

스크린리더를 사용하는 시각장애인의 한국어 합성음 청취속도 연구

A Study of Korean TTS Listening Speed for the Blind Using a Screen Reader

이 희 연¹⁾ · 홍 기 형²⁾

Lee, Heeyeon · Hong, Ki-Hyung

ABSTRACT

The purpose of this study was to evaluate the maximum and optimal listening speed of Korean TTS for the blind. Five blind participants took part in this study. The instruments used in this study were 17 sentence sets (2 sets for an exercise, 10 sets for a repeated test, and 5 sets for a random test), with short meaningful sentences (the same sentences for the repeated test, different sentences for the random test) with 15 differentiated speeds ($Range=0.8-3.6$, $SD=0.2$). Each participant's maximum and quickest listening speeds were calculated by objective recall accuracy (determined by the number of correctly recalled syllables/the total number of syllables in a sentence X 100) and subjective recall accuracy (recall accuracy judged by each participant's subjective evaluation). The results showed that the participants' recall accuracy had a tendency to increase as the TTS speed decreased. Participants' subjective recall accuracy was higher than objective recall accuracy in the repeated tests and vice versa in the random tests. The results also revealed that the participants' sentence familiarity had an influence on their Korean TTS listening speed.

Keywords: Korean TTS, the blind, listening speed, speech synthesizer

1. 서론

시각장애인들은 시각적으로 정보를 인식하고 처리하는데 제한을 가지고 있기 때문에, 부족한 시각을 보완하기 위해서 정안인에 비해 청각이나 촉각이 발달한다고 알려져 있다. 최근에는 컴퓨터와 여러 보조기기가 발달함에 따라서 많은 전맹 시각장애인들이 시각적인 정보를 청각적인 정보로 전환해주는 화면읽기 프로그램인 스크린리더(screen reader)를 사용하여 필요한 정보에 접근하며 다양한 교육적, 직업적, 사회적 활동들에 참여하는 것이 가능하게 되었다. 그러나 실제적으로 스크린리더를 사용하는 시각장애인들이 얼마나 빠르게 합성

음을 인지할 수 있는지, 어느 정도의 속도에서 별도의 노력 없이도 편안하게 합성음을 인지할 수 있는지, 합성음 인식 속도에서 시각장애인들과 정안인들이 차이를 나타내는지에 관한 연구는 별로 없는 실정이다. 따라서 구체적으로 스크린리더를 사용하는 시각장애인들이 얼마나 빠른 배속으로 얼마나 편안하게 합성음을 인지하는지에 관한 연구가 실행될 필요가 있다.

일본의 선행연구[1]에서 Asakawa와 동료들은 시각장애인들이 가장 편안하게 들을 수 있는 일본어 TTS의 최적 청취 속도와 가장 빠르게 인식할 수 있는 최고 청취 속도에 대한 연구를 객관적/주관적 평가 방법을 통해서 실행하였고, 주관적/객관적 평가 모두 동일한 문장을 다른 속도로 제시하는 반복 제시방법과 다른 문장들을 다른 속도로 제시하는 랜덤 제시방법을 사용하여 각각 평가를 진행하였다.

한국어 합성음의 인지속도의 평가방법을 도출하기 위한 선행연구에서[2] 일본의 연구에서 사용한 평가방식을 한국어 합성음 인지속도 평가에 어떻게 적용할 것인지에 관한 논의 및 전문가의 자문이 이루어 졌고, 그 내용들을 토대로 본 연구의

1) 서울대학교, heeyonlee@hotmail.com

2) 성신여자대학교, kihyung.hong@gmail.com, 교신저자

본 연구는 산업통상자원부 및 한국산업기술평가관리원의 기술혁신사업의 일환으로 수행되었습니다.

(지원번호:10036459, QoLT 산업기술기반 지원센터 구축사업).

접수일자: 2013년 8월 7일

수정일자: 2013년 9월 23일

게재결정: 2013년 9월 27일

평가도구 및 평가방법이 마련되었다.

본 연구의 목적은 일본의 선행연구를 한국어 합성음에 적용하여 시각장애인의 한국어 합성음 최고 인지속도와 최적 인지속도를 조사하는 것으로, 한국어 합성음 개발자들에게 시각장애인을 위한 스크린리더용 합성음 개발 시 필요한 배속의 지침 및 범위에 대한 정보를 제공하고, 고속합성음 개발에서 고려해야 하는 시각장애인 사용자들의 요구사항들을 도출하는데 본 연구의 의의가 있다. 일본의 선행연구에서는 합성음의 인식 속도와 문장 친숙도와의 상관관계를 제시하기는 하였지만 구체적으로 어떠한 방법으로 문장 친숙도가 평가되었는지에 관한 정보는 언급되지 않았기 때문에 이에 대한 내용을 보완하고 확장하여 본 연구를 실행하였다. 또한 시각장애인용 한국어 합성음의 배속별 인지속도에 대한 기존의 국내연구가 거의 이루어지지 않아 온 실정을 고려할 때, 이에 대한 첫 시도라는 점에서 본 연구의 의의가 있다. 구체적인 연구문제들은 다음과 같다: (1) 반복평가에서 나타나는 참가자들의 한국어 합성음 인지의 최고 배속과 최적 배속은 무엇인가? (2) 랜덤평가에서 나타나는 참가자들의 한국어 합성음 인지의 최고 배속과 최적 배속은 무엇인가? (3) (1)과 (2)에서 유의미한 차이를 나타내는가? (4) 참가자들의 한국어 합성음 최고/최적 인지속도와 평가에 사용된 문장 친숙도 간에는 어떠한 상관관계가 있는가? (5) 스크린리더용 고속합성음 개발 시 고려해야 하는 참가자들의 요구사항들은 무엇인가?

