

EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution

Sami Abboud^{1*}, Shlomi Hanassy^{1*}, Shelly Levy-Tzedek^{1,2}, Shachar
Maidenbaum¹ and Amir Amedi^{1,2,3,4}

¹Department of Medical Neurobiology, The Institute for Medical Research Israel-
Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem
91220, Israel. ²The Edmond and Lily Safra Center for Brain Sciences (ELSC), The
Hebrew University of Jerusalem, Jerusalem 91220, Israel. ³The Cognitive Science
Program, The Hebrew University of Jerusalem, Jerusalem 91220, Israel.

⁴ Correspondence should be addressed to AA (amir.amedi@ekmd.huji.ac.il)

* Equal contribution

Corresponding author details:

Amir Amedi, Ph.D., Dept. of Medical Neurobiology - The Institute for
Medical Research Israel-Canada, Faculty of Medicine, The Hebrew
University of Jerusalem, Jerusalem 91220, Israel.

Phone: +972-2-675-7259, Fax: +972-2-675-8602, Lab website:

<http://brain.huji.ac.il/>, E-mail: amir.amedi@ekmd.huji.ac.il

Abstract

Purpose: Sensory-substitution devices (SSDs) provide auditory or tactile representations of visual information. These devices often generate unpleasant sensations and mostly lack color information. We present here a novel SSD aimed at addressing these issues.

Methods: We developed the EyeMusic, a novel visual-to-auditory SSD for the blind, providing both shape and color information. Our design uses musical notes on a pentatonic scale generated by natural instruments to convey the visual information in a pleasant manner. A short behavioral protocol was utilized to train the blind to extract shape and color information, and test their acquired abilities. Finally, we conducted a survey and a comparison task to assess the pleasantness of the generated auditory stimuli.

Results: We show that basic shape and color information can be decoded from the generated auditory stimuli. High performance levels were achieved by all participants following as little as 2-3 hours of training. Furthermore, we show that users indeed found the stimuli pleasant and potentially tolerable for prolonged use.

Conclusions: The novel EyeMusic algorithm provides an intuitive and relatively pleasant way for the blind to extract shape and color information. We suggest that this might help facilitating visual rehabilitation because of the added functionality and enhanced pleasantness.

Keywords

Sensory substitution devices, SSD, Blindness, Rehabilitation, Color, Pleasant sounds

1. Introduction

Imagine a blind person sitting in a garden in the beginning of spring, looking towards the sky and appreciating the rainbow. This scene is only vaguely imaginable by the blind community numbering approximately 39 million worldwide (World Health Organization, 2012).

Several attempts at converting visual information to auditory and tactile cues have been made in the past creating a family of Sensory Substitution Devices (SSDs) which enable shape and scene perception by the blind (Reviewed in Auvray et al., 2009). Among the prevalent veteran devices are the “Tactile Vision Sensory Substitution” (TVSS) that employed four hundred stimulators which vibrate against the skin of the back conveying shapes as captured from a television camera (Bach-y-Rita et al., 1969); and the “Tongue Display Unit” (TDU), generating electro-tactile pulses displayed on the tongue (Kaczmarek, 2011). Both of these devices implement a visual-to-tactile substitution. In the visual-to-auditory domain, “The vOICe” introduced the sweep-line technique that generated auditory representations, called soundscapes, which map the x-axis of captured images to the time domain and the y-axis to the frequency domain (Meijer, 1992). The “Prosthesis Substituting Vision by Audition” (PSVA), on the other hand, used an auditory frequency for each spatial location and generated the difference between each two subsequent images relying on camera movement by the user (Capelle et al., 1998). These devices rely on the improved auditory and tactile sensations of the blind (For examples see, Goldreich et al., 2003; Gougoux et al., 2004; Sathian et al., 2010) and aim to generate a shape-preserving representation of the environment as captured by an input device, e.g., a camera.

These veteran devices transformed the visual input, which originally contained colors, into grayscale images prior to the sensory-substitution leading to the loss of part of the original visual information. However, color information is an important feature in vision for tasks such as object recognition and scene segmentation (Bramão et al., 2011; Goffaux et al., 2005; Yip et al., 2002), even if possible to perform without it. Therefore, color is expected to facilitate performing such tasks while using sensory-substitution. This motivated the creation of SSDs that encode color information. Some conveyed color as a single point, e.g., Soundview (Doel, 2003), Haptic-Colour Glove (Kahol et al., 2006), while others conveyed horizontal lines, e.g., SeeColOr (Bologna et al., 2007). SeeColOr, for example, used timbre information to represent color hues, pitch to represent saturation and an addition of two extra timbres, simultaneous to the hue timbres, to represent luminance. The SeeColOr SSD delivers one horizontal line of 25 pixels requiring the user to determine their position on the X-axis using inter-aural cues in addition to active vertical scanning for complete scene construction.

Various psychophysical experiments have demonstrated the ability of blind users to perform object and letter identification and discrimination (Auvray et al., 2007; Brown et al., 2011; Striem-Amit et al., 2012a), recognize patterns (Poirier et al., 2007), pair socks by color (Bologna et al., 2009), localize visual stimuli (Proulx et al., 2008), reach for targets (Levy-Tzedek et al., 2012a; Levy-Tzedek et al., 2012b), extract depth cues (Renier et al., 2010), navigate, and avoid obstacles (Chebat et al., 2011; Kupers et al., 2010) and even perform complex tasks such as recognizing facial expressions (Striem-Amit et al., 2012b) using SSDs, proving their feasibility (Detailed review in Auvray et al., 2009; Reich et al., 2012) and leading to the conclusion that SSDs can and should be adopted for rehabilitation

either as standalone devices or as part of combined solutions addressing the various needs of the blind.

