

Effectiveness Analysis of Structured Training Program for Sensory Substitution Systems

YunKyung Park*, Kyeongdeok Moon*, Mooseop Kim*, So Jung Park**, Chi Yoon Jeong*

* ETRI, 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, KOREA

**Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, KOREA

** American Institutes for Research, 1400 Crystal Drive, 10th Floor, Arlington, VA 22202, USA

parkyk@etri.re.kr, kdmooon@etri.re.kr, gomsim@etri.re.kr, spark@air.org, iamready@etri.re.kr

Abstract— Visual-to-auditory sensory substitution technology transforms image into sound, enabling users to perceive visual information through hearing instead of seeing. Just as humans acquire perception abilities through training according to the developmental stage of their sensory organs after birth, a continuous and efficient training process is needed to acquire Visual-to-auditory perception ability. In this research, we aim to contribute to improving the effectiveness of training methods for visual-to-auditory sensory substitution technology through systematization. We developed training materials categorized by auditory characteristics and structured by difficulty level, along with a periodic, repetitive training process, to enhance the efficiency of visual-to-auditory sensory substitution technology training, and analyzed their effects. The study's results indicate that spaced repetition training and systematic provision of materials are effective for visual-to-auditory sensory substitution technology training.

Keywords— Visual-to-Auditory, Sensory Substitution, Training Method, Evaluation, Level of Difficulty

I. INTRODUCTION

The recent advancements in sensory substitution technologies, particularly in visual-to-auditory transformation, have got attention as tools that aid in the perception of the environment through auditory signals. Various algorithms that convert visual stimuli into sound have been developed, enabling users to interpret and interact with their environment based on auditory information. However, despite their potential, these technologies have not yet gained widespread adoption, largely due to the lack of a structured framework for effective training. To effectively use sensory substitution technologies, users need to develop new cognitive pathways for processing visual-to-auditory transformations. This process is not intuitive and can only be acquired through sustained training. Moreover, the ability to develop these skills greatly depends on systematic and continuous training, making the provision of a structured training framework essential. However, most research in this field focuses on algorithm development and assessing accuracy through simple repetitive training, without adequately addressing comprehensive training systems.

In this paper, we designed an experiment to evaluate the effects of structured materials, cumulative and repetitive procedure, and various forms of activities on sensory

substitution training by dividing participants into an experimental group and a control group. The experimental group was provided with training materials categorized by image characteristics and sequenced by difficulty and engaged in training activities that included understanding principles, comparing, and contrasting images. This training was designed to be cumulative and repeated at regular intervals. In contrast, the control group underwent simple repetitive training. The results indicated that the experimental group achieved a higher accuracy rate during training compared to the control group, but a decline in accuracy was observed in delayed post-tests following the completion of the training. This suggests that while the structured training approach validated its relative effectiveness, it had limitations in ensuring long-term memory retention. Therefore, further research is needed to improve training accuracy and support lasting memory retention. In chapter 2, we review visual-to-auditory transformation algorithms and the experimental studies used to evaluate them. Chapter 3 describes the design and execution of the experiment, focusing on structured sensory substitution training method. Chapter 4 evaluates and validates the effects of the structured training method in acquiring visual-to-auditory sensory perception. Chapter 5 discusses the issues identified in the study and suggests directions for future research.

II. RELATED WORKS

Sensory substitution technology is a method that translates one type of sensory information into another, enabling the replacement of senses and providing new pathways for information delivery. Notably, algorithms that convert visual information into auditory or tactile signals are the focus of active research. Among these, the representative sensory substitution algorithms and the experiments conducted to evaluate their performance are as follows.

vOICE is a sensory substitution algorithm that converts visual information into auditory signals. This algorithm scans images captured by a camera from left to right, translating brightness into sound volume and vertical position into pitch. This allows users to 'hear' and interpret visual information through sound. In experiments, participants trained to recognize grayscale images by converting them into sound demonstrated the ability to perceive shapes, positions, and other features of objects through this system [1][2].

Additionally, using the vOICe system, it was found that participants could recognize edges, shapes, depth, and even some color information through the converted auditory signals. The findings also suggested that long-term use of the system could enhance visual perception abilities [3].