2. 연구방법

2.1 평가참가자

시각장애인들의 한국어 합성음 인지속도에 대한 평가는 서울 소재의 장애인 관련 기관에서 실행되었다. 컴퓨터 및 스크린리더의 사용에 익숙한 시각장애인 5명이 평가에 참여하였다. 참가자들의 연령은 35세부터 50세까지의 다양한 연령의 참가자들이 평가에 참여하였으며, 총 5명의 참가자들 중에서

1명은 선천적 시각장애인이었고 나머지 4명은 사고나 질병으로 인한 후천적 시각장애인이었다. 참가자들에 대한 자세한 기본 정보는 <표 1>에 제시되어 있다.

2.2 평가도구

본 연구를 실행하기 위한 선행연구[3]에서 다음의 요소들을 고려하여 평가에 사용하기 위한 문장집합들을 선정하였다: (1) 의미가 있는 단어들로 구성된 유의미 문장, (2) 주로 7 어절로 구성된 문장 길이, (3) 차별화된 문장 친숙도.

첫째, 시각장애인을 위한 스크린리더용 고속합성음의 명료도에 대한 선행연구[4]에서는 시각장애인들이 사전 지식을 활용하여 단어나 문장을 예측하는 것을 방지하기 위해서 무의미 단어들을 평가에 사용했으나 참가자들이 청취한 합성음 단어들이 의미가 없고 생소하기 때문에 들리는 대로 정확하게 리콜 하는 것을 주저하는 경향을 보였다. 이러한 제한점을 보완하기 위해서 본 연구에서는 의미전달이 잘 이루어지는 유의미 문장 집합들을 사용하였다. 둘째, 본 연구는 각 참가자들의 리콜 정확도를 이용해서 최고와 최적의 합성음 인지속도를 판단하는 평가이기 때문에, 합성음의 인지속도에 대한 리콜 능력의 영향을 최소화하기 위해서 Miller[5]의 단기 기억에 관한 연구에 근거하여 문장의 길이가 주로 7어절로 구성된 문장들을 선정하여 평가에 사용하였다. 마지막으로, 본 연구에서는 유의미한 문장들을 사용했기 때문에 문장의 친숙도가 리콜 능력에 미치는 영향을 조사하기 위해서 친숙한 문장들과 비친숙한 문장들로 구분하여 문장집합을 구성하였다. 친숙한 문장들은 한국시각장애인연합회에서 전달받은 시각장애인들이 자주가는 웹사이트와 국립국어원의 21세기 세종계획 최종보고서에 수록된 친숙한 관용표현들 중에서 선정되었고, 비친숙한 문장들은 연구자가 자주 방문하는 특수교육 및 학회 관련 학술 사이트의 전문용어나 외래어가 포함된 문장들과 21세기 세종계획 최종보고서에 수록된 관용표현들 중에서 친숙하지 않은 문장들로부터 선정되었다.

이러한 문장집합 선정과정을 통해서 최종 선정된 총 17개 (연습 평가용 2문장 집합, 반복 평가용 10문장 집합, 랜덤(random) 평가용 5문장 집합)의 문장집합들이 본 연구에 사용되었다. 표 2는 각 평가에 사용된 문장 집합의 수와 문장의 개수를 보여준다.

각 문장집합들은 TTS의 질에 따른 영향을 최소화하기 위해서 동일한 한국어 합성음 TTS를 사용하여 15단계의 차별화된 속도로 합성된 단순한 의미 있는 문장들로 구성되었다.

본 평가에 사용된 한국어 합성음 TTS는 현재 국내에서 가장 많이 사용되고 있는 스크린리더용 합성기인 (주) 보이스웨어의 TTS로 뉴스기사 약 10000음절 DB를 사용하여 생성된 현아 음색 TTS를 사용하였다. 기본 문장배속(1배속)은 (주) 보이스웨어에서 1배속으로 제공하는 실제 발화 시간과 정상적

표 1. 참가자 기본정보

Table 1. Demographic information of the participants

순번	나이	성별	장애등급 /원인	스크린리더 사용경력	하루 평균 사용시간
1	35	남	1급/선천성	10년	3시간
2	38	여	1급/후천성 (망막색소변성)	10년	10시간
3	41	남	1급/후천성 (녹내장)	3년	13-14시간
4	41	남	1급/후천성 (사고)	10년	14시간 이상
5	50	여	1급/후천성 (고열)	10년 이상	14시간 이상

휴지(pause)를 포함하는 285.96 분당음절수(SPM: syllables per minute)를 사용하였다. 이는 이전 연구에서 서울 거주성인 30 명을 대상으로 조사한 일반 성인의 말하기에서의 평균 말속도인 265 분당음절수와도 유사하다[6].