Despite all the above, SSDs are not widely used by the blind community. Hitherto, they have been considered expensive, cumbersome, not easy to operate by the blind and involved unpleasant sensations, e.g., visual-to-auditory SSDs generated unpleasant sounds, and stimulation by the visual-to-tactile SSDs as vibration or electrical currents were bothersome (Maidenbaum et al., 2012). In addition, the often unintuitive nature of these devices requires a vast amount of training hours during which the user slowly learns to generate a detailed mental image. This training period, although possible solely by self-experimentation, can be frustrating if undergone without guidance and gradual increases of difficulty.

In this work, we address user-experience and auditory pleasantness by utilizing a musical scheme which enables us to seamlessly integrate color information while conveying whole images using sounds. We further show proof-of-concept behavioral results using the EyeMusic in shape and color identification tasks, and the results of a survey tailored to assess the pleasantness of the generated soundscapes. In addition, we report the results of a task comparing the pleasantness of soundscape generated by the EyeMusic with those generated using the same images by the vOICe SSD (Meijer, 1992), which is the most used high-resolution grayscale SSD.

2. Materials and Methods

2.1 Technical description

The novel algorithm presented here is based on the image sweep-line technique where each image is processed column-by-column from left to right constructing a “soundscape”: a combination of sounds that

represents this image (an approach first introduced by Peter Meijer in the vOICe device; Meijer, 1992). Different from the vOICe however, the sounds underlying the generated soundscapes in our algorithm are recordings of musical instruments, where each instrument represents a different color (Fig. 1A). These instruments were chosen as they are clearly different from each other to avoid ambiguities, and as the time required for their notes to “build up” and reach the unique timbre which differentiates them from the rest of the instruments is minimal. Another consideration in designing the algorithm was the selection of the musical scale used for encoding the rows of each column. In soundscapes generated from images, in contrast to western music, there is a tendency for adjacent frequencies to be played together. Using a semitone western scale would then generate sounds that are perceived as highly dissonant (Wright et al., 2013). Therefore, to generate more pleasant soundscapes, we used the pentatonic musical scale, which is a subset of the western musical scale (See supplementary materials) that generates less dissonance when adjacent notes are played together. Another step to insure the pleasantness of the generated soundscapes was limiting the ceiling frequency to 1568 Hz (in contrast to the default 5000 Hz ceiling in the vOICe for instance). This was done because sounds with high energy in the 2500-5000 Hz range have been linked with unpleasantness (Kumar et al., 2008). In addition, a recent experiment investigating the perceived pleasantness of various SSD frequency ceilings showed that a ceiling of 2,500 Hz was selected as generating more pleasant soundscapes and a ceiling of 10,000 Hz was found to generate unpleasant soundscapes (Wright et al., 2013). The soundscape generating algorithm is illustrated in Fig. 1B and explained below.

2.1.1 Step-by-step Algorithm

- Capture an image from the connected camera/screen of the computer or load an image from a file (Fig. 1C).
- Resize the image to X=40 pixels (columns) by Y=24 pixels (rows; Fig. 2A).
- Run a color-clustering algorithm to get a 6-color image (Fig. 2A).
- Process the image column-by-column from left to right and create an auditory output for each column according to the following:
 - For each pixel in this column:
 - Set the musical instrument according to the pixel color.
 - Set the musical note (pitch) according to the Y-axis coordinate of the pixel.
 - Set the attenuation (volume) of the note according to the pixel luminance value.
 - Combine all the above pixels, that is, all the rows in the given column, into a single auditory output representing this column.
- Play the resulting audio file column by column, preceded by a cue sound signaling the beginning of the image.

2.1.2 Resulting audio file

The constructed audio file preserves both spatial information and image colors (Fig. 2A). The 2D image's X-axis is mapped to the time domain, i.e., pixels situated on the left side of the image will sound before the ones situated on its right side. Fig. 2B illustrates this in waveform. The Y-axis is mapped to the frequency domain, i.e., pixels situated on the upper side of the image will sound higher in frequency while those on its lower side will sound lower in frequency. Fig. 2C illustrates this in spectrogram

representation. The color information is represented by the timbre of the different instruments. This timbre, often called tone color, is characterized in part by the overtones above the fundamental frequency of each note. Fig. 2C illustrates the different overtone composition between instruments, i.e., colors.

The duration of the constructed audio file can be controlled by the user. A shorter duration will yield a higher refresh rate and should be used for simple images or highly dynamic scenes, while a longer duration will yield a slower refresh rate and should be used for complex scenes with a large amount of important details. With experience, users are expected to use higher refresh rates.

Further details are available in the Supplementary materials.

2.1.2 The EyeMusic

The EyeMusic is freely available for download with additional instructions at brain.huji.ac.il/EM.html

2.2 Experimental protocol

2.2.1 Participants

The study included twelve blind participants (7 congenitally blind, 4 early blind (vision lost by the age of 2) and one late blind (vision lost by the age of 28); 7 female and 5 male; mean age: 36 ± 7 years) and ten sighted blindfolded controls (5 female and 5 male; mean age: 25 ± 2 years). All participants had normal hearing and no known neurological conditions. All participants signed an informed consent form. The research protocol was approved by the Ethics Committee of the Hebrew University of Jerusalem, Israel.

