AUDITORY REPRESENTATION OF VISUAL STIMULI: MAPPING VERSUS ASSOCIATION

Orit Shenkar — School of Occupational Therapy, The Hebrew University of Jerusalem, Israel — shenkar@actcom.co.il

Patrice L. Weiss — Department of Occupational Therapy, University of Haifa, Israel — tamar@research.haifa.ac.il

Daniel Algom — Department of Psychology, Tel-Aviv University, Israel — algomd@freud.tau.ac.il

ABSTRACT

Two methods for representing visual images by sounds were explored. The analytic method used a rule-based representation by which values of continuous auditory variables defined each point in space. The metaphoric method governed the association of a unique sound with each image. The study demonstrated the usefulness and potential applicability of the analytic method.

1. INTRODUCTION

Sonification, the use of non-speech audio to display visual objects, provides tools for data analysis as well as for the detection of shapes. Previous studies have examined the ability to represent data graphically via sonification. Representing numeric data via auditory graphs was explored in research that studied how effectively information about statistical parameters of numeric data distribution could be conveyed [1]. Another study [2] suggested that the information conveyed by visual and auditory scatterplots is used very similarly in the two modalities. In the field of education [3], researchers explored the effectiveness of using sonified graphs of real data sets in disciplines to which students are exposed during academic courses. The effectiveness of auditory displays in science education was explored as well [4]. In the field of meteorology, a demonstration of sonification of weather records was studied [5].

Another group of studies explored the ability of people to detect geometrical shapes by sound [6, 7].

Regarding the detection of geometrical shapes, there has been no consistency in the acoustic parameters chosen to represent these shapes. In the present study, the acoustic parameters were defined based on the theory of the unity of the senses [8]. According to this theory, there are common dimensions to the different senses. We took advantage of the similarity of the senses in order to make a systematic match between the auditory to visual dimensions. Designing a sonified tool to detect a visual object was only one method we used in order to represent a visual image. We wanted to explore another method that might be helpful in detecting images. The second method used arbitrary sounds to represent those images. The objective of this study was to explore and compare two different methods for representing visual objects via non-speech audio.

2. METHOD

Two different representation methods were developed, based on the dual coding theory [9]. Both methods were aimed at representing visual stimuli via auditory output. The first method is called analytic mapping. This method is a rule-based mapping wherein images are systematically converted to sounds. A point in space is represented by a pair of acoustic parameters. The acoustic parameter of pitch represented location on the Y axis and the acoustic parameter of binaural time difference represented location on the X axis.

The second method is dubbed metaphoric. In this method, a unique sound is associated with each visual image. When images do not have "natural" sounds, then arbitrary sounds were selected to represent them. This arbitrary association of sounds to images is not uncommon. Consider, for example, Microsoft's use of sounds to announce certain events (e.g., the "tada" sound when the PC is turned on). The sounds used do not have "meaning" but, after hearing them repeatedly, users learn which event is associated with a specific sound.

The study reported in this paper explored the learning and use of the two different sonification methods. During the study, subjects learned how to match visual stimuli to auditory stimuli. On a trial, eight visual stimuli were presented on the computer monitor. Immediately following the presentation of the visual stimuli, a sound was played. The subject's task was to select the visual stimulus that matched the sound. The presentation of all eight visual stimuli and their matching sounds was defined as a cycle. The order of presentation (both visual and auditory stimuli) was random. The cycles were presented eight times.

Each subject participated in three sessions separated by at least 24 hours. At the beginning of the first session, the subject was given an explanation about the study and was introduced to one of the two mapping methods (analytic or metaphoric).

2.1. The analytic method

During the first session, a subject was exposed to either eight geometrical shapes or eight environmental objects. The task was to select the image that matched the sound that was played. The same sound was repeated until the subject made the correct matching. Then a second sound was played, and the subject was required to match it to its image. Trials continued until all of the eight sounds representing the eight images were correctly matched. A total of eight cycles was presented.

During the second session, the subject was exposed to the same set of stimuli. In order to test active learning, at the end of this session subjects were asked to draw the visual image of a sound that was played according to the rules of the analytic mapping (not one of the images presented during the session).

During the third session, a new set of stimuli was presented. We examined the extent of transfer of learning, i.e., the ability to apply the principles to novel images. Geometrical shapes were presented to subjects who had been exposed to environmental objects. Environmental objects were presented to subjects who had been exposed to geometrical shapes.

