Journal of Korean Institute of Intelligent Systems
Vol. 31, No. 3, June 2021, pp. 213-221
http://dx.doi.org/10.5391/JKIIS.2021.31.3.213

ISSN(Print) 1976-9172
ISSN(Online) 2288-2324

시각 정보의 청각 대체를 위한 청각 해상도 분석 연구

Analysis of Auditory Sensitivity for Visual to Auditory Sensory Substitution

김무섭*, 정치윤*†
Mooseop Kim and Chi Yoon Jeong†

*한국전자통신연구원 인공지능연구소 책임연구원
*Principal Researcher, Artificial Intelligence Research Lab., ETRI

요 약

시각 정보를 청각 신호로 대체하는 감각 치환(Visual-Auditory Sensory Substitution : VASS) 기술은 시각장애인이 제한적이지만 다른 사람의 도움 없이 물체를 인식하거나 기본 생활에 필요한 정보를 획득할 수 있는 가능성을 보여주었다. 그러나 이러한 가능성에도 불구하고 VASS 기술은 시각장애인 사회에 성공적으로 활용되지 않은 채 실험실 규모의 연구 또는 시제품 형태의 결과로 남아 있다. 이러한 문제의 해결을 위해서는 효율적인 기술 개발과 함께 VASS 신호에 대한 시각장애인의 청각 해상도에 대한 분석이 선행되어야 한다. 따라서 본 연구에서는 시각장애인의 청각 해상도 분석을 위한 실험을 제안하였다. 청각 해상도를 분석한 본 연구의 실험 결과는 시각 정보를 전달하는 VASS 신호의 길이를 기존 대비 50%까지 줄일 수 있음을 보여주었다. 이러한 결과는 기존 VASS 기술의 성능이 최대 2배까지 향상할 수 있음을 의미하며, 사용자를 대상으로 진행한 실험을 통하여 검증한 결과를 제시하였다.

키워드 : 감각대체, 해상도, 시각, 청각, 청각화

Received : Apr. 23, 2021
Revised : Jun. 21, 2021
Accepted : Jun. 21, 2021
† Corresponding author
(iamready@etri.re.kr)

Abstract

Visual-auditory sensory substitution (VASS), which replaces visual information with auditory signals, has demonstrated limited but significant potential to help visually impaired and blind people to recognize objects and to obtain information for daily life without the help of others. However, despite this possibility, VASS methods remain as laboratory-scale research or pilot demonstrations without successful adoption in the blind community. To solve this problem, an analysis of the auditory sensitivity of the visually impaired on the VASS signal must be preceded and combined with the development of efficient methodologies. Therefore, in this study, we proposed an experiment to analyze the auditory sensitivity of the visually impaired, which is necessary for the development of efficient VASS technology. The experimental results of this study showed that the temporal length of the VASS signal could be reduced by 50% compared to the previous studies. This result indicates the possibility of improving the performance of the conventional VASS method by up to two times. We also confirmed that our experimental results are reasonable through behavioral assessment.

Key Words : Sensory Substitution, Sensitivity, Visual, Auditory, Sonification

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [21ZS1200, 인간중심의 자율지능시스템 원천기술연구]

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

시각은 사람이 일상생활을 영위함에 있어서 필요한 환경 정보를 받아들이는 인체의 오감 중에서 가장 중요한 부분을 차지한다. 로버트 뮈르(Robert

Muir)는 사람이 외부로부터 인지하는 정보의 83%가 시각을 통해서 이루어진다고 분석하였다.

따라서 시각은 주변 환경을 지각하는 핵심적인 수용기관이며, 시각을 상실한 경우, 사람의 지각 행동에 미치는 영향은 경험의 제약, 보행의 불편, 상황 파악의 어려움 등으로 연결된다. 이러한 문제점들은 단독적으로 발생하는 현상이 아니라 일상생활 전반에서 복합적으로 동시에 발생하게 된다.

최근에 시각이 손상되거나 기능이 저하된 고령자와 장애인이 증가하면서 이들의 손상된 감각 및 지각 능력을 향상시켜 지속적인 경제 활동과 삶의 질을 향상시킬 수 있는 기술적 대안으로 시각 정보를 청각 신호로 변환하는 감각 치환(Visual-Auditory Sensory Substitution: VASS) 기술에 관한 관심이 증가하고 있다[1]. 일반적으로 감각 치환 기술은 기능이 저하되거나 손상된 감각 신호를 다른 감각 기관으로 전달 또는 대체하여 사용하는 것을 의미한다. 이러한 감각 치환 기술은 사람의 뇌가 새로운 환경에 적응하기 위하여 구조, 기능적으로 변화하고 재조직되는 뇌 가소성(Brain plasticity)에 기반하는 것으로 알려져 있다[1].

대다수 감각 정보의 경우 감각 기관의 특성과 감각 수용체가 표현할 수 있는 정보의 양이 다르므로 단일 감각 정보를 활용하여 전체 영상이나 이미지의 정보를 표시하기에는 어려움이 있다. 또한, 기존 VASS 방법의 경우, 1~2초 정도의 소리를 통해 하나의 이미지 프레임에 해당하는 정보를 전달하기 때문에 실 생활에서 변화하는 주변 환경 정보나 빠르게 접근하는 물체를 인식하기에는 어려움이 있다. 이러한 문제들은 VASS 기술들이 그동안 기술적 가능성은 입증하였지만, 여전히 실험실 수준의 연구를 벗어나지 못하는 근본적인 문제로 파악된다.