EyeMusic is another sensory substitution algorithm that converts visual information into auditory information by transforming color data into musical notes. This allows users to perceive color and positional information through sound. Each color is represented by a specific instrument, with pitch and volume determined by the image's position and brightness. Experiments using EyeMusic involved tasks where subjects distinguished colors or located objects of specific colors, assessing color recognition accuracy and object search speed [4]. The vOICe-based VBTones is a system that converts visual information into auditory signals to assist visually impaired individuals in perceiving their environment. Additionally, by removing unnecessary information and applying filters during the image-to-sound conversion, more refined sounds were generated, improving recognition accuracy. Both visually impaired and sighted participants demonstrated high accuracy in the tests [5]. A method has also been proposed that maps visual environmental elements (e.g., coordinates, object types, sizes) to audio features related to musical tones (e.g., frequency, temporal duration, note variations). In this study, a Convolutional Neural Network (CNN) was employed to remove the background and recognize objects, which were then converted into audio signals based on specific timing and pitch. This method demonstrated higher training time efficiency compared to VBTones and achieved an average accuracy of 88.05% in real object recognition [6].

These studies have demonstrated that sensory substitution algorithms can effectively convert visual information into alternative sensory modalities, allowing users to perceive and interact with the information through these different sensory channels. However, for these technologies to be effectively used in daily life, long-term training and adaptation are essential, making a systematic training program crucial to support the process. This study aims to assess the effectiveness of structured training materials and methods in developing a sensory substitution training program.

III. DESIGN AND IMPLEMENTATION OF THE EXPERIMENT

The experiment was designed and conducted to evaluate whether structuring training materials, diversifying training activities, and implementing cumulative and repetitive training processes contribute to improving the effectiveness of sensory substitution training. The study was conducted in accordance with the ethical standards laid out by the Institutional Review Board (IRB) of Korea University. The research protocol was reviewed and approved by the IRB (Approval No. KUIRB-2023-0301-01).

A. Sensory Substitution Algorithm

We used the vOICe algorithm to convert visual images into auditory input. The algorithm scans the image column by

column, encoding vertical positions as sound frequencies and horizontal dimensions as time. Each pixel's intensity is converted into sound volume, creating an audio representation of the image. The visual image was divided into 64 rows and 64 columns. Each column of 64 pixels corresponded to sound frequencies ranging from 80 to 7,600 Hz, with higher frequencies representing higher vertical positions. The image was scanned from left to right at 1.05 seconds per frame, and the sounds were generated simultaneously for all pixels in a column. The loudness was based on brightness, with white being the loudest and black being silent. Sixteen gray levels were used to represent intermediate values, and click sounds indicated the start and end of the converted sound [7][8].

B. Experiment Design

1) Structuring Training Materials: In this study, 23 black-and-white shapes were selected as training materials, and the difficulty of each shape was determined based on accuracy rates from previous research. The shapes were arranged in order of increasing difficulty and then organized into six modules according to the similarity of their sound patterns. Module 1 consists of shapes where a consistent interval is maintained between high and low pitches. Module 2 includes shapes where one pitch remains constant while the other changes in a consistent direction. Module 3 consists of shapes where one pitch remains constant while the other changes in one direction and then reverses direction. Module 4 is composed of shapes where one pitch remains constant while the other changes in the opposite direction. The sound pattern where the interval between the high and low pitches increases linearly and then decreases. Module 5 consists of shapes where the pitch changes simultaneously along the horizontal axis. Finally, Module 6 consists of shapes with patterns where the pitch converges at or near the midpoint of the overall pitch range, or diverges towards the extremes. The shapes categorized according to the similarity of their sound patterns are presented in Table 1.

TABLE 1. STRUCTURED SHAPES

Module	Shape	Module	Shape
1		4	
2		5	
3		6	

2) Diversification of Training Activities: Training activities included sequential playback of shape sounds, comparison and contrast of sounds for presented shape pairs, as well as quizzes, reviews, and tests. For the test, the sound of a specific shape was played, and the trainee was asked to select one of the four presented choices.

3) Cumulative and Repetitive Training Process: The training process consists of six sessions, with modules arranged to facilitate step-by-step training, starting from simpler shapes and progressing to more complex ones.

Additionally, it is designed for cumulative repetition, allowing participants to repeatedly review and practice shapes trained in previous sessions. The pre-stage of session 1 includes an explanation of the experiment, a music ability assessment, and a pre-test. In each session, various training activities are conducted for newly introduced shapes. Furthermore, sessions 2 through 4 include cumulative review of content trained in previous sessions. A midterm evaluation is conducted in session 3, and a post-evaluation in session 5. In session 6, a delayed post-evaluation is conducted to assess the effect of long-term memory retention. Throughout the entire training process, each shape is practiced a total of 40 times. Figure 1 illustrates the training procedure designed for the experimental group.

