Development of a Visualised Sound Simulation Environment: An e-Approach to a Constructivist Way of Learning

Jingjing Zhang[1], Beau Lotto[2], Ilias Bergstom[2], Lefkothea Andreou[2], Youzou Miyadera[3], and Setsuo Yokoyama[4]

[1] Department of Education, University of Oxford, UK
jingjing.zhang@bnc.ox.ac.uk
[2] Lottolab Studio, University College London, UK
[3] Division of Natural Science, Tokyo Gakugei University, Japan
[4] Information Processing Centre, Tokyo Gakugei University, Japan

Abstract. In this paper, the design and implementation of a visualised sound simulation environment is presented as an initial step to further laboratory experimentation. Preliminary laboratory experiments showed a positive learning curve in human auditory perception. Learning occurred when new information was processed with relevant existing knowledge in this simulation environment. While the work towards the truth of the empirical hypothesis is still under discussion, this project has been expanded beyond the scope that was originally envisaged and the developed environment showed its potential to be adopted on mobile devices for many educational purposes. This initiative not only brings scientists and educators together, but it is also hoped that it represents a possible e-approach to a constructivist way of learning.

Keywords: visualisation, simulation, constructivist, learning, mobile, visual, auditory.

1 Introduction

John Dewey (1933), the educational philosopher, defined learning as "the continual process of discovering insights, inventing new possibilities for action, and observing the consequences leading to new insights". The classical view of cognition considers humans as processors of knowledge. Knowledge as an object can be transferred from one mind to another. In contrast, the constructivist view of cognition sees humans as constructors of knowledge. Knowledge as a mental representation exists in the human mind, which cannot be moved from one mind to another. Construction plays the role in integrating the new material with relevant existing knowledge in a coherent structure (Mayer, 2003).

Despite the battle between cognitivism and constructivism continuing for centuries, we accepted that both approaches have their role to play and took an innovative approach to understanding learning from where the process of learning begins. Gibson (1991, p. 493) argued that the acquisition of knowledge actually “begins with and depends upon knowledge that is obtained through perception, which extracts

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2009, LNCS 5613, pp. 266–275, 2009.

information from arrays of stimulation that specify the events, layout and objects of the world". In this paper, we put focus on parts of the learning process that look to be somewhat independent from conscious or critical forms of learning (such as learning to solve a mathematical problem) but that are involved with rather low-level acquisition through the central perception system.

Perception has long been considered as a modular function, with the different sensory modalities working separately from each other (Schlottmann, 2000). Recent studies in neuroscience and psychology, however, challenge this view by suggesting that cross-modal interactions are in fact more common than originally thought (Martino & Marks, 2000). A number of experiments involved with cross modality have been carried out (e.g. Butler & Humanski, 1992; Carlile, 1990). Recently, research on sensory perception has become both popular and integrated with fast development of computer simulation technology. However, the fundamental problem for any sensory-guided system (natural or artificial) is that the information it receives from its environment is ambiguous. Particularly, most computer simulated sound systems fail under natural conditions, or when forced to contend with an environment that is not explicitly represented in the system. How natural systems resolve this challenge is currently not known, though one increasingly popular hypothesis is that the problem is resolved empirically. That is, the system encodes the statistics of its experience with past sources of stimuli[1]. Despite of the increasing support for this view, however, there is currently no one that has tested it. One reason is that to truly test this view we would have to provide someone with a wholly new sensory modality, and then look to see how that person's brain deals with the underlying ambiguity of the information it receives. This is of course impossible. What is possible, however, is to present one of the senses with a new kind of statistical experience in a simulated environment.

Therefore, a visualised sound simulation environment was designed and implemented, as an initial step to further laboratory experimentation. The environment presented the visual stimuli with a new auditory experience in order to prove the ambiguity problem in vision is resolved empirically. Preliminary laboratory experiments showed a positive learning curve in human auditory perception. Learning occurred when new information was process with relevant existing knowledge in this sound simulation environment. While the work towards the truth of the empirical hypothesis is still under discussion, this project is extended beyond its initial establishment and the developed environment shows its potential to be adopted in different real-world educational settings, e.g. a music lesson and a drawing course. This initiative not only brings scientists and educators together, but it is also hoped that it represents a possible e-approach to a constructivist way of learning.