Because the analytic method used is new, let us illustrate it in somewhat more detail. First, we selected "natural" auditory variables to represent location along the horizontal and the vertical axis. Thus, horizontal location and/or change in horizontal location was represented by subjective change in the interaural space. The subject perceived a right-hand point by a sound localized close to the right ear. A left-hand visual point was heard as a sound close to the left ear. Sound location was created by interaural time differences (in the microsecond range, the sound closer to the lead ear). Vertical location and/or change in location was represented by change in pitch. The higher the point in space, the higher the pitch. The pitch of the tone was affected by corresponding frequency.

Consider the sound representation of a square, ABCD, by the analytic method. The constant-pitch tone at starting point A at the lower left starts moving to the right till B, at which point it remains at a constant right-hand location but increases in pitch. At point C, pitch is held constant (at high value) but the sound is felt moving to the left till point D. The sound then remains at this left-hand location, but decreases in pitch until it returns to A.

This "musical graph" represents the square ABCD and, by extension, any other two-dimensional figure. In the metaphoric method, an arbitrary sound is repeatedly paired with the square ABCD.

2.2. The metaphoric method

During the first session the subject was exposed to eight geometrical shapes and was asked to select the shape that matched the sound played. The same sound was repeated until the subject made the correct matching. Then a second sound was played, and the subject was required to match it to its image. The cycle continued until all of the eight sounds representing the eight images were correctly matched. A total of eight cycles were presented.

During the second session, the subject was exposed to the same set of stimuli. In order to test active learning, at the end of this session subjects were asked to draw an image of a sound that was played according to the rules of the analytic mapping (not one of the images presented during the session). They were given a brief introduction to the method of analytic sonification, but not an extensive training. The sound was played according to the rules of the analytic mapping.

During the third session, the subject's ability to transfer learning was examined. The subject was exposed to novel geometrical shapes and their arbitrary matching sounds.

This paradigm allowed us to examine evidence about learning and generalization of the two different mapping methods. They included (1) the learning curves obtained for each subject during the first two sessions, (2) the images drawn on the basis of the sound stimuli showing ability to apply the rules, and (3) assessing transfer of learning.

3. RESULTS

3.1. The analytic method

Twelve subjects (eight men and four women), paid volunteers, participated in the experiment. Their mean age was 16.5 years (SD=0.67). The participants took part in three sessions in which they were exposed to the analytic method. Analytic sounds were played in order to represent either geometrical shapes or environmental objects. During the first two sessions, half of the subjects were exposed to geometrical shapes and the other half were exposed to environmental objects. During the third session, subjects were exposed to novel images from the other group.

Figure 1 shows the mean number of incorrect responses made by the 12 subjects for each of the eight cycles during the first, second and third experimental sessions. As shown in this figure, there was a reduction in the number of incorrect responses from the first (circles) to the second (squares) sessions. During the third session (triangles), in which novel images were presented, the number of incorrect responses was much smaller than the number recorded during the first session, and similar to the number of incorrect responses recorded during the second session.

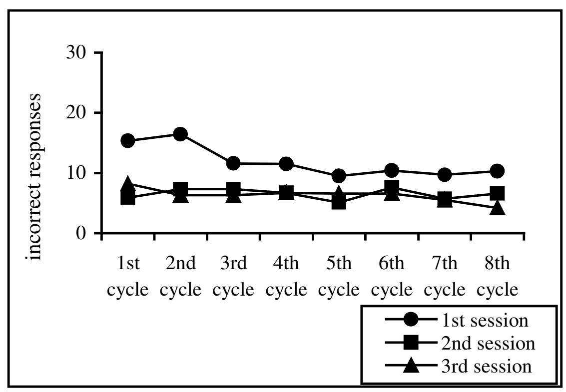

Figure 1. Mean number of incorrect responses obtained under the condition of the analytic method

3.2. The metaphoric method

Twelve subjects (six men and six women), paid volunteers, participated in the experiment. Their mean age was 16.4 years (SD=0.51). Participants took part in three sessions during which they were exposed to the metaphoric method. Metaphoric sounds were played to represent geometrical shapes. During the first two sessions, half of the subjects were exposed to shapes from set A and the other half were exposed to shapes from set B. During the third session, subjects were exposed to shapes from the other set.

Figure 2 shows the mean number of incorrect responses made by the 12 subjects for each of the eight cycles during the first, second and third experimental sessions. As shown in this figure, there was a reduction in the number of incorrect responses during the first session (circles), from the first to the eight cycles. This phenomenon is seen more clearly during the second session (squares). During the third session (triangles), when different shapes were presented, the number of incorrect responses was much higher then the number at the second session.