표 2. 문장집합의 구성
Table 2. Organization of sentence sets

평가유형	문장 집합 수	문장 개수	제시 방법
반복 연습용	1집합	1문장	동일 문장을 15개의 다른 속도로 제시
반복 평가용	10집합	10문장	
랜덤 연습용	1집합	15x1=15문장	다른 문장을 15개의 다른 속도로 제시
랜덤 평가용	5집합	15x5=75문장	
총계	17집합	101문장	

2.3 평가절차

본 연구는 각 참가자들에 대해서 개별평가 방식으로 진행되었다. 시각장애인 참가자들을 평가가 진행되는 자리로 안내한 후, 전반적인 연구의 목적과 절차에 관한 오리엔테이션을 진행하였다. 합성음의 소리크기가 인지속도에 영향을 미치는 것을 최소화하기 위해서 평가를 시작하기 전 음성파일 샘플을 이용해서 각 참가자들마다 최적의 볼륨을 설정하게 한 후 평가를 시작하였다.

평가는 크게 합성음 문장 리콜 평가, 문장 친숙도 평가, 자유 피드백으로 구성되었으며, 참가자당 평균 1시간 30분 정도의 평가시간이 소요되었다 (표 3 참조). 참가자들은 평가가 진행되는 동안 계속해서 합성음 문장들을 집중해서 듣고 리콜을 해야 하기 때문에 시간이 지날수록 피로도가 증가하고 집중력이 하락하는 것을 방지하기 위해서 원하는 경우 평가의 중간에 휴식을 요청하도록 안내되었다.

모든 평가과정동안 두 명의 연구 보조원들이 각 참가자들의 응답을 개별적으로 응답지에 기록하여 추후 객관적 리콜

표 3. 평가절차
Table 3. Evaluation procedures

평가유형	세부 항목
문장 리콜 평가	객관적 반복 평가
	주관적 반복 평가
	객관적 랜덤 평가
	주관적 랜덤 평가
문장 친숙도 평가	객관적 평가의 문장 친숙도 평가
	주관적 평가의 문장 친숙도 평가
자유 피드백	15배속의 합성음에 대한 자유 의견

정확도 분석 및 주관적 평가분석에 사용하였다. 모든 평가과정과 참가자의 응답은 비디오카메라에 녹화되고 녹음기에 녹음되어서 두 명의 연구 보조원들의 응답기록이 상이한 경우 확인 자료로 활용되었다.

2.3.1 반복 평가절차(Repeated Test)

객관식 평가를 실행하기 전 평가방식에 관한 참가자들의 이해를 돕기 위해서 연습용 문장집합을 사용한 반복 연습평가가 실행되었다. 참가자들은 동일한 문장을 가장 빠른 속도에서 가장 느린 속도의 순서로 15번을 듣고 각 속도에서 들은 그대로 말하도록 요청되었다. 쉬운 속담처럼 익숙한 관용표현의 경우에는 참가자들이 미리 추측해서 대답할 수도 있기 때문에, 이미 아는 관용표현이라도 다른 단어나 조사로 표현될 수도 있기 때문에 미리 예측해서 대답하지 말고 반드시 잘 들은 후에 한 음절씩 들리는 대로 응답하도록 요청되었다.

참가자들이 반복 연습평가를 마치고 평가에 대한 준비가 되었다고 응답한 경우, 실제 반복평가가 실행되었다. 반복평가에서는 총 10개의 문장집합에 대해서 각 문장집합 내의 동일한 문장을 15개의 다른 배속으로 들은 후에 들은 그대로 리콜 하는 방식으로 진행되었다.

객관적 반복평가에서는 참가자들이 “빠~”하는 신호음이 울린 후에 나오는 합성음 문장을 듣고 단 한 글자라도 들리는 그대로 말하도록 요청되었다. 각 참가자들의 응답은 리콜 정확도로 계산되어 리콜 정확도(바르게 리콜된 음절 수/문장 내 전체 음절 수 x 100)가 50%이상인 속도는 최고 속도로 90% 이상인 속도는 최적 속도로 조작적으로 정의되었다.