2 Mapping Model

A Mapping Model (M) was proposed for translating any pixel in a real 2D image to a unique virtual sound pixel of an imaginary sound panel in two continuous steps. That

[1] Constructivists such as Helmholtz H. and Gregory R. argue that external world cannot be directly perceived because of the poverty of the information in the retinal images. Since information is not directly given, we have to interpret the sensory data in order to construct perception.

is, a real image pixel is mapped to a 'visual pixel', which is subsequently converted as a 'sound pixel' by using this Mapping Model (M).

$$\boldsymbol{M} \quad = \quad \boldsymbol{\{R, I, S, M_{ri}, M_{is}\}}$$

The elements of Mapping Model (M), are justified as follows:

$\boldsymbol{R}$: a $w \times h$ matrix, which represents the real image input (which can be anything from hand drawn images to photographs or streaming video).

$\boldsymbol{I}$: a $xNum \times yNum$ matrix, which represents an imaginary visual panel.

$\boldsymbol{S}$: a $(panMAX - panMIN) \times (freqMAX - freqMIN)$ matrix, which symbolises a virtual sound panel.

$\boldsymbol{M_{ri}}$: a mapping function from $\boldsymbol{R}$ to $\boldsymbol{I}$. A scaled and projected representation $\boldsymbol{I}$ of the real image R is produced by this mapping function $\boldsymbol{M_{ri}}$: Map any point $P_r(x, y) \in \boldsymbol{R}(w \times h)$ to $P_i(xth, yth) \in \boldsymbol{I}(xNum \times Num)$ by:

$$xth = \frac{x}{xunit} \qquad\qquad yth = \frac{y}{yunit}$$

$\boldsymbol{M_{is}}$: a mapping function from $\boldsymbol{I}$ to $\boldsymbol{S}$. It creates a sound panel isomorphic to the visual panel: Map any point $P_i(xth, yth) \in \boldsymbol{I}(xNum \times Num)$ to $P_s(pan, freq) \in \boldsymbol{S}((panMAX - panMIN) \times (freqMAX - freqMIN))$ by:

$$freq = \frac{freqMIN + (freqMAX - freqMIN) \times (yNum - yth - 1)}{yNum - 1}$$

$$pan = \frac{panMIN + (panMAX - panMIN) \times xth}{xNum - 1}$$

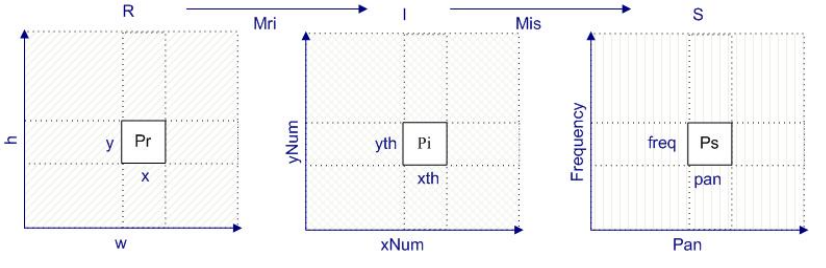

Fig. 1. Mapping Model - w is the width of the matrix $\boldsymbol{R}$, while h is the height of the matrix $\boldsymbol{R}$. xNum is the width of the matrix $\boldsymbol{I}$; yNum is the height of the matrix $\boldsymbol{I}$; (panMAX -panMIN) is the width of the matrix $\boldsymbol{S}$; y(freqMAX -freqMIN) is the height of the matrix $\boldsymbol{S}$.

As we can see, The Mapping Model shows the process of converting a real image pixel (Pr) in an image $\boldsymbol{R}$ to a visual pixel (Pi) on a visual panel $\boldsymbol{I}$ and then subsequently mapped to a sound pixel (Ps) on a sound panel $\boldsymbol{S}$. The mapping function $\boldsymbol{M_{ri}}$ is a scale formula to trim the input of a real image into a required standard image, which

can be then mapped to a virtual sound panel. The mapping function $\boldsymbol{M_{is}}$ then converts each pixel of the product of the mapping function $\boldsymbol{M_{ri}}$, to a sound pixel with an unique property, such as loudness, frequency, and pan.