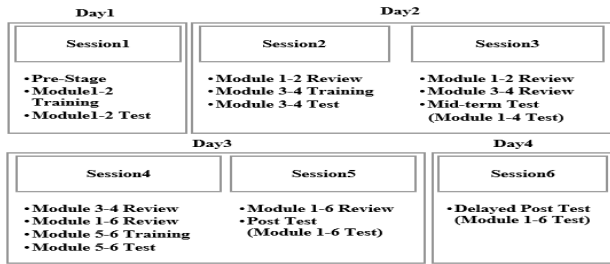


Figure 1. Training Procedure for Experimental Group

C. Experiment Implementation

The experiment involved 64 adults without visual or auditory impairments, who were randomly assigned to either an experimental group or a control group, with 32 participants in each. One participant from the experimental group withdrew, resulting in a final total of 63 participants. According to WWC standards, with a 2% dropout rate, the risk of bias due to participant withdrawal is negligible [9]. Musical ability was measured using the Micro-PROMS. Micro-PROMS is a test developed and validated to assess various music perception abilities within 10 minutes. The study results showed that Micro-PROMS has adequate internal consistency and test-retest reliability, as well as demonstrated convergent and discriminant validity. This tool can be administered online in a short time and has established itself as an important tool for objectively evaluating musical abilities [10]. The experimental group underwent a training according to the designed procedure, while the control group participated in an equal number of random, repetitive training sessions. The experiment took place twice a week over two weeks, with each session lasting 60 minutes.

D. Analysis of Experimental Results

Throughout the experiment, a total of seven tests, including quizzes, were conducted. In the test, participants were asked to listen to a sound corresponding to a shape image and then select the correct image from the four provided examples. A pre-test to assess the participants' baseline competencies, quizzes for reviewing the training content, a mid-term test, a post-test, and a delayed post-test were conducted. The post-test was conducted to evaluate the training performance, while

the delayed post-test was conducted to assess the extent of long-term memory retention

1) Sensory Substitution Accuracy Rate: Experimental results showed that both the experimental group and the control group demonstrated higher accuracy rates in all tests compared to the pre-test. Specifically, the accuracy rate improved by 78.14% in the experimental group and 60.38% in the control group compared to the pre-test. In the module tests conducted for each session, a relatively large disparity in accuracy rates between the experimental and control groups was observed in the higher difficulty Module 5-6 tests. This indicates that more systematic training is needed as the difficulty increases. In the delayed post-test conducted after the post-test, no statistically significant differences were observed between the experimental group and the control group. This suggests that while the training program was effective in the short term, it was not substantially superior to traditional methods in terms of long-term memory retention. Figure 2 shows the accuracy rates of the tests by group.

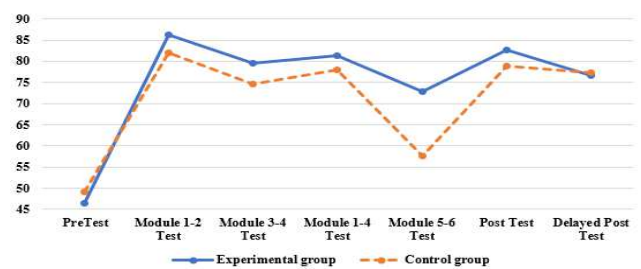


Figure 2. Accuracy Rates by Group

2) Effectiveness of Structured training Method in Sensory Substitution: To evaluate whether structured training is more effective than simple repetitive training and to validate its statistical significance, a regression analysis was conducted. To eliminate the effects of other variables that could influence the results, three models were defined and controlled for. Model 1 compares the performance differences between the experimental group that received structured training and the control group that received simple repetitive training. (1) Model 2 controls for music ability scores to more accurately assess the true effect of the training program. (2) Model 3 incorporates personal background variables to more precisely measure the impact of the training program. (3).

$$Y = \beta_0 + \beta_1(Treatment) + \epsilon \quad (1)$$

$$Y = \beta_0 + \beta_1(Treatment) + \beta_2(Music Ability) + \epsilon \quad (2)$$

$$Y = \beta_0 + \beta_1(Treatment) + \beta_2(Music Ability) + \beta_3(Gender) + \beta_4(Undergraduate Ratio) + \beta_5(Stem Ratio) + \epsilon \quad (3)$$