주관적 반복평가에서는 참가자들이 본인이 스스로 생각하기에 각 문장을 몇 %나 들었다고 생각하는지 0%부터 100%까지의 범위 내에서 10%단위로 응답하도록 요청되었다. 개인들 간의 주관적인 판단 기준이 너무 상이한 경우에는 일관성이 없는 응답을 나타낼 수 있기 때문에 다음과 같은 대략적인 주관적 평가의 가이드라인을 제시해 주었다. 청취한 문장을 전혀 인식하지 못했으면 0%, “~다”와 같이 한 음절 정도만 인식했으면 10%, 한 단어 정도를 인식했으면 20%, 반 정도 인식했으면 40-60%, 한 단어 정도만 빼고 어느 정도 다 인식했으면 70-80%, 문장은 어느 정도 다 들리는데 조사나 한 음절 정도가 불확실한 경우에는 90%, 자신 있게 모든 음절을 다 인식한 경우에는 100%로 응답하도록 요청되었다.

참가자의 객관적/주관적 리콜 정확도가 모두 100%인 경우에는 동일 문장집합 내에서 더 느린 배속으로 넘어가지 않고 바로 다음 문장집합으로 넘어가는 방식으로 평가를 진행하였다. 참가자의 주관적인 판단으로는 청취한 문장을 100% 다 인지했다고 하더라도 실제 객관적 리콜 정확도가 100%가 아닌 경우나, 반대로 참가자의 객관적 리콜 정확도가 100%임에도 불구하고 주관적 판단으로는 100% 다 인지했다고 확신하

지 못하는 경우에는 객관적/주관적 리콜 정확도 모두가 100%에 도달할 때까지 동일 문장집합 내에서 계속 더 느린 배속으로 평가를 진행하였다.

2.3.2 랜덤 평가절차(Random Test)

랜덤평가는 반복평가와 동일한 방식으로 진행되었는데 동일한 문장이 가장 빠른 속도에서 가장 느린 속도의 순서로 15번 제시되는 반복평가와는 달리 랜덤평가에서는 서로 다른 문장들이 15번 제시된다는 차이점이 있다. 참가자들은 연습용 랜덤평가를 통해 랜덤평가의 진행방식을 익힌 후에, 총 5개의 문장집합들에 대해서 각 문장집합 내의 다른 문장들을 15개의 다른 배속으로 들은 후에 들은 그대로 리콜 하는 객관적 랜덤평가와 참가자 본인이 생각하기에 각 문장을 몇 %나 들었다고 생각하는지 10%단위로 응답하는 주관적 랜덤평가에 참여하였다.

랜덤평가에서는 각 참가자의 객관적/주관적 리콜 정확도가 특정 배속에서 모두 100%에 도달했다고 하더라도, 각 문장집합 내에서 서로 다른 문장들을 사용하기 때문에 다음 문장 집합으로 넘어가지 않고, 가장 느린 속도까지 모두 15개의 문장들에 대한 지속적인 평가가 진행되었다.

2.3.3 문장 친숙도 평가절차

총 10개의 문장집합에 대한 반복 평가를 실행한 후에, 반복 평가에 사용된 총 10개의 문장들에 대한 문장 친숙도 평가를 실행하였다. 참가자들은 평가 진행자가 불러주는 반복평가에 사용된 각 문장을 들은 후 그 문장들이 평소에 많이 들어본 쉽고 친숙한 문장이면 “쉬움”, 보통이면 “보통”, 평소에 잘 못 들어본 낯설고 어려운 문장이면 “어려움”으로 응답하도록 요청되었다.

랜덤평가에서는 각 문장집합 내에서 서로 다른 15개의 문장들을 사용했기 때문에, 랜덤평가에 사용된 총 75개(5문장집합 x 각 15개의 문장)의 문장들에 대한 문장 친숙도 평가는 랜덤평가와 동시에 진행되었다. 참가자들은 각 랜덤평가 문장 집합에 대한 객관적 평가와 주관적 평가를 마친 후에 그 문장 집합에 사용된 각 15개의 문장들에 대한 친숙도를 ‘쉬움’, ‘보통’, ‘어려움’으로 평가하도록 요청되었다.

2.3.4 자유 피드백 평가절차

각 참가자들은 가장 빠른 속도에서부터 가장 느린 속도까지 15배속으로 합성된 여러 합성음 문장들을 들으면서 느낀 점이나 합성음 청취시의 어려움, 최고 및 최적의 배속에 대한 의견, 그 외 시각장애인을 위한 고속, 고품질의 합성음 개발과 관련된 의견들을 자유롭게 말하도록 요청되었다.

3. 연구결과

3.1 반복평가 결과

반복평가에서의 참가자들의 합성음 인지속도를 평가한 결과, 참가자들의 합성음 리콜 정확도는 합성음 청취속도가 느려짐에 따라서 향상되는 양상을 나타냈다. 객관적 평가에서의 최고 및 최적 합성음 인지속도는 각각 2.8배속과 1.6배속으로 나타났으며, 주관적 평가에서의 최고 및 최적 합성음 인지속도는 각각 2.8배속, 1.8배속으로 나타났다 (그림 1 참조).