3 System Design and Implementation

The invention of the screen, either on a computer or a mobile device, has completely changed the traditional work of designing. However, while it adds more interactivity and flexibility into most of our products in the daily life, it also unavoidably brings in complexity and ambiguity. It is said that miscellaneous designs has given most users a difficult time to get used to the interface. Building upon the proposed model, the designed system updated from the first working prototype with pop-up panels to a unique three-tier structure for maximising such seamless translation. The final design was based on and has been further refined by the feedback obtained during the usability testing. The look of such design was thus simple and appealing to users.

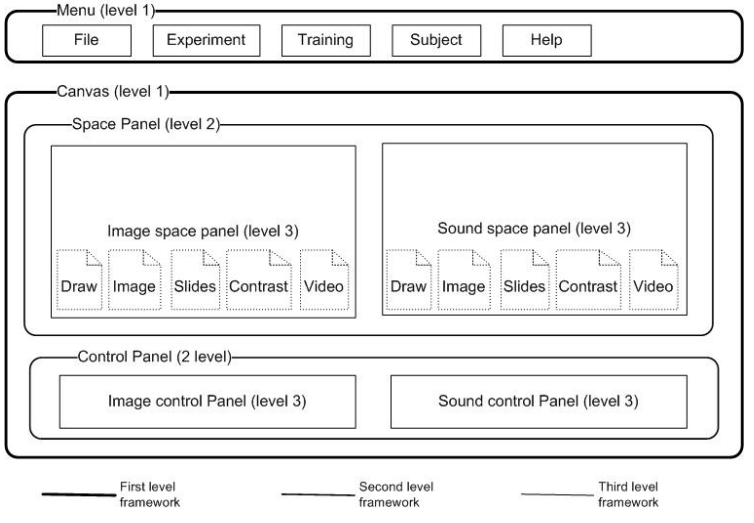

Fig. 2. Three Hierarchy Design - - showing the three-tier architecture

The top architectural level of the system was a two part top-and-bottom skeleton (Fig. 2), which included (in the top part) a menu bar and a painting canvas (in the bottom part). The second architectural level within the 'painting' canvas also consisted of a top-and-bottom structure. On the top, there was a 'Space Panel', which included an 'Image Space Panel' and a 'Sound Space Panel, and at the bottom a 'Control Panel' (i.e. an 'Image Control Panel' and a 'Sound Control Panel'). The third level was outlined as a left-and-right structure contained within the painting canvas, separated as 'image' and 'sound' (from left to right) in both the 'Space Panel' and 'Control Panel'. That is, on the painting canvas, there were four panels, from top left - an 'Image Space Panel', top right - a 'Sound Space Panel', bottom left – an

'Image Control Panel' for controlling the 'Image Space Panel', bottom right - a 'Sound Control Panel' to control the 'Sound Space Panel'.

The four panels were independently resizable and can be switched on and off for maximising the screen of any device, such as a computer or a mobile phone. The 'Image Control Panel' allowed users to change the resolution of the above image panel and the sound panel. There was also an image colour chooser panel consisted of two parts: the Preview panel and the Colour palette. The Colour Palette consisted of four tabbed panels: Swatches, HSB, RGB, and Gray Scale. The Gray Scale was a slider bar which adjusted the colour intensity of image pixels. The 'Sound Control Panel' provided different selections of 128 MIDI sounds and 1 sine sound predefined for sound pixels above with the option to choose Loop or not.

The system was implemented in Java, and so can be run as a standalone program on a local computer or an applet on the Web (the interface is shown in 2). The success of the developed system was tested in three methods: Code Reading testing, Black Box testing, and Integration testing. In addition, the system has been used to carry out several experiments to test the empirical hypothesis on nearly 20 subjects. Both the testing of the system and the results of the experiments indicated that the algorithms behind the Mapping Model successfully mapped simple visual images into sound 'images'. It is suggested that it can be further developed and installed on mobile devices in the future. It is also believed that this sound simulation would work better on a touch screen and could receive better learning outcomes whilst on the move.