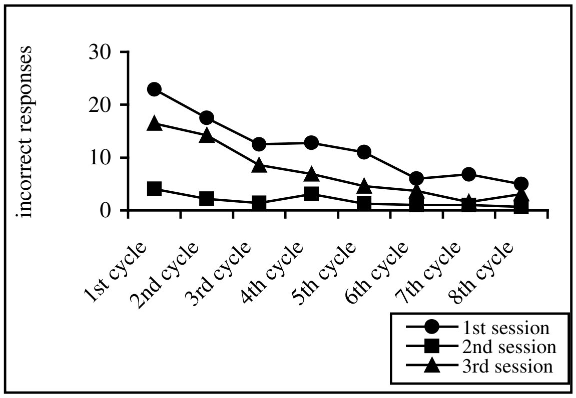

Figure 2. Mean number of incorrect responses obtained under the condition of the metaphoric method

The results show that learning takes occurs during both analytic and metaphoric mapping methods. This is demonstrated by the differences in the number of incorrect responses between the first and the second sessions, both for the analytic method ($p<.016$, Cohen's $d=2.68$) and the metaphoric method ($p<.016$, Cohen's $d=2.25$). Subjects' ability to transfer learning from one set of stimuli to a novel set of stimuli, on the other hand, took place only for the analytic method, in which there was a difference between the first and the last session ($p< .016$, Cohen's $d=2.69$).

4. CONCLUSIONS

The purpose of the study was to compare between two methods that use non-speech audio to represent visual objects. The first method, the analytic mapping, was a rule-based mapping. The second method, the metaphoric mapping, was an association-based mapping. The results show the potential of both methods in learning the language of auditory representation. Although this potential was demonstrated by both methods, there is a major difference between them. Under the condition of the analytic rule-based method, transfer of learning took place. Subjects took advantage of the principles that underlie this method and were able to apply them to a wholly different set of images. The analytic method is almost unlimited in its ability to represent two-dimensional visual objects. In contrast, the metaphoric method is limited. The arbitrary association of sounds to visual images can be applied only for a pre-defined set of images. Once these images were replaced, subjects were unable to apply any learned principle to different images.

The findings presented in this report were obtained from normally sighted subjects. As a feasibility study, we have also run the analytic method with a small group of subjects with visual impairment. The performance of these subjects supports the conclusions drawn from the healthy subjects. The visual impaired subjects showed good learning with the analytic method. We therefore anticipate that further exploration of the analytic method used in this study is important for sonifying materials for people with visual impairment.

5. REFERENCES

[1] J.H. Flowers, and T.A. Hauer, "The ear's versus the eye's potential to assess characteristics of numeric data: are we too visuocentric?" *Behavior Research Methods, Instruments, & Computers.,* vol. 24, no.2, pp. 258-264, 1992.

[2] J.H. Flowers, D.C. Buhman, and K.D. Turnage, "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples," *Human Factors.,* vol. 39, no. 3, pp. 341-351, 1997.

[3] T.L. Bonebright, M.A. Nees, T.T. Connerley, and G.R. McCain, "Testing the effectiveness of sonified graphs for education: a programmatic research project", in *Proc. Seventh Int. Conf. (IC)*, Helsinki, Finland, July 2001, pp. 62-66.

[4] J.M. Keller, E.E. Prather, W.V. Boynton, H.L. Enos, L.V. Jones, S.M. Pompea, T.F. Slater and M. Quinn, "Educational testing of an auditory display regarding seasonal variation of martian polar ice caps", in *Proc. Nineth Int. Conf. (IC)*, Boston, MA, USA, July 2003, pp. *212-215.*

[5] J.H. Flowers, L.E. Whitwer, D.C. Grafel, and C.A. Kotan, "Sonification of daily weather records: issues of perception, attention and memory in design choices", in *Proc. Seventh Int. Conf. (IC)*, Helsinki, Finland, July 2001, pp. *222-226.*

[6] H.M. Kamel, and P. Roth, "Graphics and User's Exploration via Simple Sonics (GUESS): providing interrelational representation of objects in a non-visual environment", in *Proc. Seventh Int. Conf. (IC)*, Helsinki, Finland, July 2001, pp. 261-266.

[7] K. van den Doel, D. Smilek, A. Bodnar, C. Chita, R. Corbett, D. Nekrasovski, and J. McGrenere, "Geometric shape detection with soundview" in *Proc. Tenth Int. Conf. (IC)*, Sydney, Australia, July 2004, pp. 1-8.

[8] L.E. Marks, The unity of the senses: Interrelations among the modalities. New York: Academic Press, 1978.

[9] A. Paivio, *Mental representations: a dual coding approach.* New York: Oxford University Press, 1990.