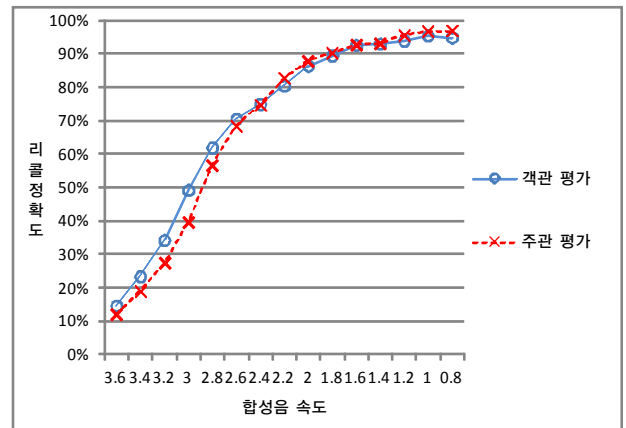


그림 1. 반복평가 결과

Figure 1. Results of the repeated test

3.2 랜덤평가 결과

랜덤평가에서의 참가자들의 합성음 인지속도를 평가한 결과, 객관적 평가에서의 최고 및 최적 합성음 인지속도는 각각 3.0배속과 2.2배속으로 나타났으며, 주관적 평가에서의 최고 및 최적 합성음 인지속도는 각각 2.8배속과 2.0배속으로 나타났다(그림 2 참조).

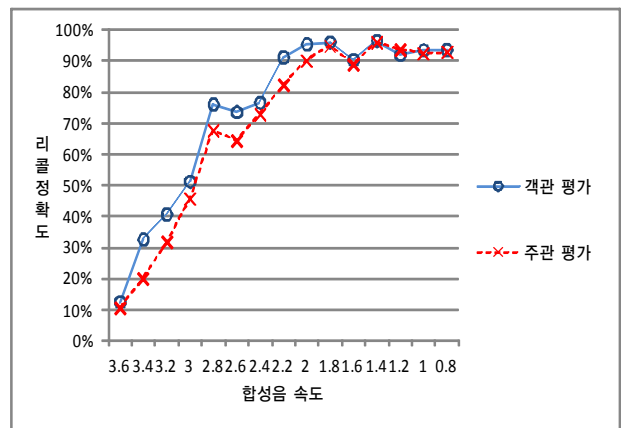


그림 2. 랜덤평가 결과

Figure 2. Results of the random test

3.3 반복/랜덤평가 결과 비교

참가자들의 반복평가와 랜덤평가에서의 합성음 인지속도를 비교한 결과, 랜덤평가에서의 객관적 평가에서 가장 빠른 최고 인지속도를 나타냈고, 그 외의 반복 객관/주관적 평가와 랜덤 주관적 평가에서는 모두 동일하게 2.8배속에서 최고 인지속도를 나타냈다. 참가자들의 합성음 최적 인지속도에 있어서는 객관적/주관적 평가 모두에서 랜덤평가에서의 최적 인지속도가 반복평가에서의 최적 인지속도보다 높게 나타났다. 참가자들의 합성음 최적/최고 인지속도는 <표 4>에 정리되어 있다.

표 4. 반복/랜덤평가에서의 최고/최적 청취속도

Table 4. The maximum/optimum listening speed in the repeated/random test

	반복객관	반복주관	랜덤객관	랜덤주관
최고 속도	2.8	2.8	3.0	2.8
최적 속도	1.6	1.8	2.2	2.0

3.3 문장 친숙도 평가결과

3.3.1 반복평가에서의 상관관계 비교

반복평가용 10개의 문장집합들에 사용된 10개의 문장들에 대한 참가자들의 문장 친숙도 평가결과와 각 문장에 대한 참가자들의 객관적 리콜 정확도를 비교한 결과, 최고 인지속도와 최적 인지속도는 각각 3.0배속, 2.0배속으로 나타났다. 보통 난이도의 문장에 대한 최고 인지속도와 최적 인지속도는 각각 2.8배속, 2.2배속으로 나타났다. 어려운 문장에 대한 최고 인지속도는 2.6배속으로 나타났으나 이후 2.4 배속에서는 유지되지 못하고 2.2배속부터 유지되었으며, 최적 인지속도에는 가장 느린 배속에서도 도달하지 못하였다 (그림 3 참조).

반복평가에서의 문장 친숙도에 따른 최고/최적 속도는 <표 5>에 제시되어 있다.

표 5. 반복평가 문장 친숙도에 따른 최고/최적 청취속도

Table 5. The maximum/optimum listening speed by sentence familiarity in the repeated test

	쉬움	보통	어려움
최고 속도	3.0 (3.0)	2.8 (2.8)	2.6 (2.2)
최적 속도	2.0 (2.0)	2.2 (2.2)	미성취

*괄호안의 숫자는 지속적으로 유지되는 최고속도를 의미함.