2.2.2 Behavioral Experiment

Two identification tasks preceded by a training procedure were performed as a proof of concept for our algorithm. The participants were first introduced to the visual-to-auditory transformation concept and then exposed to 15 simple stimuli. An instructor guided each participant on decoding shape and color information from the audio stimulation generated by the EyeMusic. Each participant was trained on two sets of stimuli (40x24-pixel bitmap images). The first set was of white stimuli containing vertical and horizontal lines with variations in length, thickness and location (34 stimuli). The second set was of colored stimuli (red, green, blue and white) containing vertical lines, horizontal lines and combinations of them forming simple shapes and letters (33 stimuli). See Fig. 3 for examples. Participants completed approximately 2-3 hours of training prior to testing. The training was conducted using an implementation of the EyeMusic algorithm installed on a laptop (Fig 1C).

Shape test - Upon completion of the first set of stimuli, each participant took a test with 15 randomly ordered white stimuli, a subset of the first set, and had to answer the question “What shape is it?” when presented with each stimulus. All the sighted and 11 out of the 12 blind participants took part in this test.

Color test - Upon completion of the second set of stimuli, each participant took a test with 15 randomly ordered colored stimuli, a subset of the second set, and had to answer the question: “What color is it?” when presented with each stimulus. All participants took part in this test.

Twenty-one participants completed both shape and color training and testing, out of which 19 did so in one single session. The remaining 3 participants (2 blind and 1 sighted) completed the training and testing in two separate sessions, one for shape and the other for color. Finally, a

single participant only took part in the color training and testing and did so in one session.

While it is possible to alter the refresh rate of the algorithm, the stimuli used in the current training and testing were all generated using a constant refresh rate of 2 seconds per presentation. Task performance and response times were recorded.

2.2.3 Pleasantness survey

Participants were presented with a short survey upon the completion of the behavioral experiment. The survey consisted of two questions:

A) To what extent did you find the sounds, you have listened to, to be pleasant? (1 – Little extent/Not pleasant, 5 – Large extent/Very pleasant)

B) To what extent do you think that you would be able to listen to these sounds for prolonged periods? (1 – Little extent; 5 – Large extent)

All participants took part in this survey except the 5 sighted participants who were first to participate in the experiment and were not presented with the survey at all.

Additionally, we conducted a more objective measure of pleasantness directly comparing the pleasantness of the soundscapes generated by the EyeMusic with those generated by the vOICe (See Section 2.3).

2.2.4 Data Analysis

Wilcoxon rank sum test was used for assessing whether the test performance percentages were above chance level. The chance level for the shape test is $1/34$, where 34 is the number of stimuli presented. The chance level for the color test is $1/4$, where 4 is the number of colors used during the experiment. Wilcoxon rank sum was also used to assess whether either of the groups had significantly higher performance percentages than

the other. A t-test was used in order to assess if any of the groups had faster response times than the other.

2.3 Comparison of pleasantness task

2.3.1 Participants

This task included 23 participants; all signed an online informed consent form. No personal details about the participants were stored.

2.3.2 Task design

We chose the leading vOICe SSD for the comparison of pleasantness because, to the best of our knowledge, it is the SSD with most blind users. We designed an online task with 30 stimuli consisting of 3 categories: indoor scenes (Adapted from Quattoni et al., 2009), outdoor scenes (Adapted from Gould et al., 2009) and simple shapes that we designed. Each category contained 10 stimuli. Two-second soundscapes were generated from all stimuli using the EyeMusic and the vOICe SSDs in their default modes. In each of the 30 trials, the user was presented with a pair of soundscapes conveying one stimulus encoded using both SSDs, separated by a 2-second break. When the playback of both sounds was over, a scale of 11 bars was presented using which the participants rated the relative pleasantness of the two sounds played. The middle bar signified that both sounds were equivalently pleasant. The 5 left bars signified the extent to which the first sound was more pleasant, while the 5 right bars signified the extent to which the second sound was more pleasant. Trial order and soundscape order within trials were pseudo-randomly drawn for each participant.

2.3.3 Data analysis

User ratings were collected and scaled between -5 and 5, where -5 is the highest pleasantness score for a vOICe soundscape and 5 is the highest pleasant score for an EyeMusic soundscape. The mean score was calculated for each participant. A histogram of the means was plotted. For technical reasons, the last two trials by each participant were not recorded by the system and therefore were not included in the analysis.

3. Results

3.1 Behavioral experiment

Shape test – Correct response rate stood on $91.5 \pm 10.8\%$ for the blind group and on $86.7 \pm 8.9\%$ for the sighted group (mean \pm SD; Fig. 4A). High rates of correct responses were recorded for both groups, significantly above chance (Wilcoxon, $p < 0.00005$ for the blind and $p < 0.0001$ for the sighted). The difference between the blind and sighted groups was not significant (Wilcoxon, $p = 0.26$). Therefore, we also present the combined data across the entire participants (Combined correct response rate stood on $89 \pm 10\%$; Fig. 4A). Response time was 16.6 ± 13.9 sec for the blind group and 18.8 ± 13 sec for the sighted group. No significant group difference was observed (t-test, $p = 0.14$).