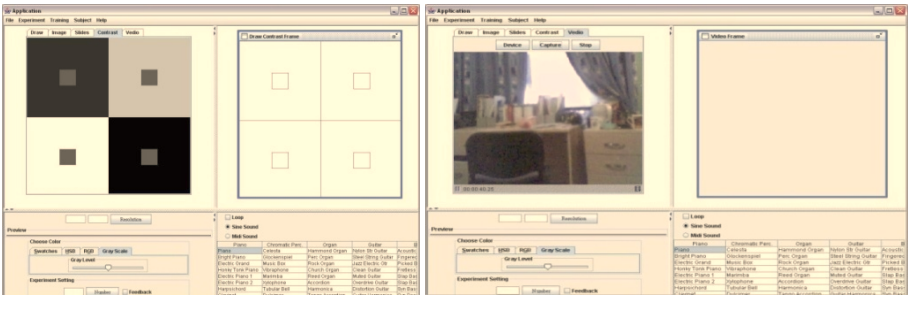

Fig. 3. System Interface (two screen captures) - showing the main windows of the user interface, which consists of four split panels with changeable size. They are: Visual Space Panel (on the top left), Sound Space Panel (the top right), the Visual Control Panel (on the bottom left), the Sound Control Panel (on the bottom right). The five menu bars are on the top of the interface: file, experiment, training, subject and help.

4 Laboratory Experiments

Early work shows that vision dominates in multi-sensory processing. This means vision is always considered as active, but hearing is usually thought of as passive. However, research has shown that while vision may dominate spatial processing, hearing dominates temporal processing (Guttman, Gilroy, & Blake, 2005). Therefore, the success of testing the 'empirical' basis of learning possibility by translating visual

information into auditory information lies in the right design of appropriate experiments, which can take best advantage of the temporal processing of hearing. A series of experimental tasks were developed, and the data from such experiments can be analysed in a great many ways, resulting in the discovery of many unknown aspects of learning. In this paper attention focuses mainly on two sets of experimental tasks accomplished by eleven subjects. There were 6 males and 5 females of them, ranging in age from 20 to 30 years old. The hearing abilities of the 11 subjects were at average level.

4.1 Experiment 1 - Sound Localisation

The sound localisation experiment was designed to measure the ability of the brain to process spatial information. This experiment differed from other sound localisation experiments in several ways. Firstly, earlier experiments always used special facilities to organise the experiment such as geodesic spheres housing an array of 277 loudspeakers (e.g. Hartmann & Acoust, 1983). This increased the cost of experiments and decreased the practicality. In contrast, this experiment used a computer simulated system, which can automatically accomplish the experiment without manual input and supervision. In addition, the system was developed in Java and can be easily extended to a mobile device in the future. Secondly, real sound sources (e.g. speakers) were usually used in other experiments earlier whereas our experiments used virtual 2D sound space, in which each 'sound pixel' represented a unique virtual sound source thereby again saving on cost, practicality and time. Finally, unlike 3D cave-like virtual environments where sound localisation system involved with head-related transfer functions (HRTFs), each image pixel in this developed system was translated directly to a sound pixel. To add more detail, in the horizontal direction, the inter-aural time difference (i.e. Inter-aural level difference and Inter-aural time difference) was used to determine horizontal position. In the vertical direction, frequency was used to perceive vertical position in this experiment. These two parameters are used to distinguish each 'sound pixel' in experiment with respect to the Mapping Model (M). All the mouse-click responses were recorded and stored in a text file.

In the series of this experiment, the resolutions of the sound panel increased from 3×3, 7×7, to 11×11. In each one, four different combinations of trials were used to

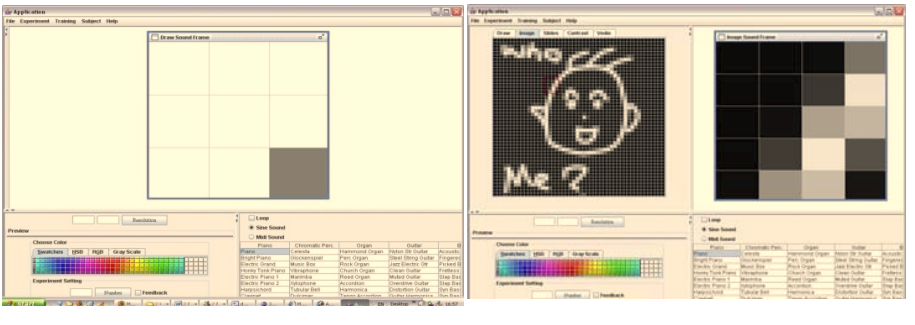

Fig. 4. Experiment 1 Sound Localisation - the Square with gray colour is the square where the subject clicked to predict the location of the sound

