
	R	Superior Temporal Gyrus	60, -26, 0	7.23
	L	Middle Temporal Gyrus	-56, -8, -10	7.63
	L	Middle Temporal Gyrus	-52, -18, -6	6.99
	L	Sub-Gyral	-48, -36, -4	6.98
	L	Superior Temporal Gyrus	-64, -42, 6	4.83
	L	Left Cerebellum	-12, -40, -42	4.67
	L	Left Cerebellum	-18, -52, -36	4.55
	L	Left Cerebellum	-4, -60, -34	4.53
	L	Inferior Temporal Gyrus	-54, -54, -6	4.57
	R	Right Cerebellum	22, -64, -34	4.54
	R	Right Cerebellum	16, -58, -32	4.48
	L	Left Cerebellum	-4, -64, -20	4.49
	L	Superior Frontal Gyrus	-8, 14, 56	4.46
	R	Inferior Temporal Gyrus	42, -12, -32	4.44
L	Putamen	-14, 6, -8	4.36	
자연음 LNR	L	Middle Temporal Gyrus	-60, -6, -8	6.21
	L	Superior Temporal Gyrus	-50, -26, 0	5.04
	L	Sub-Gyral	-44, -32, 4	4.79
	R	Middle Temporal Gyrus	58, -10, -6	4.78
	R	Middle Temporal Gyrus	60, -2, -14	4.68
	R	Middle Temporal Gyrus	68, -18, -10	4.41



HR: High frequency Regular word (ex: 목표, 공부, 분석)
 HNR: High frequency Non Regular word (ex: 독립/동남/, 불안/부란/, 철학/치락)
 LR: Low frequency Regular word (ex: 공핍, 영단, 경직)
 LNR: Low frequency Non Regular word (ex: 물증/물증/, 직함/지람, 격렬/경멸)

<그림 2> 합성음과 자연음을 들었을 때의 뇌 영상 사진

그림 2에서 볼 수 있듯이 고빈도 단어일 경우에는, 규칙어휘이든 불규칙어휘이든 관계없이 자연음과 합성음간에 차이가 없었다. 자연음이든 합성음이든 1차 및 2차 청각피질에서의 활성화가 있었으며, 또한 중측두피질(middle temporal gyrus)에서도 유의미한 활성화가 있었다. 그러나 저빈도 단어의 경우에는 다른 결과가 나타났다. 자연음의 경우에는 음변화가 없는 규칙어휘의 경우, 고빈도 단어의 활성화와 거의 동일한 활성화가 나타났으나 음변화가 있는 불규칙 단어의 경우에는 활성화 수준이 떨어졌다. 반면에 합성음의 경우에는 저빈도 단어가 고빈도 단어에 비해 전반적인 활성화 수준이 떨어졌으며, 음변화가 있는 단어에 비해 음변화가 없는 단어에서 활성화 수준이 더 떨어졌다.

이러한 연구 결과는 자연음과 합성음에서 활성화된 대뇌 영역의 질적인 차이는 없지만 두 조건 모두에서 활성화된 대뇌 영역의 활성화 정도에서 차이가 난다는 것을 의미한다. 이 결과는 합성음이 자연음과 질적으로 다른 방법으로 이해되기 보다는 전반적인 과정은 유사하나 하위수준의 미세한 정보처리 과정에서 차이가 있다는 것을 의미한다.

fMRI를 이용한 실험 결과는 어느 정도 반응시간 자료와 일치한다. 반응시간 실험 결과에서도 저빈도 단어인 경우에 특히, 자연음과 합성음간의 차이가 있었다. fMRI 결과에서는 저빈도인 경우에 청각 담당 영역의 저조한 활성화로 나타났다. 특히 합성음에서 저빈도이면서 음변화가 없는 규칙어휘인 경우에 청각 영역에서

의 활성화가 낮았는데 이런 결과는 실험 1의 반응 시간 자료에서 규칙어휘인 경우에 더 느린 반응으로 나타나 일맥상통하는 것이다. 즉, 합성음의 저빈도 규칙 단어를 인지하는데 특별히 어려움이 있다는 것을 암시한다.