Color test – Correct answer rate stood on $85.6 \pm 14.4\%$ for the blind group and on $74.7 \pm 15.7\%$ for the sighted group (Fig. 4A). As in the shape task, these rates were significantly higher than chance level (Wilcoxon, $p < 0.00001$ for the blind and $p < 0.0005$ for the sighted). The difference between the blind and sighted groups was not significant (Wilcoxon, $p = 0.12$). Therefore, we also present the combined data across the entire participants (Combined correct response rate stood on $81 \pm 16\%$; Fig. 4A). Response time was 9.3 ± 9.3 sec for the blind group and 10.8 ± 9.5 sec for the

sighted group. No significant group difference was observed (t-test, $p=0.15$).

3.2 Survey of pleasantness

All participants rated the pleasantness of the soundscapes generated by the EyeMusic as pleasant to a medium extent and above, except one participant which gave a lower rate. The mean rate stood on 3.6 ± 0.9 out of 5 (mean \pm SD; Fig. 4B – gray bar). The vast majority of the participants rated the expected ability to listen to the soundscapes for prolonged periods as medium extent and above, except three participants which gave a lower rate. The mean rate stood on 3.5 ± 1.2 out of 5 (Fig. 4B – black bar).

3.3 Comparison of pleasantness task

All but two participants scored the soundscapes generated by the EyeMusic as more pleasant on average (Fig. 4C – Gray bars). One of these two participants scored the vOICe soundscapes as equally pleasant on average to the EyeMusic soundscapes and the other found the vOICe soundscapes to be slightly more pleasant on average (Fig. 4C – Black and white bars). Specifically, fifteen of the participants gave the EyeMusic a score of 3 and above on a scale from -5 (the highest pleasantness score for a vOICe soundscape) to +5 (the highest pleasantness score for an EyeMusic soundscape).

4. Discussion

We developed the EyeMusic, an SSD aiming to provide pleasant musical image representations preserving shape information and seamlessly integrating color information. We showed that encoded shapes and colors are interpretable by all 22 participants tested following as little as 2-3 hours of training. The participants included individuals spanning the

spectrum of vision loss, i.e., fully congenitally blind, early-blind and late-blind, in addition to sighted blindfolded individuals. A survey of pleasantness showed that the majority of participants found the sounds to be pleasant. Finally, most users found the soundscapes generated by the EyeMusic to be more pleasant on average than the ones generated by the vOICe SSD. This was shown using a variety of image types, stretching from very simple lines to daily indoor and outdoor complex scenes.

This study is a demonstration of abilities showing that it is possible to encode the basic building blocks of shape using the EyeMusic, and more importantly, these can be decoded from the generated soundscapes by blind users. Furthermore, the success in associating color to musical timbre holds promise for facilitating the representation of more complex shapes. Variability in performance can be observed in both shape and color tests and it is attributed to inter-participant differences. We believe that these differences are due to the short training period which, sufficient for achieving ceiling performance for some participants, could have been too short for others which found the algorithm less intuitive. Higher variability is observed in the color task, which was not easy for all participants, especially when decoding the color of vertical lines lasting for 50 msec.

The behavioral protocol we employed utilizes only simple shapes which are the basic building blocks of shapes in the system. Therefore, tasks of greater complexity and richer images should be assessed before drawing conclusions about the practical use of the device. However, we can anecdotally report that blind users of the system, following a larger number of training hours, were able to use a web camera as the main source for input images and distinguish between different types of food and drinks, e.g., identify the location and reach for the red apple from a bunch of green apples, or identify the location and reach for the soda

bottle of choice out of several options presented ([Movie S1](#)). This is facilitated by the added color information.

In the future, the EyeMusic can be used together with other visual restoration approaches (for examples of such approaches see Dagnelie 2012; Reich et al., 2012; Kasten et al., 1999) such as the evolving visual prostheses. It can be used in augmenting the capabilities of the “bionic-eyes” using information arriving from the same image, e.g., adding color, to create a unique and novel entity: an SSD-bionic eye hybrid. In other words, the prosthesis will provide visual sensation or qualia and the SSD will convey detailed visual information with higher resolution or enhanced color information (Reich et al., 2012). Importantly, we have already demonstrated that spatial information is shared between vision and the EyeMusic SSD, showing the potential in the use of an SSD to augment partial vision (Levy-Tzedek et al., 2012b). This approach may also augment other rehabilitation attempts of users with partial vision, as previously shown using the vOICe SSD (Meijer et al., 1992) for augmenting partial vision with higher resolution (Ward et al., 2010). In addition, the EyeMusic could also be used by the color-blind providing the missing color information.

A meta-analysis of several studies inspecting the contribution of color to object recognition reached the conclusion that color improves object recognition wherever it is an attribute of diagnostic value, i.e., yellow banana vs. yellow hair comb (Bramão et al., 2011). It has also been shown that color cues lead to better performance even in face recognition when shape cues are degraded (Yip et al., 2002). It is further hypothesized that the contribution of color in image perception aids low-level image analysis processes such as segmentation (Yip et al., 2002). Therefore, we anticipate that this novel SSD would facilitate object recognition with the help of the

provided color information, albeit a resolution lower than some veteran SSDs, e.g, The vOICe (Meijer, 1992). Users are expected to perform better in tasks such involving color-driven reaching (Levy-Tzedek et al., 2012a) and figure-ground segregation compared with veteran SSDs that do not provide color information. In contrast, tasks that require higher resolution such as assessing facial expressions (Striem-Amit et al., 2012b) could benefit from an SSD with a higher resolution, such as the vOICe, which will have the upper hand.