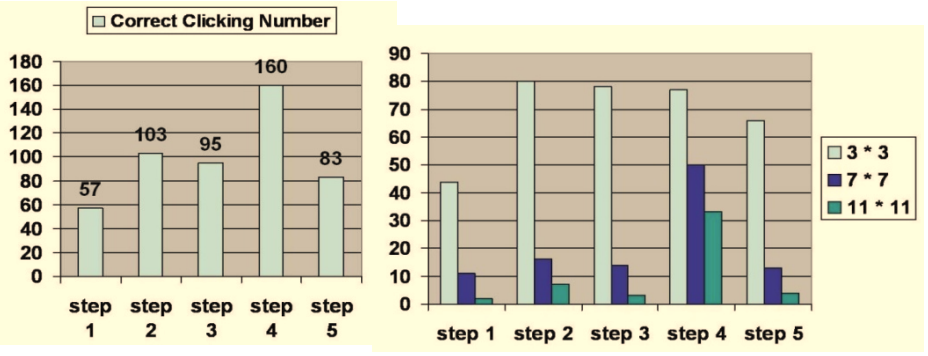

Fig. 5. Results of Experiment 1 - showing the total number of correct clicks from step 1 to step 5. The step 1 and step 5 were the test stages, while the step 2, 3, and 4 served as the training. The number of correct clicks of the step increased after 3 different kinds of training. The step 4 (adjacent sound with a reference sound) achieved the best results. The number of correct clicks decreased as the sound space dimension increased.

explore a better pattern for identifying sound. Subjects were first introduced to a random sound, and then were given a reference sound from the pixel in the centre of the sound panel followed by a random sound. After this, subjects were given a series of sounds which were adjacent to each other. Then, subjects were given a reference sound followed by an adjacent sound to it. Last, subjects were provided by a random sound again to find out whether learning occurred from the training.

As we can see in Fig. 5, the sound localisation ability is clearly improved with training. Subjects were also better able to locate sound if given an initial reference sound. The chart on the right shows that the ability of the subjects to locate sound decreased with increased resolution. By comparing the ability of identifying sound horizontally and vertically, if the space dimension is smaller than 7×7, subjects can find the horizontal position of a sound better than its vertical position in the space; if more than 7, it seemed to be easier to identify a sound in the vertical axis. As expected, the subject's ability to specifically locate sound correctly in the 11×11 space was very low. Despite this, as shown in Fig. 6, while the absolute position was incorrect, subjects were very good at finding the relative location of the sound source, given the previous test sound. Thus, in deciding on the location of the pixels in this sound space, subjects used their previous responses to previous stimuli in order to estimate the positions of future test sounds. This indicates that the human auditory system predicts new information based on past information. Thus the past information can be used to aid in the ambiguity that they experience when trying to predict the new test sound, and enhance the current attempt.

The positive parts of the experiments have shown that humans are able to learn from the wholly new translated sensory information (as shown by the increasing learning curves from the sound localisation experiment data). We suspect a long-term training process would help to better develop subjects' visual and audio sense with this new kind of audio-visual experience.

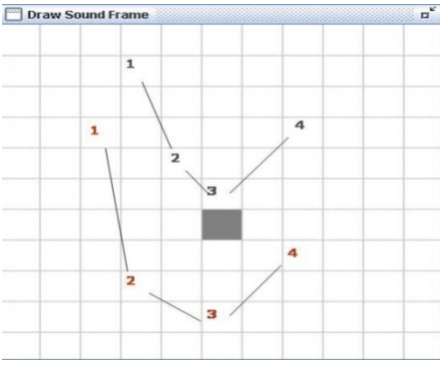

Fig. 6. Total Number of Correct Clicks (General Pattern of Movement) – The red numbers show the sequential locations of the test sounds sound generated by the experiment. The black numbers represent an example subjects clicking responses.)