따라서 효율적인 VASS 기술의 개발을 위해서는 감각 수용체인 청각의 한계 해상도와 특성을 고려하여 적절하게 데이터를 변환하고 정보를 전달하는 방법이 필요하다. 이를 위해 VASS 기술의 잠재적 사용자인 시각장애인들의 인지 및 행동 특성에 대한 정확한 이해와 함께 시각장애인이 어느 정도까지 축약된 소리에서 시각 정보를 인식할 수 있는지 확인하고 분석하는 과정이 선행되어야 한다. 시각장애인에 대한 지각 인식도 및 청각 해상도 분석은 선천적 시각장애인(Congenital Blind: CB), 후천적 시각장애인(Late Blind: LB) 및 비장애인(Sighted User: SU)을 포함하는 다양한 그룹의 실험 참가자들에 대하여 VASS 신호의 학습 진행, VASS 신호에 대한 지각 정도 등 시각장애인과의 직접적인 소통을 통한 시험 및 조사가 필요하다.

본 연구에서는 효율적인 VASS 기술 개발을 위하여 필수적인 VASS 신호에 대한 시각장애인의 청각 해상도 및 지각 인식도를 분석하였다. 이를 위해 VASS 신호에 대한 시각장애인의 지각 정확도를 측정하고 적용 가능성과 요구사항을 도출하기 위해 2단계에 걸친 시험을 진행하였다. 1단계 시험에서는 축약된 일반 음향 정보에 대하여 비장애인과 시각장애인의 청각 신호에 대한 인식도를 비교, 분석하여 사용자 그룹별로 나타나는 청각 해상도의 차이와 특징을 파악하였다. 2단계 시험에서는 1단계의 시험 결과를 반영하여 VASS 신호의 축약 정도에 대한 비장애인과 시각장애인의 지각 인식도 및 유효해상도를 비교, 분석하여 시각장애인이 VASS 신호로부터 시각 정보를 인식할 수 있는 최대 신호 압축률을 도출하였다.

본 논문의 구성은 다음과 같다. 2절에서는 VASS 기술과 관련된 선행연구들에 관하여 소개하고, 3절에서는 일반 음향 신호의 압축율 변화에 따른 시각장애인과 비장애인의 지각 인식도 및 해상도 분석 시험을 소개한다. 4절에서는 VASS 신호의 축약 정도에 대한 비장애인과 시각장애인의 지각 인식도 분석을 위한 시험을 소개한다. 마지막으로, 본 연구의 대한 결론 및 향후 연구 방향을 5절에 제시하였다.

2. 시각-청각 변환 감각대체 관련 연구

시각 정보를 소리로 변환하여 전달하는 VASS 기술은 정보를 시각에 의존하지 않고 청각을 이용하여 시각 정보를 새로운 감각의 형태로 정보를 접할 수 있는 방식을 제공한다. VASS 기술 중 초기에 제안되었으나 현재까지도 가장 광범위하게 사용되는 방법은 1992년에 피터 메이어르(Peter Meijer)가 제안한 vOICe이다[2]. vOICe는 그레이 스케일의 64x64 해상도의 2차원 이미지를 소리로 변환하여 사용자에게 전달함으로써 제한적이지만 사용자가 소리를 통해 주변 정보를 지각할 수 있는 기능을 제공한다. 단일 프레임의 이미지를 사운드스케이프(Sound scape)라는 청각 신호로 표현하기 위해 vOICe는 이미지 픽셀의 수평 및 수직 위치와 밝기 정보를 사용한다.

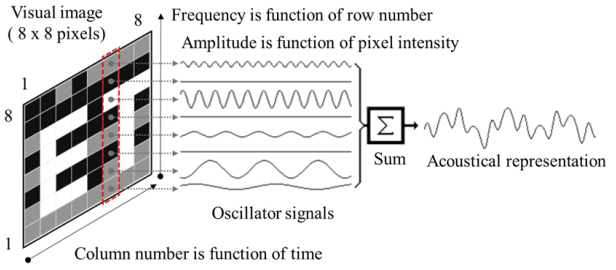

그림 1. vOICe의 시각-청각 변환 개념도[1]

Fig. 1. Concept for Visual to Auditory conversion of the vOICe[1]

그림 1에서 예시한 것과 같이 vOICe는 시간 축을 기준으로 해당 이미지 프레임을 열 단위로 구분하여 각 열에서 픽셀의 위치와 밝기 정보를 소리로 변환하는 과정에 사용한다. 시간 축으로 구분된 각 열에서 픽셀의 위치에 따라 최저 500Hz에서 최고 5,000Hz까지 지수적으로 변화하는 구간에서 각각 다른 주파수를 할당하고, 픽셀의 밝기에 따라 다른 크기의 소리로 변환한다. 이렇게 이미지 프레임의 열 단위로 개별 픽셀에 할당되는 주파수와 크기를 가지는 음향들을 모두 합하여 그 열을 나타내는 소리 정보로 변환되고 샘플링되어 사용자에게 전달된다. vOICe는 이러한 열 단위의 변환 과정을 하나의 이미지 프레임을 구성하는 모든 열에 대하여 왼쪽에서 오른쪽으로 진행하여 1개의 이미지 프레임을 1초 또는 2초 단위의 사운드스케이프 신호로 변환하여 전달한다. vOICe에서 사용하는 사운드스케이프 방법은 신호의 변환 방법이 직관적이고 단순하여 시각장애인이 VASS 신호를 학습하기 쉽고 이미지 정보를 분석하여 의미를 추출하는 등의 추가적인 처리 과정이 필요 없다는 장점이 있다. 실제로 vOICe를 활용하여 일정 기간(최소 수십 시간에서 최대 수 개월) 동안 학습한 시각장애인들이 물체를 구분하고, 주변 정보를 지각하게 되는 실험적 결과와 사례들이 발표되었다[3-5].