The analysis results indicated that Model 1 was not statistically significant across all tests. However, Models 2 and 3, which controlled for other variables to purely isolate the effects of the structured training, showed that the experimental group scored higher on both the mid-test and post-test, and these differences were statistically significant at the 95% confidence level. This suggests that the structured training program had a clear effect on performance in the mid-

test and post-test. For the delayed post-test, none of the models showed statistically significant results at the 95% confidence level. The regression analysis results for each model are presented in Table 2.

TABLE 2. REGRESSION ANALYSIS RESULTS

	Module 1-4 Test			Post Test			Delayed Post Test		
	Model1	Model2	Model3	Model1	Model2	Model3	Model1	Model2	Model3
Effect Size	0.2	0.52	0.55	0.25	0.55	0.57	-0.03	0.27	0.26
R ²			0.37			0.44			0.35

The adjusted accuracy results of the experimental and control groups, along with the original accuracy results, are shown in Figure 3.

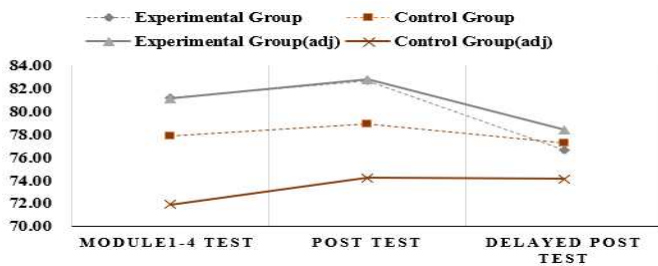


Figure 3. Adjusted Accuracy Rates by Group

3) Analysis of Response Time: In this study, accuracy rates were used to evaluate difficulty. Additionally, to determine whether response time could serve as a new metric for difficulty assessment, a correlation analysis and regression analysis were conducted between the accuracy rates and the average response times for each shape. The Pearson correlation coefficient between Average Response Time and Correct Rate is approximately -0.81, indicating a strong negative correlation between these two variables. Figure 4 visually represents the results of the regression analysis, with blue dots denoting the actual data points and the red line representing the regression line. The graph clearly illustrates that as the Average Response Time increases, the Correct Rate tends to decrease. The p-value of 3.03e-06 indicates that the effect of Average Response Time on the Correct Rate is statistically significant. Additionally, the R-squared value of 0.654 demonstrates the model's relatively high explanatory power.

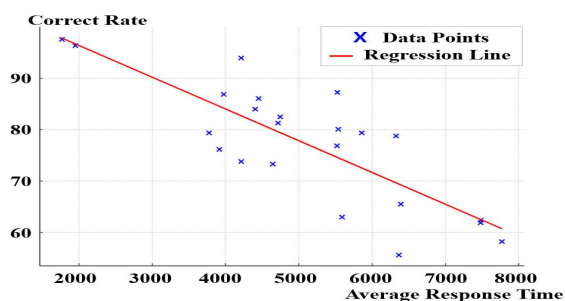


Figure 4. Regression Analysis of Average Response Time and Correct Rate

IV. CONCLUSIONS

This study designed a training process that includes cumulative repetitions, structured training materials based on difficulty levels, and various training activities to evaluate the effectiveness of systematic training in visual-to-auditory sensory substitution. The experimental results demonstrated that systematic training enhances the efficiency of visual-to-auditory sensory substitution training. These findings offer valuable insights into improving training methods for visual-to-auditory sensory substitution and assist users in adapting more effectively to this new sensory modality. However, the limited training time highlighted challenges in transferring the trained content to long-term memory. Therefore, ongoing research is necessary to assess the long-term effects of training and to develop strategies that facilitate the transfer of knowledge to long-term memory within shorter training periods. In particular, it is crucial to explore methods for resetting already acquired perceptual traits to further enhance perceptual abilities for alternative senses.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [24ZB1200, Fundamental Technology Research for Human-Centric Autonomous Intelligent Systems].