Furthermore, the survey results, showing that almost all participants scored the extent of pleasantness as medium to high, combined with the results of the comparison between the vOICe soundscapes and the EyeMusic soundscapes suggest that the EyeMusic could be a step forward in terms of user-experience. Notwithstanding, we wish to stress that the increased pleasantness and potential for prolonged use come at the expense of image resolution. Therefore, different devices could be optimal for different scenarios depending on the input characteristics. For example, SSDs with higher resolution yet less pleasant soundscapes could be used for tasks demanding finer detail, while more pleasant SSDs with lower resolution and additional visual features, e.g., color, could be used for tasks that require prolonged use or when the additional visual features are valuable to the user.

Previous studies investigated the role of pleasant music when treating patients with unilateral neglect after stroke (Chen et al., 2013; Sarkamo et al., 2012; Soto et al., 2009; Tsai et al., 2013). Findings show that free exploration of natural scenes included more fixation and viewing time in the neglected side when listening to pleasant music versus unpleasant music and white noise (Chen et al., 2013). In a picture scanning test, performance in finding items on the neglected side increased while

listening to classical music versus white-noise and silence (Tsai et al., 2013). It is hypothesized that pleasant music, inducing positive emotions and increased arousal reduces the clinical symptoms of unilateral neglect (Chen et al., 2013; Soto et al., 2009). The auditory soundscapes generated by the EyeMusic, currently far from being music, might still encourage users to use the system if subjectively perceived as pleasant.

Finally, the utilization of the EyeMusic and its added color information in the field of neuroscience could facilitate exploring several questions in the blind with the potential to expand our understanding of brain organization in general. It would be intriguing to explore the plastic changes associated with learning to decode color information from auditory timbre in the congenitally blind, who never experience color in their life. Particularly interesting, would be the comparison of the neural correlates of color as encoded by timbre between the different etiologies of blindness, e.g., congenital versus late-blindness. In addition, color, being a cue for figure-ground segmentation, could facilitate testing the invariance of sensory modality in the neural processing of contours.

To conclude, the novel EyeMusic algorithm provides an intuitive and relatively pleasant way for the blind to extract visual shape and color information using auditory soundscapes. We suggest that this might aid visual rehabilitation for two main reasons. First, the added color information is expected to facilitate object recognition and scene segmentation. Second, the enhanced pleasantness of the generated soundscapes holds promise for potential prolonged use. Finally, the EyeMusic is expected to be an interesting tool for future neuroscience research.

Acknowledgements

The authors wish to thank D. Tauber for running part of the behavioral protocol and filming Movie S1. This work was supported by a European Research Council grant to A.A. (grant number 310809); The Charitable Gatsby Foundation; The James S. McDonnell Foundation scholar award (to AA; grant number 220020284); The Israel Science Foundation (grant number ISF 1684/08); The Edmond and Lily Safra Center for Brain Sciences (ELSC) Vision center grant (to AA) and an ELSC fellowship (to SL); SA was supported by a scholarship from the Israeli Ministry of Science.

References

- Auvray, M., Hanne-ton, S. and O'Regan, J.K. (2007). Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with 'the vOICe'. *Perception*, 36, 416-30.
- Auvray, M. and Myin, E. (2009). Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cogn Sci*, 33, 1036-58.
- Bach-y-Rita, P., Collins, C.C., Saunders, F.A., White, B. and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221, 963-4.
- Bologna, G., Deville, B., Pun, T. and Vinckenbosch, M. (2007). Transforming 3D Coloured Pixels into Musical Instrument Notes for Vision Substitution Applications. *J Image Video Process*, 2007, 076204.
- Bologna, G., Deville, B. and Pun, T. (2009). Bioinspired Applications in Artificial and Natural Computation. Springer Berlin Heidelberg.
- Bramão, I., Reis, A., Petersson, K.M. and Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychol*, 138, 244-253.
- Brown, D., Macpherson, T. and Ward, J. (2011). Seeing with sound? exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception*, 40, 1120-35.
- Capelle, C., Trullemans, C., Arno, P. and Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans Biomed Eng*, 45, 1279-93.
- Chebat, D.R., Schneider, F.C., Kupers, R. and Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport*, 22, 342-7.

- Chen, M.C., Tsai, P.L., Huang, Y.T. and Lin, K.C. (2013). Pleasant music improves visual attention in patients with unilateral neglect after stroke. *Brain Inj*, 27, 75-82.
- Dagnelie, G. (2012). Retinal implants: emergence of a multidisciplinary field. *Current opinion in neurology*, 25(1), 67-75.
- Doel, K. (2003). Soundview: Sensing color images by kinesthetic audio. *Proc of ICAD*, 303-306.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. and Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Vis Cogn*, 12, 878-892.
- Goldreich, D. and Kanics, I.M. (2003). Tactile acuity is enhanced in blindness. *J Neurosci*, 23, 3439-45.
- Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R.J. and Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature*, 430, 309.
- Gould, S., Fulton, R. and Koller, D. (2009). Decomposing a scene into geometric and semantically consistent regions. *IEEE I Conf Comp Vis*, 1-8.
- Kaczmarek, K.A. (2011). The tongue display unit (TDU) for electrotactile spatiotemporal pattern presentation. *Sci Iran*, 18, 1476-1485.
- Kahol, K., French, J., Bratton, L. and Panchanathan, S. (2006). Learning and perceiving colors haptically. *Proc of ASSETS*, 173-180.
- Kasten, E., Poggel, D. A., Müller-Oehring, E., Gothe, J., Schulte, T., & Sabel, B. A. (1999). Restoration of vision II: residual functions and training-induced visual field enlargement in brain-damaged patients. *Restorative neurology and neuroscience*, 15(2), 273-287.