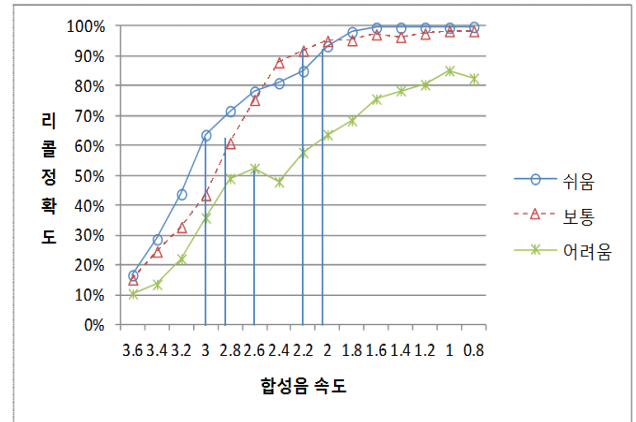


그림 3. 반복평가에서의 합성음 리콜 정확도와 문장 친숙도 간의 상관관계

Figure 3. Correlations between recall accuracy and sentence familiarity in the repeated test

3.3.2 랜덤평가에서의 상관관계 비교

랜덤평가용 5개의 문장집합들에 사용된 75개의 문장들에 대한 참가자들의 문장 친숙도 평가결과와 각 문장에 대한 참가자들의 객관적 리콜 정확도를 비교한 결과, 쉬운 문장에 대한 최고 인지속도는 3.2배속으로 나타났고, 최적 인지속도는 2.2배속으로 나타났다. 보통 난이도의 문장에 대한 최고 인지속도는 3.4배속으로 나타났으나 이후 3.3배속과 3.0배속에서는 유지되지 못하고 2.8배속부터 최고 속도가 유지되었고, 최적 인지속도는 2.0배속으로 나타났으나 1.6배속과 가장 느린 배속인 0.8배속에서는 유지되지 않았다. 어려운 문장에 대한 최고 인지속도는 2.6배속으로 나타났으며, 최적 인지속도는 2.2 배속으로 나타났으나 이후의 2.0, 1.6, 1.2배속에서는 유지되지 못하고 불규칙한 양상을 나타냈다 (그림 4 참조).

랜덤평가에서의 문장 친숙도에 따른 최고/최적 속도는 <표 6>에 제시되어 있다.

표 6. 랜덤평가에서의 문장 친숙도에 따른 최고/최적 청취속도

Table 6. The maximum/optimum listening speed by sentence familiarity in the random test

	쉬움	보통	어려움
최고 속도	3.2 (3.2)	3.4 (2.8)	2.6 (2.6)
최적 속도	2.2 (2.2)	2.0 (불규칙)	2.2 (1.0)

* 괄호안의 숫자는 지속적으로 유지되는 최고/최적 속도를 의미함.

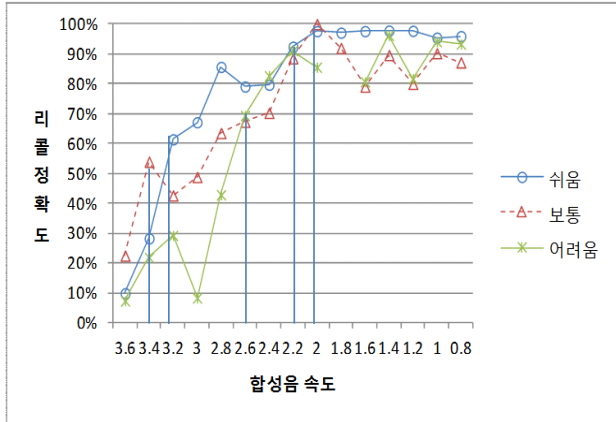


그림 4. 랜덤평가에서의 합성음 리콜 정확도와 문장 친숙도 간의 상관관계

Figure 4. Correlations between recall accuracy and sentence familiarity in the random test

3.4 자유 피드백 평가결과

평가에 사용된 합성음에 대한 개인의 의견을 자유롭게 제시하는 자유 피드백 평가결과 참가자들은 합성음의 속도가 빨라서 인지하지 못하는 경우도 있지만 합성음이 명료하지 않아 잘 인지하지 못하는 경우도 있다는 의견들을 제시하며 다음의 사항들에 대한 합성음의 명료도 개선을 요구하였다: (1) 본 평가에 사용된 합성음 뿐 아니라 대부분의 TTS에서 ‘ㅅ’이 첫 자음으로 나오는 경우에는 ‘ㅈ’이나 ‘ㅎ’과 혼동이 되는 경우가 많고, (2) ‘S’와 ‘X’와 같이 비슷한 발음의 영어 구분이 어려우며, (3) ‘ㅈ’과 ‘ㄷ’의 구분이 명료하지 않으며, (4) 받침 ‘ㄴ’이 잘 안 들리는 경향이 있다.

또한 빠른 속도의 합성음에서 기본 잡음이 있어 명료성 개선이 요구되며, 문장의 띄어쓰기를 자연스럽게 읽어 줄 필요가 있다는 의견들을 제시하였다. 느린 속도의 합성음의 경우에는 속도가 느리기 때문에 합성음 인지가 쉬울 것이라고 생각하지만, 실제로는 억양의 변화 없이 속도만 느려지기 때문에 느린 속도임에도 오히려 잘 안 들리는 경우도 있음을 나타냈다 (예: ‘회원만이 가입할 수 있습니다’에서 ‘만이’의 억양이 없어 잘 안 들림).