이미지나 영상에서 색상은 물체의 인식이나 상황을 인지하는 과정에 사용되는 중요한 정보이다. 그러나 위에서 살펴본 vOICe의 경우, 이미지나 영상을 그레이 스케일로 변환하여 처리함으로써 영상에서 중요한 요소인 색상 정보의 손실이 발생한다. 또한, 시각 정보를 청각 신호로 변환하는 과정에서 생성하는 소리는 임의로 할당된 주파수 대역의 소리들을 합성하여 생성하므로, 일정 시간 동안 지속하여 듣기에는 거부감이 있다는 단점이 있다.

vOICe가 가진 이러한 문제점을 해결하기 위하여 이스라엘 히브리대 연구진은 EyeMusic을 개발하였다[6]. EyeMusic은 vOICe에서 제안한 사운드스케이프 방식을 사용하지만, 이미지의 색상 정보를 6개의 대표적인 색상 정보(빨강, 파랑, 녹색, 노랑, 흰색, 검정색)로 군집화하고, 분류된 대표 색상에 각각 다른 악기를 사용하여 사운드 정보로 표현하였다. 최대 주파수 대역이 5,000Hz인 vOICe와 달리 사운드스케이프 생성에 사용하는 주파수 대역을 사람들이 청취하기 좋은 대역인 최대 1,568Hz로 설정하고 열의 픽셀 위치에 할당하는 주파수를 5음계로 매핑하여 오디오 신호로 전달하여 사용자의 거부감을 줄였다. EyeMusic을 사용하여 지속적인 반복훈련을 거치면 선천적 시각장애인도 눈앞에 놓인 물체의 모양과 크기, 색깔까지 구분할 수 있으며, 방 전체의 풍경을 구별할 수 있음을 실험적으로 검증하였다.

지금까지 살펴본 사운드스케이프 방식을 사용하는 기존 VASS 기술들의 특징을 정리하면 표 1과 같다. 표 1의 특징들에 따르면, 기존 연구들은 시각 정보를 청각 신호로 변환하는 방법 위주로 연구되어왔으며, 효율적인 VASS 기술 개발을 위해 선행되어야 하는 사용자의 청각 해상도에 대한 연구는 간과되어 왔음을 알 수 있다.

표 1. 기존 시각-청각 변환 방법의 특징
Table 1. Comparison of conventional VASS methods

Methods	Input image	Soundscape length (sec.)	Frequency range (Hz)	Conveyed information
vOICe [2]	64x64 (Gray)	1~2	500~5,000	Shape
EyeMusic [6]	24x40 (Color)	2	65~1,568	Shape, Color

3. 음향의 축약도 변화에 대한 청각 해상도 분석

일반적으로 시각장애인은 비장애인보다 더 뛰어난 청각을 가진다고 알려져 있다. 이러한 특성은 같은 소리라도 시각장애인의 경우 비장애인과 비교해 더 축약된 소리에서도 정보를 이해할 수 있다는 가설의 수립이 가능하다. 기존 VASS 기술의 경우, 시각-청각 변환에 이러한 사용자의 특성이 고려되지 않았다. 따라서 본 연구에서는 같은 소리라도 비장애인과 시각장애인이 축약되어 전달되는 소리에서 정보를 인식하는 정도의 차이 유무를 확인하는 청각 해상도 분석을 선행 시험으로 진행하였다. 만일 비장애인과 시각장애인의 청각 해상도에 차이가 존재한다면, 이러한 특성을 반영하여 기존의 VASS 기술에 적용하는 경우 같은 시간에 시각장애인에게 더 많은 정보를 전달할 수 있을 것이다.

3.1 실험 방법

이러한 가설의 검증을 위하여, 먼저 소리로 표현되는 일반 문장을 축약했을 때 각 그룹별(CB, LB, SU)로 인지하는 축약 정도의 평균치를 분석하고, 축약된 소리를 인지하는 정도가 그룹별로 차이가 있는지를 확인하였다. 축약된 일반 음향에 대한 청각 해상도 분석을 위한 실험과정은 그림 2에 나타내었으며, 전체 실험과정은 본 연구의 실험과정을 위탁하여 진행한 강남대학교 기관생명윤리위원회의 승인하에 진행하였다(승인번호: KNU_HR2020010).

일반 음향의 축약도 변화에 대한 청각 해상도 분석을 위한 실험 참가자는 만 18세 이상의 선천·후천 시각장애인(CB, LB)으로 각 10명씩 2개 그룹

과 비장애인(SU) 10명으로 1개 그룹 총 3개의 그룹 30명으로 구성하였다.

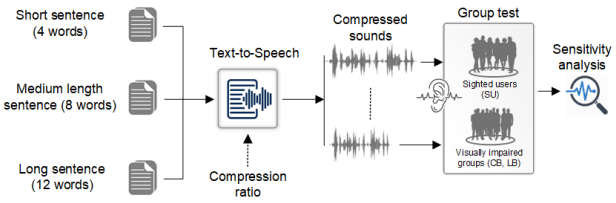

그림 2. 청각 해상도 분석 실험 절차
Fig. 2. Procedure for auditory sensitivity analysis

실험에 사용한 소리 정보로는 일상생활에서 들을 수 있는 주변 환경 소리의 경우, 소리의 주파수나 피치 등의 특징에 따라 축약도와 상관없이 정보를 파악할 수 있으므로 청각 해상도 분석에 사용하기에는 부적합하다. 따라서 의미를 내포하는 문장을 축약한 소리로 변환하였을 때, 각 문장의 의미를 이해할 수 있는지를 청각 해상도 분석 지표로 설정하였다. 이러한 기준에서 실험에 사용한 문장의 길이와 음절의 구성을 위하여 아래의 선행연구의 결과를 적용하였다.