4. 종합논의

본 연구에서는 자연음의 어휘를 정보처리할 때와 합성음의 어휘를 정보처리할 때 어떤 차이가 있는지를 조사하고, 합성음의 품질에 따른 차이 여부를 알아보기 위해 실시되었다.

실험 결과에 따르면, 어휘 판단 시간에서는 세 종류의 음성에 따른 차이는 나타나지 않았으며, 빈도효과에서도 큰 차이가 없었다. 그러나 단어규칙성효과에서는 자연음 조건과 합성음 조건 간에 질적인 차이가 나타났다. 즉, 자연음 조건에서는 단어 규칙성효과가 나타나지 않았지만, 합성음 조건에서는 단어 규칙성효과가 뚜렷이 나타났다. 또한 fMRI를 이용한 실험에서도 규칙성이 문제가 되었다. 즉, 단어 빈도와 관련해서는 자연음을 들을 때나 합성음을 들을 때나 큰 차이가 없었지만, 단어 규칙성에서는 단어 빈도와 상호작용하여 차이가 있었다. 특히 저빈도 이면서 규칙 단어인 경우에 자연음보다 합성음 조건에서 청각 영역의 활성화가 작았다.

따라서 단어규칙성효과만을 고려하면, 자연음과 합성음 간에는 질적인 차이가 있는 것으로 보인다. 음성 어휘를 듣고 일차적으로 단어인지 아닌지를 판단하는데에는 자연음이나 합성음이나 큰 차이가 없지만, 어휘 의미와 같은 속성 등의 심층적인 정보를 처리하는 데에서는 자연음과 합성음간의 차이가 있는 것이 아닌가 추측된다. 최근에 발표된 Delogu, Conte, and Sementina(1998)의 결과를 보면, 문장 이해 시에 합성음이 자연음에 비해 심적 부담(mental load)이 더 커서 더 어렵고, 합성음을 청취할 때 더 큰 주의(attention)가 요구된다고 한다. 이런 맥락에서 본 연구의 결과를 생각해 볼 때 합성음 어휘 정보처리 시에, 단어규칙성과 같은 어휘의 심층적 특성을 다루기 위해서는 더 큰 인지적 부담(cognitive load)이 요구되는 것으로 해석할 수 있겠다.

흥미로운 현상은 어휘판단시간과 단어규칙성효과에서 고품질 합성음과 저품질 합성음 간에는 차이가 없다는 것이다. 표 1에서 볼 수 있듯이 어휘판단시간이 두 조건 간에 유사하고 단어규칙성효과에서도 유사하다는 것이다. 이런 결과를 놓고 볼 때, 합성음을 어느 정도 식별할 수만 있으면 어휘를 표면적인 수준에서 이해하는 데에는 큰 차이가 없는 것으로 생각된다. 이런 특성은 음성 정보의 애매모호함(ambiguity) 때문에 나타나는 현상이 아닌가 추측된다. 잘 알려진 것처럼 음성 정보는 상당히 애매모호하다. 이런 이유로 때로는 적절한 문맥이 없으면 무슨 단어인

지를 이해하기 힘든 경우가 많다. 본 연구에서는 문맥 속에 어휘가 제시된 것이 아니고 고립 음성어휘(isolated word)만이 제시되었기 때문에 음성 어휘의 애매모호함이 그대로 유지되었을 것으로 예측된다. 이런 애매모호함이 큰 상황에서 일부 정교한 음성 정보가 더 주어지더라도 심층적인 정보를 이해하는 데에는 별 도움이 안 된다면, 고품질 합성음이 저품질 합성음에 비해 더 많이 제공하는 정보가 별 영향력을 가지지 못할 것이다. 문제는 합성음의 분절음 수준보다 더 미세한 음성학적 차이가 합성음의 품질과 관련된 것으로 추측된다.

본 연구 결과를 종합해 볼 때, 합성음은 자연음에 비해 어휘 자체를 이해하는 과정에서는 큰 차이가 없지만, 좀 더 인지적이고 어휘의 규칙성과 같은 상위의 심층적인 어휘 정보처리가 요구되는 과정에는 어려움을 주는 것으로 볼 수 있겠다.