합성음의 인지속도는 합성음의 속도뿐 아니라 합성음의 성별에 따라서도 달라진다는 의견도 제시되었다. 여성 합성음은 음의 피치가 높아서 남성 합성음으로 청취한 경우 인지할 수 있는 속도에서도 여성 합성음에서는 인지하지 못하는 경우가 있기 때문에 합성음의 성별에 따른 청취 시의 인지속도에 대한 객관적인 연구가 이루어 질 필요가 있다는 의견도 제시되었다.

참가자들은 빠른 속도의 자연스럽게 명료한 합성음, 장시간 청취 시에도 피로하지 않은 고속, 고품질의 합성음이 개발되어 상용화되면 좋겠다는 공통된 의견들을 나타냈다.

4. 결론 및 제언

4.1 결론

각 문장집합 내에서 동일한 문장이 가장 빠른 속도에서 가장 느린 속도 순서로 15번 제시되는 반복평가에서는 참가자들의 합성음 인지속도가 배속이 느려짐에 따라 증가하는 경향을 나타냈다. 반복평가에서의 합성음 최고 인지속도는 객관적 평가와 주관적 평가에서 동일하게 나타났으며, 최적 인지속도는 참가자들의 주관적 평가가 객관적 평가보다 한 단계 더 빠른 배속으로 나타났다. 객관적 평가에서는 각 문장 집합 내에서 동일한 문장들이 반복적으로 제시되기 때문에 참가자들이 실제 자신들이 리콜 한 것보다 본인의 리콜 정확도를 주관적으로 더 높게 판단하는 경향이 있음을 나타낸다.

랜덤평가에서 참가자들의 합성음 인지속도 또한 배속이 느려짐에 따라 전반적으로 증가하는 경향을 나타냈다. 랜덤평가에서의 합성음 인지속도는 객관적 평가에서의 최고/최적 인지속도가 주관적 평가에서의 최고/최적 인지속도보다 각각 한 단계 빠른 배속으로 나타났다. 이는 랜덤평가에서는 각 문장 집합 내에서 서로 다른 문장들이 제시되기 때문에 참가자들이 그들이 실제로 리콜을 한 것보다 스스로 생각하기에 본인의 리콜 정확도를 더 낮게 평가하는 경향이 있음을 보여준다.

이러한 랜덤평가에서의 객관적/주관적 평가결과격차의 격차는 속도가 너무 빨라서 거의 듣지 못하는 가장 빠른 배속이나 속도가 느려서 대부분 잘 인식할 수 있는 느린 배속에서는 격차가 좁아지고, 그 외의 배속에서는 실제 객관적 리콜 정확도와 스스로의 주관적 판단에 의한 리콜 정확도간의 격차가 있음을 보여준다. 실제 평가과정에서 참가자들은 자신들이 청취한 합성음 문장의 일부 혹은 전체를 정확히 리콜 하였음에도 불구하고 자신들의 응답에 확신을 갖지 못하고 자신 없어 하는 모습을 종종 보이거나, 특정 배속에서 실제 100% 정확하게 리콜을 하더라도 주관적인 생각으로는 100% 정확하게 인지하지 못한 것으로 응답하는 모습을 보이기도 하였다.

반복평가와 랜덤평가의 결과를 비교해 보았을 때, 예상과는 달리 서로 다른 문장들을 배속별로 열 번 청취하는 랜덤평가에서의 최고 및 최적 인지속도가 동일한 문장을 배속별로 열 번 청취하는 반복평가에서의 최고 및 최적 인지속도보다 더 높게 나타났다. 이는 반복평가와 랜덤평가에서 사용된 합성음 문장들의 친숙도 및 문장길이(문장 내 음절 수)가 최고/최적 합성음 인지속도에 영향을 미쳤기 때문인 것으로 생각된다.

참가자들의 합성음 인지속도와 문장 친숙도간의 상관관계를 평가한 결과, 반복평가에서는 모든 배속에서 참가자들의 객관적 리콜 정확도와 문장의 친숙도간의 높은 정적 상관관계를 나타냈다. 참가자들은 평가에 사용된 문장이 자주 접해 온 친숙하고 쉬운 문장일수록 쉽게 인지하였고, 전문용어나 외래어, 숫자 등이 들어간 친숙하지 않은 문장일수록 인지에 어려

음을 나타냈다. 이는 시각장애인들이 합성음을 인지할 때 속도뿐만이 아니라, 청취하는 내용에 관한 배경지식이나 친숙도 또한 중요한 요인으로 작용함을 보여준다.