일반 성인들의 구어 속도 선호도를 연구한 선행연구[7]에 따르면, 일반 성인들은 일상 발화에서 구어 속도로 250~300 SPM(Syllables Per Minute)을 가장 선호하는 것으로 나타났다. SPM은 1분당 발화하는 음절 수로 이를 초당 음절 수(Syllables Per Second: SPS)로 환산하면 4~5 SPS, 즉 4~5 음절이 된다. 또한 지역, 성별, 세대와 같은 사회언어학적 변수가 한국어의 발화 속도에 어떠한 영향을 미치는지를 고찰한 연구[8]에 따르면 한국어의 전체 평균 말 속도는 4.82 SPS, 평균 조음 속도는 5.99 SPS로 나타났다. 한편 Baddeley[9]는 단어나 숫자와 같이 친숙한 자료에 대한 즉시적 기억폭은 일정 시간 동안 의식되지 않는 상태로 발성적으로 되새길 수 있는 자료의 양과 동등하다고 강조였는데 이는 대략 1.5초 또는 2초 정도로 추정된다[10]. 또한, 한국어 단어의 글자 수, 받침의 수와 위치, 합성어 여부가 기억폭에 미치는 효과를 검증한 연구[10]에 따르면 전반적으로 글자가 많은 단어일수록 기억폭이 작았고, 받침이 많을수록 기억폭이 더 작았으며, 합성어보다 단일어의 기억폭이 더 크게 나타났다. 본 연구에서는 이상의 선행연구 결과를 검토하여 표 2와 같이 1초에 4글자, 2초에 8글자, 3초에 12글자로 난이도를 구분하였으며, 난이도마다 10개의 문장으로 구성하였다.

시험 음향 제작은 먼저 난이도별로 구분한 각 문장을 TTS(Text to Speech) 소프트웨어로 재생하면서 어색한 발음을 교정하여 문장 파일로 생성하였다. 이때 각 문장의 음성은 정해진 압축률을 넘지 않도록 하였다. 문장을 소리로 변환할 때 일반적인 발음 속도와 비슷하도록 변환하였으며, 이를 기본 속도로 설정하였다. 기본 속도의 경우 4글자는 0.75초, 8글자는 1.5초, 12글자는 2.25초의 길이를 갖는다. 이 기준은 Baddeley[9]가 제시한 즉시적 기억폭인 1.5초에서 2초 사이에 근접하는 시간이다. 청각해상도 분석을 위한 소리의 길이는 기본 재생 시간의 20%, 25%, 30%, 35%, 40%, 50%, 75% 가 되도록 압축하여 사용하였다. 오디오 편집으로 소리 재생속도를 빠르게 하였을 때 음성이 깨지는 현상이 발생하여 추가적인 보정 작업을 진행하였다.

표 2. 시험 사운드 생성을 위한 문장
Table 2. Sentences for generation of test sound

No	4 char.	8 characters	12 characters
	easy	medium	difficult
1	고마워요	커피 마시러 오세요	차 마시러 우리 가게로 오세요
2	예쁘네요	기차 타고 대구 가요	피노키오가 차 타고 무주 가요
3	서러워라	아파트로 이사 가요	누나네 가게로 과자 사러 가요
4	미안해요	나의 꿈은 요리사다	너는 자라서 무엇이 되고 싶니
5	사랑해요	우리는 꼭 만나야 해	개와 고양이는 사이가 좋아요
6	힘이나네	바람이 불면 떠나요	비가 오면 우산을 쓰도록 해요
7	멋지구나	난 승마를 좋아한다	우리 가족은 서로 많이 사랑해
8	새롭네요	담배는 건강에 나빠	나의 소망은 함께 지내는 거야
9	출발하자	오늘은 무엇을 먹지	내일 그는 영화를 보기로 했다
10	행복하니	서울역에 도착했다	건강을 위해 검진을 받습니다

축약된 소리의 재생은 팟플레이어 64비트 프로그램을 활용하였으며, 시험 과정 중 주변 환경 음과의 차단을 위해 헤드셋(음량 50%)을 착용한 후 진행하고 보청기 착용자의 경우 헤드셋 대신 노트북 스피커로 직접 음향 재생하였으며, 비장애인(SU) 참여자의 경우 테스트 시작 단계에서 안대를 착용한 후 진행하였다. 일반 음향의 인식도 분석은 반복되는 소리 정보의 청취에서 발생할 수 있는 학습효과의 부작용을 최소화하기 위하여 가장 많이 축약된 음향(레벨7, 20%)부터 가장 적게 축약된 소리(레벨1, 75%)로 순차적으로 청취하며 진행하였으며, 각 레벨의 소리에 대하여 시험자의 응답을 연구진이 기록하여 인식도를 측정하였다. 추가

로, 실험 후 인터뷰에서 가장 편안하게 들린 속도의 소리를 확인하고 축약 정도에 따른 청취 과정의 불편함과 사용자 의견을 조사하였다.

청각 해상도의 평가 방법으로는 문장의 각 레벨과 난이도에 해당하는 가중치 점수를 서로 곱한 합계 점수를 사용하였다. 모든 테스트 과정은 하위 난이도(쉬움, easy)의 상위 레벨(레벨7, 20%)부터 실시하였다. 난이도와 레벨에서 정답으로 인정되는 경우, 그 난이도의 하위 레벨들은 평가를 진행하지 않고 해당 점수로 산정하였다. 단, 해당 난이도의 문장에서 가장 낮은 레벨 1까지 청취 후에도 문장의 의미를 이해하지 못한 경우 0점으로 처리하였다.

3.2 실험 결과

일반 음향의 축약도 변화에 대한 청각 해상도 분석을 위해 사용자 대상으로 시행된 시험 동안 수집된 자료는 실증적 통계분석을 위해 통계프로그램 IBM SPSS Statistics 21을 이용하여 통계 처리하였다.