REFERENCES

- [1] Peter B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112-121, 1992.
- [2] A. Haigh, D. Brown, P. Meijer, and M. J. Proulx, "How well do you see what you hear? The acuity of visual-to-auditory sensory substitution," *Frontiers in Psychology*, vol. 4, p. 330, Jun. 2013. doi: 10.3389/fpsyg.2013.00330.
- [3] J. Ward and P. B. L. Meijer, "Visual experiences in the blind induced by an auditory sensory substitution device," *Consciousness and Cognition*, vol. 19, no. 1, pp. 492-500, 2010.
- [4] 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.
- [5] N. Riahi, S. F. Mirhoseini, and E. Mehrbani, "VBTones: A visual to auditory device for the blind," *International Journal of Advanced Computer Research*, vol. 7, no. 33, pp. 233-243, 2017.
- [6] E. Mehrbani, S. F. Mirhoseini, and N. Riahi, "A Novel Mapping for Visual to Auditory Sensory Substitution," *arXiv preprint arXiv:2106.07448*, 2021. [Online]. Available: <https://arxiv.org/abs/2106.07448>
- [7] M. Kim, Y. Park, K. Moon, and C. Y. Jeong, "Impact of Device and Environment on Visual-Auditory Sensory Substitution: A Comprehensive Behavioral Analysis Using the vOICe Algorithm," *IEEE Access*, vol. 12, pp. 90501-90510, July 2024.
- [8] M. Kim, Y. Park, K. Moon, and C. Y. Jeong, "Deep Learning-Based Optimization of Visual-Auditory Sensory Substitution," *IEEE Access*, vol. 11, pp. 14169-14180, Feb. 2023.
- [9] What Works Clearinghouse, "What Works Clearinghouse Procedures and Standards Handbook," *Institute of Education Sciences*, 2022.
- [10] H. Strauss, S. Reiche, M. Dick and M. Zentner, "Online assessment of musical ability in 10 minutes: Development and validation of the Micro-PROMS," *Behavior Research Methods*, pp. 1-16, 2023.



YUNKYUNG PARK received the B.S. degree in Mathematics Education from Korea University, South Korea, in 1987, and the M.S. degree in Telecommunications Engineering from Chungnam National University, South Korea, in 2006. She has been working as a Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, since 1987, where she is currently a Principal Researcher. Her current research interests

include wearable computing and sensory substitution.



KYEONGDEOK MOON received the B.S. and M.S. degrees in computer science from Hanyang University, South Korea, in 1990 and 1992, respectively, and the Ph.D. degree in information engineering from KAIST, South Korea, in 2005. From 1992 to 1996, he was a Researcher at the System Engineering Research Institute, where he worked on high-performance computing and clustering computing. Since 1997, he has been a Principal Researcher with the Electronics and

Telecommunications Research Institute, where he develops the home network middleware, deep learning for video analysis, framework for autonomously navigated ships, and human augmentation architecture. His research interests include sensory substitution technology, human augmentation technology, deep learning architecture, and autonomous ship



MOOSEOP KIM received the M.S. degree in electrical engineering from Kyungpook National University, South Korea, in 1998, and the Ph.D. degree in computer science and engineering from Chungnam National University, South Korea, in 2008. He was a Research Engineer at the Device and Materials Laboratory, Organic LED (OLED) Group, LG Electronics Institute of Technology (LG Elite), Seoul, South Korea, from 1998 to 1999. He has been with the Electronics and Telecommunications Research Institute

(ETRI), Daejeon, South Korea, since 1999, where he is currently a Principal Researcher. His current research interests include wearable computing, activity recognition, and sensory substitution.



SO JUNG PARK earned a Ph.D. in educational leadership and policy analysis from the University of Wisconsin-Madison in 2014, an M.S. in education policy from the University of Pennsylvania in 2008, and a B.A. in education (with a minor in public administration) from Korea University in 2005. Dr. Park's work focuses on policy and program evaluation in K-12 education, with an emphasis on enhancing teacher capacity, instructional practices, and educational conditions for disadvantaged students. Her research has

published in leading journals such as the American Educational Research Journal, Journal of Research on Educational Effectiveness, and Teachers College Record.



CHI YOON JEONG received the B.S. and M.S. degrees in electronic and electrical engineering from the Pohang University of Science and Technology, Pohang, South Korea, in 2002 and 2004, respectively, and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2018. He is currently a Principal Researcher with the Artificial Intelligence Laboratory, Electronics and Telecommunications Research Institute, Daejeon. His main research

interests include computer vision, pattern recognition, machine learning, and sensory substitution.