자유 피드백에서는 한국어 합성음 청취 시 인식의 어려움이 단지 합성음 속도가 빨라서가 아니라 합성음의 명료도와 관련이 있다는 의견들이 제시되었다. 이를 보완하기 위해서는 혼동이 되기 쉬운 특정 자음이나 모음, 알파벳 합성음에 대한 명료도를 개선하고 합성음 문장의 띄어쓰기와 억양을 자연스럽게 표현할 필요가 있다. 또한 빠른 속도의 합성음의 경우 음이 깨지는 소리나 울리는 소리와 같은 잡음이 있다는 의견을 반영하여 고속 합성음의 잡음을 최소화 할 필요가 있다.

4.2 후속 연구를 위한 제언

본 연구에서는 합성음 인지속도가 리콜 능력에 영향을 받는 것을 최소화하기 위해서 주로 7어절로 구성된 문장들을 평가에 사용함으로써 최대한 객관성을 유지하고자 하였지만 평가에 사용된 각 문장 내의 음절수는 동일하게 통제하지 못한 제한점을 가지고 있다. 따라서 후속 연구에서는 각 문장 내의 음절수와 어절 수 등이 동일한 객관적인 평가문장의 집합을 사용한 합성음 인지속도에 대한 평가가 이루어질 필요가 있다.

또한 합성음 청취 시 특정 자음이나 모음이 문장이나 단어 내에서 어디에 위치하는지에 따라서 정확한 인식 속도가 달라진다는 참가자들의 의견을 고려하여, 특정 자음이나 모음의 문장 혹은 단어 내 위치에 따른 인식 속도나 명료도 패턴을 분석하는 후속 연구가 실행 될 필요가 있다.

본 연구는 컴퓨터 및 스크린리더를 업무상 일상적으로 사용하는 참가자들을 대상으로 실행되었기 때문에 본 연구의 결과를 컴퓨터나 스크린리더의 사용 경험이 거의 없는 시각장애인들에게 일반화하기에는 제한이 있다. 그러므로 후속 연구에서는 컴퓨터 및 스크린리더의 사용경험이 거의 없는 시각장애인들을 대조 집단으로 포함하여 참가자들의 컴퓨터 및 스크린리더 사용경험이 실제 합성음의 인지속도에 영향을 미치는지에 관하여 비교 조사해 볼 필요가 있다.

마지막으로 여성 합성음은 음의 피치가 높아서 남성 합성음으로 청취한 경우 인지할 수 있는 속도에서도 여성 합성음은 인지하지 못하는 경우가 있다는 참가자들의 의견을 반영하여, 실제로 한국어 합성음의 인지속도가 합성음의 성별에 의해 영향을 받는지에 관한 객관적이고 체계적인 연구가 실행될 필요가 있다.

참고문헌

[1] Asakawa, C. Takagi, H. Ino, S., & Ifukube, T. (2003). Maximum listening speeds for the blind. *Proceedings of the*

2003 International Conference on Auditory Display, Boston: MA., 276-279.

[2] Lee, H., & Hong, K-H. (2012). Issues on how to evaluate maximum and best listening speed of the Korean TTS for the blind. *Proceedings of 2012 Fall Conference of The Korean Society of Speech Sciences*, 273-275.

(이희연 · 홍기형 (2012). 시각장애인의 한국어 합성음 인지속도 평가 방법 도출을 위한 과제. 2012 한국음성학회 가을 학술대회발표논문집, 273-275.)

[3] Lee, H., Hong, K-H., & Park, D. (2013). A sentence set for evaluation on maximum and best listening speed of the Korean TTS for the blind. *Proceedings of 2013 Spring Conference of The Korean Society of Speech Sciences*, 242-244.

(이희연, 홍기형 · 박단비 (2013). 시각장애인의 한국어 합성음 인지속도 평가를 위한 문장집합. 2013 한국음성학회 봄 학술대회발표논문집, 242-244.)

[4] Lee, H., & Hong, K-H. (2012). A study on usability and needs of speech synthesizer for the blind. *Special Education*, 11(2), 55-75.

(이희연 · 홍기형 (2012). 시각장애인용 음성합성기에 대한 사용자 요구분석. 특수교육, 11(2), 55-75.)

[5] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63(2), 81-97.

[6] Shin, M., & Han, S-J. (2003). A study of speech rate and fluency in normal speakers. *Speech Sciences*, 10(2), 159-168.
(신문자 · 한숙자 (2003). 정상 성인의 말속도 및 유창성 연구. 음성과학, 10(2), 159-168.)

• 이희연 (Lee, Heeyeon)

서울대학교 QoLT센터
서울시 성북구 동선동 2가 169-1 미디어정보관 402호
Tel: 02-877-1268

Email: heeyeonlee@hotmail.com

관심분야: 발달장애, 지적장애, 보완대체의사소통시스템
2011~현재 서울대학교 QoLT 센터 선임연구원

• 홍기형 (Hong, Ki-Hyung) 교신저자

성신여자대학교 IT 학부
서울시 성북구 동선동 2가 169-1 미디어정보관 403호
Tel: 02-920-7525 Fax: 02-929-7525

Email: Kihyung.hong@gmail.com

관심분야: 삶의 질 향상 기술, 웹정보시스템, 사용자 인터페이스

1998~현재 성신여자대학교 IT학부 교수