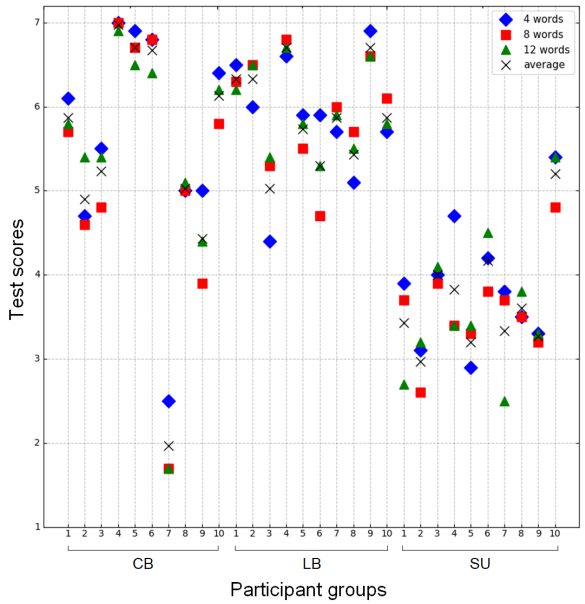

그림 3. 참가자 별 시험 점수 평균 분포
Fig. 3. Test score distribution of participants

그림 3은 전체 실험 참가자들의 시험 결과 분포를 나타내었다. CB 그룹에서는 전체 평균 6.97점(참여자 CB_04), LB 그룹에서는 6.70점(참여자 LB_04), SU 그룹에서는 5.20점(참여자 SU_10)이 가장 높았다. 시각장애인과 비장애인 참여자를 구분하여 비교한 결과 전체적으로 시각장애인 참여자(CB, LB)들의 평균 점수가 높게 나타났으나, 예외적으로 선천적 시각장애인 중 CB_07(40대 후반, 여성, 보청기 착용)의 경우 전체 참여자 중 가장 낮은 점수를 나타내었다.

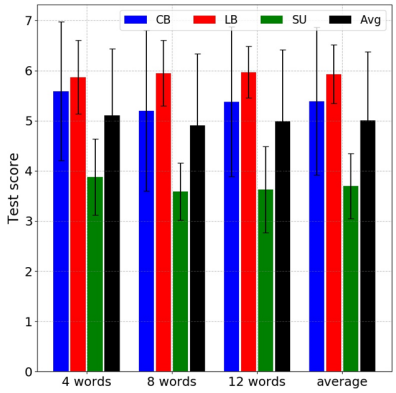

그림 4. 참여 그룹별 시험 결과
Fig. 4. Test score distribution of participants groups

그림 4는 전체 참여자들에 대하여 표 2에서 정의한 3가지 난이도에 대하여 그룹별 청각 해상도의 평균치 결과를 나타낸다. 실험 참가그룹 중에서 LB 실험군의 전체 평균이 5.93점으로 가장 높았으며, SU 실험군이 3.70점으로 가장 낮았다. 전체 참여자들의 청각 해상도 평균은 5.01점이었다. 축약된 일반 음향에 대한 청각 인식도를 시험한 결과, 전반적으로 시각장애인 그룹(CB, LB)이 비장애인 그룹(SU)보다 50% 정도 뛰어난 청각 해상도를 가짐을 확인할 수 있었다.

4. 축약된 VASS 신호의 지각 인식도 분석

기술의 적용 관점에서 시각 정보에 대한 청각 해상도의 실험 결과가 실제 VASS 신호에 대한 사용자의 지각 인식도에 연관성이 있는지 확인할 필요가 있다. 따라서 일반 사운드의 축약 정도에 대해 참여 그룹별 실험에서 얻은 청각 해상도의 분석 결과를 실제 VASS 신호에 적용하여 사용자의 지각 인식도를 분석하였다.

4.1 실험 방법

축약된 VASS 신호의 지각 인식도 분석은 일반 음향의 축약도 변화에 대한 청각 해상도 분석 실험에 참가하였던 참가자 중 희망자 18명을 포함하여 총 35명(여성 15명)의 실험 참가자를 모집하여 진행하였다. 실험 참가자들은 표 3에서와 같이 선천·후천적 전맹 시각장애인(CB, LB)으로 각 15명씩 2개 그룹과 비장애인(SU) 1개 그룹(5명)을 포함해서 총 3개의 실험군으로 구성하였으며, vOICe를 활용하여 인코딩된 VASS 신호의 길이에 따라 0.5초(A그룹), 1초(B그룹), 2초(C그룹)의 3개 실험군으로 랜덤하게 배정하였다. 특히 일반 사운드의 축약 정도에 대한 참가 실험군에 대한 실험에서 얻은 청각 해상도 결과를 반영하여 A그룹의 학습에 사용하는 VASS 신호의 축약 비율을 50%로 설

정하였다. 실험의 주요 목적이 각기 다른 길이로 인코딩된 VASS 신호에 대한 지각 인식에 대한 검증이므로 시각장애인 그룹(CB, LB) 중심으로 진행하였으며, 5명으로 구성된 SU 실험군은 다른 그룹과 인지도 분석을 위한 참조 지표의 목적으로 일반 길이(1초, B그룹)의 VASS 신호에 한정하여 실험을 수행하였다.

표 3. 실험 참가자의 그룹 분류
Table 3. Group classification of participants

Participants	Group A (0.5 sec)	Group B (1 sec)	Group C (2 sec)	sum
CB	5	5	5	15
LB	5	5	5	15
SU	0	5	0	5
sum	10	15	10	35

축약된 VASS 신호의 지각 인식도 분석을 위한 실험과정은 그림 5에 나타내었다. 실험 참가자의 대부분이 시각장애인으로 구성되었으므로 실험이 진행되는 과정은 통제된 실험실 환경에서 수행되었으며, 전체 실험과정은 기관생명윤리위원회의 검증과 승인하에 진행하였다(승인번호: KNU_HR2020010).

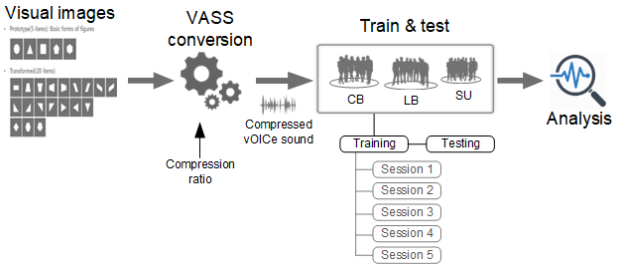

그림 5. VASS 신호의 지각 인식도 시험 절차
Fig. 5. Procedure for perception test of VASS signals

실험에 사용한 시각 이미지는 축약된 VASS 신호의 지각 정도를 분석하기 위한 목적을 최대화하기 위하여 단순한 형태의 도형으로 선정하였으며, 그림 6과 같이 원, 정삼각형, 정사각형, 오각형, 육각형으로 구성된 기본도형 5개와 기본도형을 회전시키거나 변환시킨 변형 도형 20개로 구성하였다.