- Kumar, S., Forster, H.M., Bailey, P. and Griffiths, T.D. (2008). Mapping unpleasantness of sounds to their auditory representation. *J Acoust Soc Am*, 124, 3810-7.
- Kupers, R., Chebat, D.R., Madsen, K.H., Paulson, O.B. and Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proc Natl Acad Sci U S A*, 107, 12716-21.
- Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S. and Amedi, A. (2012a). Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. *Restor Neurol Neurosci*, 30, 313-23.
- Levy-Tzedek, S., Novick, I., Arbel, R., Abboud, S., Maidenbaum, S., Vaadia, E. and Amedi, A. (2012b). Cross-sensory transfer of sensory-motor information: visuomotor learning affects performance on an audiomotor task, using sensory-substitution. *Sci Rep*, 2, 949.
- Maidenbaum, S. and Amedi, A. (2012). *Plasticity in Sensory Systems*. Cambridge University Press.
- Meijer, P.B. (1992). An experimental system for auditory image representations. *IEEE Trans Biomed Eng*, 39, 112-21.
- Poirier, C., De Volder, A., Tranduy, D. and Scheiber, C. (2007). Pattern recognition using a device substituting audition for vision in blindfolded sighted subjects. *Neuropsychologia*, 45, 1108-1121.
- Proulx, M.J., Stoerig, P., Ludwig, E. and Knoll, I. (2008). Seeing 'where' through the ears: effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution. *PLoS One*, 3, e1840.
- Quattoni, A. and Torralba, A. (2009). Recognizing indoor scenes. *Proc CVPR IEEE*, 413-420.

- Reich, L., Maidenbaum, S. and Amedi, A. (2012). The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches. *Curr Opin Neurol*, 25, 86-95.
- Renier, L. and De Volder, A.G. (2010). Vision substitution and depth perception: early blind subjects experience visual perspective through their ears. *Disabil Rehabil Assist Technol*, 5, 175-83.
- Sarkamo, T. and Soto, D. (2012). Music listening after stroke: beneficial effects and potential neural mechanisms. *Ann N Y Acad Sci*, 1252, 266-81.
- Sathian, K., and Randall Stilla. "Cross-modal plasticity of tactile perception in blindness." *Restorative neurology and neuroscience* 28.2 (2010): 271-281.
- Soto, D., Funes, M.J., Guzman-Garcia, A., Warbrick, T., Rotshtein, P. and Humphreys, G.W. (2009). Pleasant music overcomes the loss of awareness in patients with visual neglect. *Proc Natl Acad Sci U S A*, 106, 6011-6.
- Striem-Amit, E., Cohen, L., Dehaene, S. and Amedi, A. (2012a). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. 76, 640-52.
- Striem-Amit, E., Guendelman, M. and Amedi, A. (2012b). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS One*, 7, e33136.
- Tsai, P.L., Chen, M.C., Huang, Y.T., Lin, K.C., Chen, K.L. and Hsu, Y.W. (2013). Listening to classical music ameliorates unilateral neglect after stroke. *Am J Occup Ther*, 67, 328-35.
- Ward, J., Meijer, P., 2010. Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and cognition* 19, 492-500.

World Health Organization, (2012). Fact sheet no. 282

Wright, T. and Ward, J. (2013). The evolution of a visual-to-auditory sensory substitution device using interactive genetic algorithms. *Q J Exp Psychol*, 66, 1620-38.

Yip, A.W. and Sinha, P. (2002). Contribution of color to face recognition. *Perception*, 31, 995-1003.

Figure captions

Fig. 1 – Concept, Algorithm & Setup – A) The Musical-instrument-to-color mapping. Specifically, Red is mapped to a Reggae organ, Green to a Rapman’s Reed, Blue to Brass instruments, Yellow to string instruments, White to a Choir and Black to silence. **B)** The visual-to-auditory sensory substitution algorithm. **C)** The EyeMusic application running on a laptop, showing sample stimuli. **D)** A sample setup with a camera mounted on sunglasses and bone-conductance headphones.

Fig. 2 – Samples – A) Sample images. Left column: Original images resized to 40x24 pixels. Middle column: Color-clustered resized images. Right column: Black and White versions of the input images. **B)** Left: Sample input image. Right: A waveform representation of the resulting audio file generated by the algorithm. **C)** Left: Sample input images. Right: A spectrogram representation of the resulting audio file generated by the algorithm.

Fig. 3 – Experiment stimuli – Left: Examples of stimuli used during the Shape Task. Right: Examples of stimuli used during the color task.

Fig. 4 – Results – A) Behavioral results of the shape and color tests. Shape test: n=10 sighted, n=11 blind participants and n=21 for both groups. Color test: n=12 blind and n=10 sighted participants and n=22 for both groups.

Mean percentage of correct responses \pm SD. Horizontal line depicts the chance level in each task. **B)** The results of the survey conducted after the shape and color tests. A total of 17 blind and blindfolded sighted participants rated the extent of pleasantness and ability to listen to the soundscapes for prolonged periods. Mean rate \pm SD for each question is plotted. **C)** A histogram of the average scores per user binned into 11 scores, showing how many participants fell in each bin. Positive scores (gray bars) indicate that EyeMusic soundscapes were perceived as more pleasant while negative scores (black bar) indicate that the vOICe soundscapes were perceived as more pleasant. The white bar indicates a neutral average score between the two SSDs.