참고문헌

- [1] 서상규, *현대 한국어의 어휘빈도*, 연세대학교 언어정보연구원, 1998.
- [2] 이광오, “한글 글자열의 음독과 음운 규칙”, *한국심리학회지: 실험 및 인지*, 8, 1, pp. 1-24, 1996.
- [3] 윤용욱, 정한민, 이근배, “어휘의미패턴을 이용한 음성인식 오류 검출 및 수정”, *제14회 한글 및 한국어 정보처리 학술대회 논문집* pp. 62-68, 2002.
- [4] 조혜숙, 남기춘, “실어증환자에서 보이는 단어 규칙성 효과”, *언어청각장애연구*, 7, 3, pp. 77-94, 2000.
- [5] D. A. Balota, “Visual word recognition: The journey from features to meaning”, In M. a. Gernsbacher (Eds.), *Handbook of Psycholinguistics*: Academic Press.1994.
- [6] D. Bauer, K. E. Stanovich, “Lexical access and spelling-to-sound regularity effect”, *Memory & Cognition*, 8, pp. 424-432, 1980.
- [7] M. Coltheart, “Lexical access in simple reading tasks”, In G. Underwood (Eds.), *Strategies of information processing*, London/New York, Academy Press, 1978.
- [8] C. Delogu, S. Conte, & C. Sementina, “Cognitive factors in the evaluation of synthetic speech”, *speech communication*, 24, pp. 153-168, 1998.
- [9] K. I. Forster, “Accessing the mental lexicon”, In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam, North-Holland, 1976.
- [10] M. Goldstein, “Classification of methods used for assessment of text-to-speech systems according on the demand placed on the listeners”, *speech communication*, 16, pp. 225-244, 1995.
- [11] P. A. Luce, “A computational analysis of uniqueness points in auditory word recognition”, *Perception and Psychophysics*, 39, pp. 155-158. 1986.
- [12] S. Monsell. “The nature and locus of word frequency effects in reading”, In D. Besner and G. W. Humphreys (Eds.), *Basic processes in reading*. Lawrence Erlbaum Associates. Hillsdale, New Jersey. 1991.
- [13] D. Pisoni, S. Hunnicutt, “Perceptual evaluation of MITalk: The MIT unrestricted

text-to-speech system”, *Proceedings of ICASSP*, 80, 3, 1980.

- [14] H. B. Savin, “Word-frequency effect and errors in the perception of speech”, *Journal of Acoustic Society of America*, 35, pp. 200-206, 1963.
- [15] M. S. Seidenberg, G. S. Waters et al., “When does irregular spelling or pronunciation influence word recognition?”, *Journal of Verbal Learning and Verbal Behavior*, 23, pp. 383-404, 1984.

▶ 남기춘(Kichun Nam)

주소: 136-701 서울시 성북구 안암동 5가 1번지 고려대학교 심리학과
 소속: 고려대학교 심리학과
 전화: 02) 3290-2068
 FAX: 02) 3290-2662
 E-mail: kichun@korea.ac.kr

▶ 최원일(Wonil Choi)

주소: 363-849 충북 청원군 남일면 쌍수리 사서함 335-2 공군사관학교 교수부 인사처 국방학과
 소속: 공군사관학교 국방학과
 전화: 043) 290-6396
 E-mail: cerouno@afa.ac.kr

▶ 김충명(ChoongMyung Kim)

주소: 136-701 서울시 성북구 안암동 5가 1번지 고려대학교 심리학과
 소속: 고려대학교 행동과학연구소
 전화: 02) 3290-2068
 FAX: 02) 3290-2662
 E-mail: czykim@korea.ac.kr

▶ 최양규(Yang-Gyu Choi)

주소: 689-784 울산광역시 울주군 웅촌면 곡천리 산72-10번지 춘해대학교 유아특수치료학과
 소속: 춘해대학교 유아특수치료학과
 전화: 052) 270-0100
 E-mail: ygchoi@choonhae.ac.kr

▶ 김종진(Jongjin Kim)

주소: 305-350 대전광역시 유성구 가정동 161번지
 소속: 한국전자통신연구원(ETRI)
 전화: 042) 860-6114
 E-mail: kimjj@etri.re.kr