축약된 VASS 신호 생성은 그림 5에서와 같이 시각 이미지를 청각 신호로 변환하기 위해 사용하는 vOICe 프로그램에 3가지 유형의 인코딩 길이(0.5, 1, 2초)를 제어 변수로 사용하여 3가지 다른 길이의 사운드스케이프를 생성하였다. 사운드 변환을 위한 vOICe 프로그램에 사용한 주파수는 500Hz에서 5000Hz 구간을 64개의 선형 로그 스케일로 할당하였으며, 생성된 VASS 신호는 소니 MDR7506 헤드폰으로 실험 참가자에게 전송하였다.

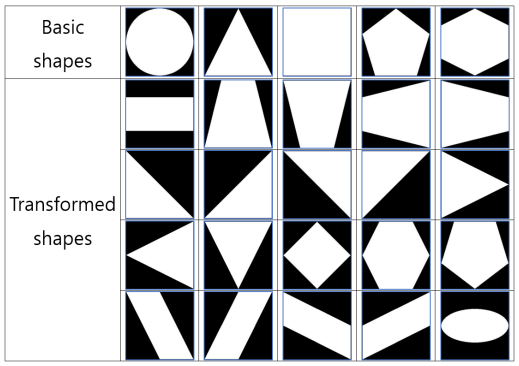

그림 6. 학습 및 시험에 활용한 이미지
Fig. 6. Visual images used for training and testing

축약된 VASS 신호의 지각 인식도 분석을 위한 학습 과정은 5회의 세션에 걸쳐 진행하였으며, 각 세션은 학습 및 평가 단계로 구성되었다. 학습 진행을 위한 세션별 간격은 최소 2일 이상 최대 4일 이내로 설정하였다. 각 세션에 사용된 이미지는 5개의 기본 모양 이미지와 변형된 모양 이미지에서 무작위로 선택된 10개의 이미지로 구성하였다. 이렇게 선택된 15개의 이미지는 의사 무작위(Pseudo random)로 10회에 걸쳐 반복하여 세션 당 총 150개의 훈련 세트로 훈련이 이루어지도록 프로그래밍하였다. 기존 연구에서는 실험 참가자들이 신호의 변환 알고리즘이나 방법을 이해하기 위해 사전 설명이나 절차에 대한 사전 교육을 진행하였으나, 본 연구에서는 이러한 사전 교육이나 설명 없이 VASS 신호를 학습하였다.

학습 세션 동안 실험 참가자들이 청취하는 VASS 신호들은 그림 7과 같은 순서로 생성하였다. 참가자 대부분이 시각장애인들이므로 헤드폰으로 전달되는 VASS 신호가 언제 시작하는지 알 수 있도록 시작 음을 먼저 알려주고 1초 후에 VASS 신호를 청취하도록 하였으며, VASS 신호 0.5초 후에 종료음으로 학습 사운드가 끝났음을 알려주었다. 이와 별도로 각 VASS 신호의 시작 부분에 참가자에게 사운드 시작을 다시 한번 알리기 위해 50ms 길이의 클릭 사운드를 추가하였다. VASS 신호 청취 후 시각장애인들이 소리에 대응하는 이미지를 손으로 탐색하여 이해할 수 있도록 해당 도형을 점자로 인쇄하여 추가로 제공하였다.

각 세션에서 학습이 끝나면 실험 참가자의 학습 효과를 평가하기 위해 학습하였던 15개의 도형 외에 학습하지 않은 10개의 도형을 인지할 수 있는지 확인하기 위하여 전체 25개 이미지를 의사 무작위 순서로 5회 반복하여 세션 당 총 125개의 테스트 예제를 생성하였다. 각 학습 세션의 평가 단계에서 각 참가자에게는 의사 랜덤으로 재생되는 VASS 신호(1회 재생)와 점자로 인쇄된 4개의 도형 예제를 제공하였다. 헤드폰을 통해 전송되는 VASS

신호를 청취한 실험 참가자는 손으로 점자 출력물을 탐색하여 모양을 확인하고 제시된 4가지 도형에서 소리에 대응하는 이미지를 선택하도록 하였다.

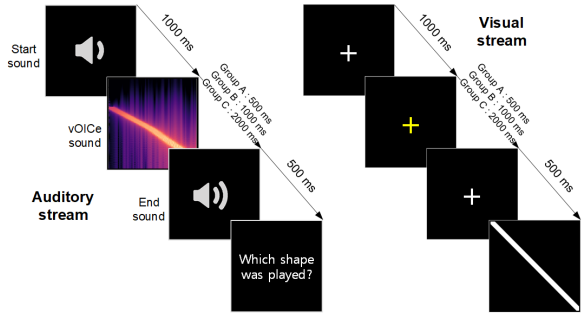

그림 7. 학습 세션의 VASS 신호 순서
Fig. 7. Sequence of VASS signals for training sessions

학습이 진행되는 동안 각 세션의 교육 시간과 각 세션의 평가에 대한 답변이 기록되었으며, 세션 평가 중에는 실험자에게 선택에 대한 정답 여부의 피드백을 제공하지 않았다. 참가자들은 vOICe의 신호 변환에 관련하여 사전 훈련이나 설명 없이 실험을 수행하였으므로 도형의 선택에 충분한 시간을 제공하였으며, 평가에서 응답 시간은 인식도 분석에 고려하지 않았다.

VASS 신호의 학습 과정에서의 평가는 0.25의 기회 수준(Chance level)으로 각 테스트의 정답 비율을 계산하여 측정하였으며, 학습 과정에 대한 모든 결과 분석은 참가자 및 그룹의 평균 테스트 점수로 산정하였다.