Figures

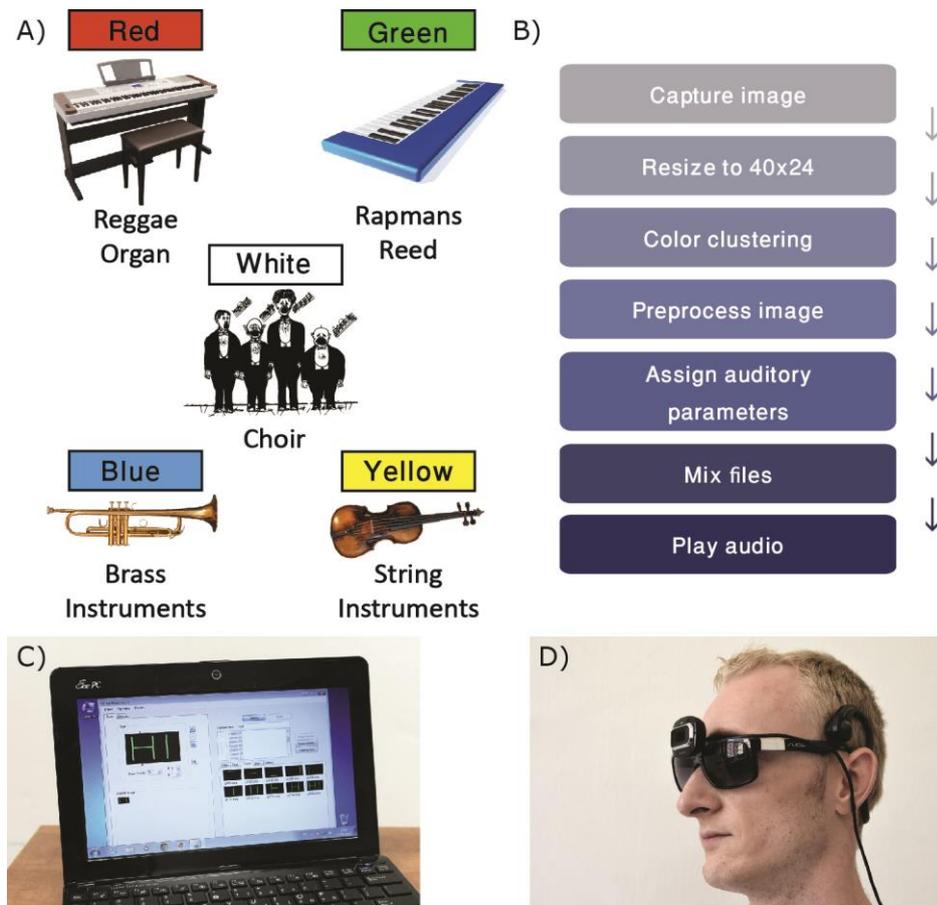


Fig. 1 – Concept, Algorithm & Setup

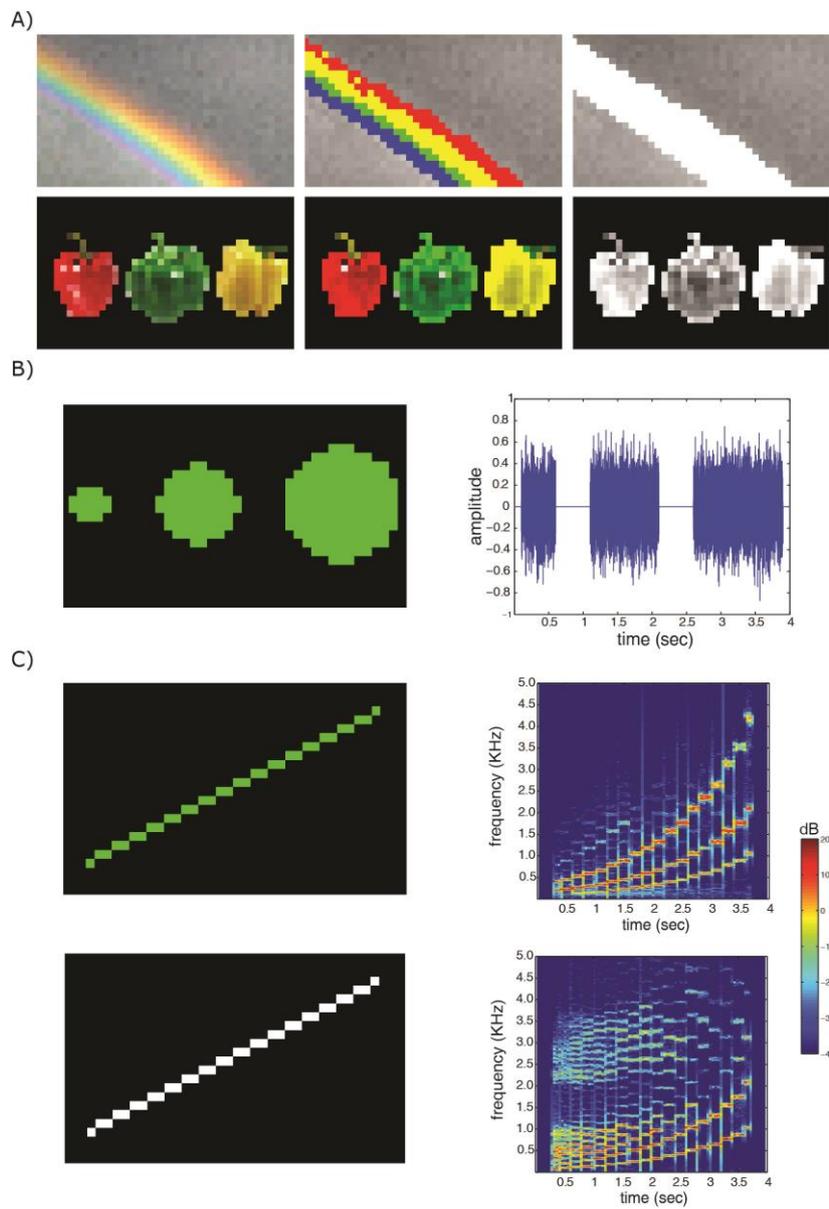
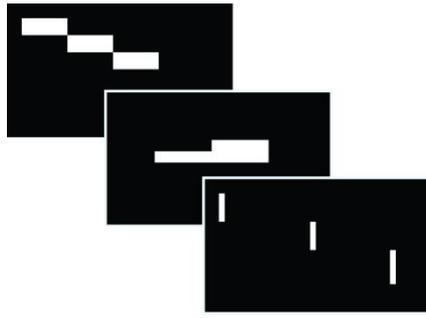
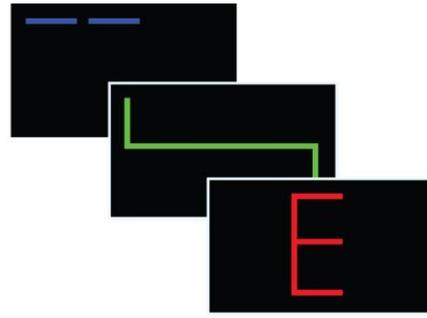