4.2 실험 결과

축약된 VASS 신호에 대하여 학습 세션이 진행되는 동안 실험 참가자 그룹들의 세션당 평균 훈련 시간의 변화와 각 훈련 세션에서 수행한 평가의 정답 비율을 그림 8에 나타내었다. 그림 8(a)는 총 5번의 세션에 걸쳐 진행된 학습 과정에서 각 참가자 그룹별로 VASS 신호를 학습하는 데 걸리는 평균 시간의 변화를 보여준다. 그래프를 통해 인코딩된 VASS 신호의 길이와 관계없이 학습이 진행됨에 따라 각 세션에서 학습에 필요한 학습 소요 시간이 현저하게 감소하는 현상을 모든 참가자 실험군에서 확인할 수 있다.

또한, 그림 8(b)는 각 세션의 학습 단계 이후에 수행된 평가에서 점자로 인쇄된 네 가지 모양에서 VASS 신호에 대응하는 모양 정보를 선택하는 인지도의 그룹별 평균값을 나타낸다. 각 그룹의 참가자가 VASS 신호로 전달되는 모양을 인식하는 정도는 학습이 진행됨에 따라 점차 증가하였다. 주목할 결과는 실험 참가 그룹 중에서 LB 실험군이 VASS 신호의 길이에 상관없이 학습 과정이 진행됨에 따라 가장 높은 성능 향상을 보였다. 이러한 현상은 LB 실험군의 경우, 시각 정보에 대한 이전 경험 또는 기억에 잔존하는 시각적 이미지의 영향인 것으로 판단된다.

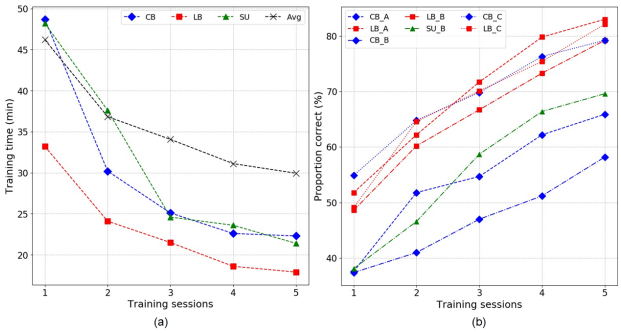

그림 8. 학습 세션당 훈련 시간 및 인지도의 평균 변화
Fig. 8. Change in average training time and the proportion of correct answers for each training session

특히, 기존 연구들과 달리 본 실험에서 발견된 흥미로운 특징은 인코딩된 VASS 신호를 기존 VASS 신호의 절반으로 압축하여 훈련을 받은 A 그룹의 참가자들이 다른 그룹의 실험 참가자들과 동일한 학습 결과의 변화 경향을 보였다는 점이다. 이러한 결과는 사운드스케이프 방식을 사용하는 기존 VASS 기술의 한계였던 시각-청각 신호 변환을 두 배로 향상할 수 있는 가능성을 제시한다.

총 5회에 걸친 학습 세션 후에는 축약된 VASS 신호의 지각 정도를 분석하기 위하여 최종적으로 테스트 세션을 진행하였다. 테스트 세션은 실험 참가자가 학습 과정에서 듣지 못한 다른 그룹들의 두 가지 길이의 VASS 신호에서도 이미지 정보를 인식할 수 있는지 확인하기 위한 목적으로 교차 테스트를 수행하였다. 이러한 교차 테스트를 통해 VASS를 기술을 위한 훈련 절차 또는 프로토콜에 관한 중요한 단서를 찾을 수 있을 것으로 기대하였다.

그림 9는 교차 테스트 수행 후, 각 실험 참가군의 평균 지각 인식도 결과를 나타내었다. 그림에서 점선으로 표시한 수평선은 5회 학습 수행 후 평가한 실험 참가군들의 평균 인식도를 의미한다.

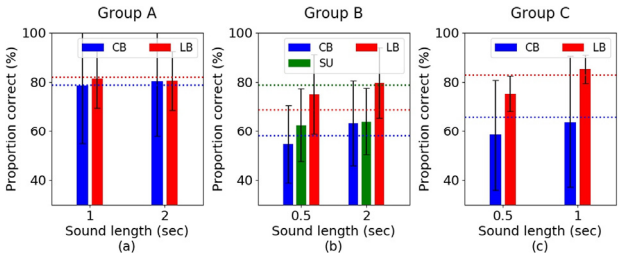

그림 9. 교차 테스트에 대한 실험 그룹의 평균 인식도
Fig. 9. Averaged performance of the crossover test

교차 테스트 결과, 학습 세션 동안 사용한 VASS 신호의 길이와 관계없이 모든 실험 참가그룹의 인식도는 기회 수준 이상을 나타내었다. 학습 세션에서 가장 높은 인식도를 나타내었던 LB 실험군의 경우, VASS 신호의 길이와 관계없이 일관되고 높은 인식도를 나타내었다.

또한, 그림 9의 결과는 가장 압축된 VASS 신호로 학습한 A그룹의 실험 참가자들(CB, LB 실험군)은 소리의 길이와 관계없이 높은 인식도를 보였다. 이와 대조적으로, 가장 긴 시간으로 인코딩된 VASS 신호로 학습한 C그룹의 CB 실험군은 짧은 길이의 VASS 신호로부터 시각 정보를 인식하는 데 어려움을 나타내었다. 이러한 현상은 정상적인 시간적 길이의 VASS 신호로 학습 후, 더 압축된 소리를 들었을 때 SU 실험군들보다 더 낮은 성능을 보인 B그룹의 CB 실험군의 성능 결과에서 더 명확하게 설명된다. 따라서 가능하면 짧게 인코딩된 VASS 신호를 사용하여 학습하는 것이 더 효율적이라는 사실을 교차 테스트의 결과에서 확인할 수 있다.