Fig. 2 – Samples



What shape is it?



What color is it?

Fig. 3 – Experiment stimuli

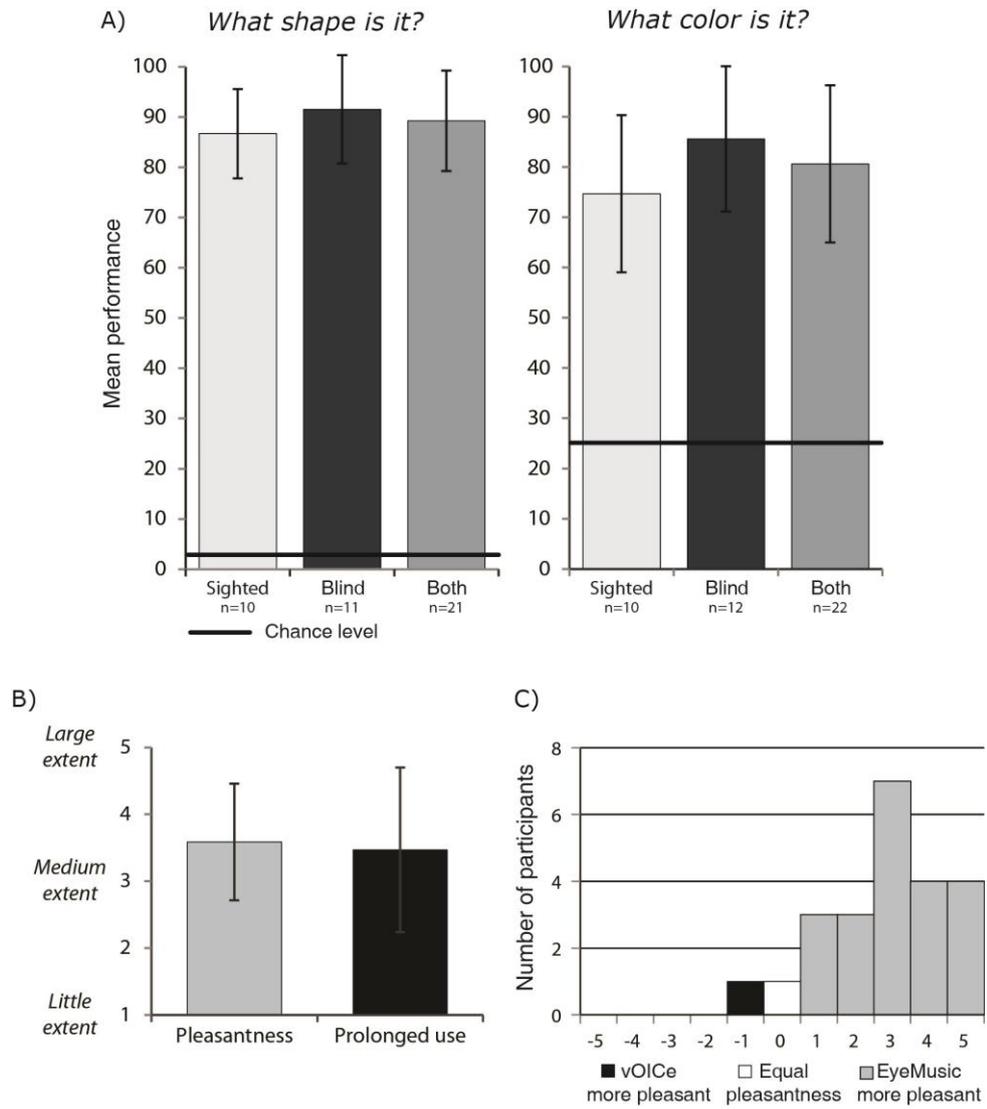


Fig. 4 – Results