5. 결론 및 향후 연구

VASS 기술은 시각 정보를 청각 신호로 변환하여 전달함으로써 시각장애인이 제한적이지만 타인의 도움 없이 공간을 지각·인지하고 행동할 가능성을 제시하였으나 여전히 실제 생활에서 적용되기에는 어려움이 있다. 본 연구에서는 효율적인 VASS 기술 개발을 위한 선행적 연구로 시각장애인의 청각 해상도와 지각 인식도 분석을 위한 실험을 2단계로 구성하고 실험 결과를 제시하였다.

먼저 축약된 일반 음향 신호를 이용하여 청각 해상도를 분석하였으며, 실험 결과 시각장애인은 비장애인보다 청각 감도가 최대 50% 더 높은 것을 확인할 수 있었다. 또한 청각 해상도 분석 결과를 활용하여 결정한 압축비를 VASS 인코딩에 적용한 후 7개의 실험군으로 구성된 실험 참가자들을 대상으로 청각 신호를 통해 전달되는 시각 정보를 인지하는지를 확인하는 실험으로 확장하였다. 다양한 실험 참가자 그룹에 대하여 확장 실험을 수행한 결과, 학습에 사용하는 VASS 신호의 길이를 50% 줄일 수 있음을 확인하였다. 이러한 결과는 기존 VASS 기술의 신호 전달에 걸리는 지연 시간을 절반으로 줄여 200%의 성능 향상이 가능함을 의미한다.

본 연구는 시각장애인의 청각 해상도 분석을 통하여 VASS 기술의 개선 가능성을 확인하였지만, 실험 결과의 일반화를 위해서는 추가적인 연구가 필요하다. 특히, 학습 과정에서 전반적으로 실험 참여자들의 학습 속도가 단축되고, 지각 인식도가 향상함을 확인하였다. 그러나 일부 고연령 실험 참가자의 경우, 일반적으로 학습 효과가 두드러지게 나타나지 않는 특성이 나타났다.

따라서 추후 연구에서는 나이에 따른 음역과 수용도를 확인하여 모든 연령층에서 학습 가능한 VASS 신호의 생성 또는 각 연령대의 특성에 맞는 VASS 신호의 전달에 필요한 기술 개발이 필요할 것으로 보인다. 추가로 본 연구의 결과들은 효율적인 VASS 신호의 학습을 위한 시각 정보의 선택, 학습을 위한 세션의 최적 시간 산출 및 충분한 시각 정보의 지각에 필요한 학습 세션 추정 등 관련 분야의 연구와 VASS 기술의 체계적인 훈련 프로토콜 개발에도 활용할 수 있을 것으로 기대된다.

Conflict of Interest

저자는 본 논문에 관련된 어떠한 잠재적인 이해상충도 없음을 선언한다.

References

[1] K. D. Moon, M. Kim, C. Y. Jeong, Y. K. Park, S. Y. Shin, C. M. Oh, J. S. Park, and H. C. Shin, “Technological Trends in Sensory Substitution,” *Electronics and Telecommunications Trend*, vol. 34, no. 4, pp. 65-75, 2019.

[2] P. B. Meijer, “An experimental system for auditory image representations,” *IEEE transactions on bio-medical engineering*, vol. 39, no. 2, pp. 112-121, 1992.

[3] M. Auvray, S. Hanneton, and J. K. O’Regan, “Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with ‘the voice’,” *Perception*, vol. 36, no. 3, pp. 416-430, 2007.

[4] S. Maidenbaum, S. Abboud, and A. Amedi, “Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation,” *Neuroscience & Biobehavioral Reviews*, vol. 41, pp. 3-15, 2014.

[5] A´. Kristja´nsson, A. Moldoveanu, O´. I. Jo´hannesson, O. Balan, S. Spagnol, V. V. Valgeirsd´ottir, and R. Unnthorsson, “Designing sensory-substitution devices: Principles, pitfalls and potential 1,” *Restorative neurology and neuroscience*, vol. 34, no. 5, pp. 769-787, 2016.

[6] S. Abboud, S. Hanassy, S. Levy-Tzedek, S. Maidenbaum, and A. Amedi, “Eyemusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution,” *Restorative neurology and neuroscience*, vol. 32, no. 2, pp. 247-257, 2014.

[7] S. K. Kim and M. S. Shin, “A Study of the Speech Rate Preference of Laypeople,” *Journal of Speech-Language & Hearing Disorders,* vol. 25, no. 2, pp. 57-64, 2016.

[8] N. Lee, J. Shin, D. Yoo, and K. Kim, “Speech rate

in Korean across region, gender and generation," *Phonetics and Speech Sciences*, vol. 9, no. 1, pp. 27-39, 2017.
[9] A. D. Baddeley, N. Thomson, and M. Buchanan, "Word length and the structure of short-term memory," *Journal of verbal learning and verbal behavior*, vol. 14, no. 6, pp. 575-589, 1975.
[10] T. Park, S. Park, and T. Kim, "The Effects of Korean Lexical Characteristics on Memory Span," *Korean Journal of Cognitive Science*, vol. 17, no. 1, pp. 15-27, 2006.

저 자 소 개

김무섭 (Mooseop Kim)
1999년~현재: 한국전자통신연구원 책임연구원
1998년~1999년: LG종합기술원 연구원
2008년: 충남대학교 컴퓨터공학과 박사
1998년: 경북대학교 전자공학과 석사

관심분야 : Wearable computing, Activity recognition, Sensory substitution
ORCID Number : 0000-0003-4914-0584
E-mail : gomskim@etri.re.kr

정치윤 (Chi Yoon Jeong)
2004년~현재: 한국전자통신연구원 책임연구원
2018년: KAIST 컴퓨터공학부 박사
2004년: POSTECH 전기전자공학부 석사
2002년: POSTECH 전기전자공학부 학사

관심분야 : Computer vision, Pattern recognition, Machine learning, Sensory substitution
ORCID Number : 0000-0001-7089-2516
E-mail : iamready@etri.re.kr