4.2 Experiment 2 - Object Recognition

While Experiment 1 was based on one sound, the object recognition experiment was further designed to increase the sound sources. Nine 20 × 20 simple black and white images, consisted of the same number of pixels (and therefore of the same overall intensity), were used. The images were randomly loaded from a local folder on a computer. There was the same number of white pixels in each image. Pixels in white were the one able to produce sound simultaneously, while pixels in black were silent ones. Subjects were first introduced by a compound sound from a random image, and asked to pick up a right image to pair with the sound. They were then presented again with a compound sound with feedback. That is, if subjects chose a wrong image, the right image would be highlighted to correct them. Then, an identical image panel would be switch on automatically on the left to allow subjects to mark a small area (in which sound pixels were able to be generated) and navigate to different directions. Sound would be generated if subjects moved onto corresponding white pixels hidden underneath. Last, subjects were presented with a random compound sound from images again to test their learning experiences.

It is clear that the ability of object recognition without training was poor (16%). However, their ability to recognise the object increased sharply when they were given the opportunity to explore it (72%), which increased further after four trial explorations (where the subjects were correct 100% of the time). The statistical data also showed that motion aided in recognition of the masked object as opposed to polyphonic compound sounds generated by the static object. That is, motion was easier to recognise than the static object. By moving around, the brain could construct an internal map of the image from experience. Although it is clear that in moving the rectangular window around in a sometimes arbitrary fashion, only a temporary memory was utilised to intuitively construct a mental model of the 2D environment. We believe that without using any exploration of the masked Sound Space to obtain temporary experience and perceptions, the brain instead needs the experience stored in the brain by long-term training.

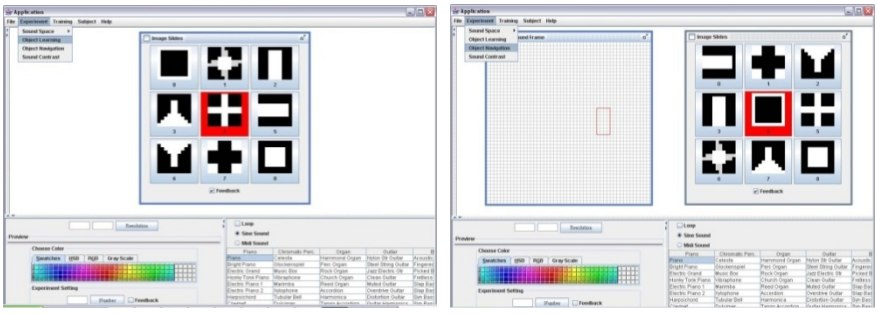

Fig. 7. Experiment 2 – the interface on the left shows the object recognition experiment in Step 1, 2 and 4. The interface on the right shows the object recognition with exploration in step 3. There is the masked image panel on the left containing one of the nine images on the right hand side underneath. The subject can drag a rectangle by mouse and move it by pressing the up, down, left, right arrow keys to explore this image. After exploration, the subject clicked the one he/she thought was correct on the right hand side. The red image button then showed the correct answer.

5 From Laboratory to Real World

As stated earlier, this study, was at first intended to be a three-year project in a neuroscience laboratory to prove the empirical vision theory, but later expanded beyond the scope that was originally envisaged, as the adoption of this developed sound simulation in real educational settings became increasingly feasible. Especially, when mobile technologies hold out the promise of ubiquitous leverage to this Java-based application, the study moved on to the discussion of the social values of such sound simulation on mobile devices. Statistical data showed that global mobile phone use was set to pass 3.25 billion by 2007 (Reuters, 2007) - around half the world's population. With regard to this, it is widely accepted that mobile and ubiquitous computing can genuinely support learning, but it is crucially important to ask how such simulated environment can be combined with mobile technologies to be immersed into various kinds of learning experiences in different contexts.

It is a very complex undertaking to design interfaces for mobile devices. Visual interfaces have unpleasant limitations in small-screen devices, as they have a confined space on which to display information (Rinott, 2004; Smith & Walker, 2005; Walker & Lindsay, 2006). While mobile devices were getting smaller and lighter, this developed environment with its three-tier structures was able to switch on and off four different panels in the real time. That is, it potentially enlarged the small screen four times. Such an interface with layered view also promised seamless switching between different layers to a great extent of graphic rending. More importantly, the designed interface was sonically enhanced, and required less visual attention. The use of audio feedback thus to certain extent augmented the display of a mobile device, and helped users to be able to navigate in unfamiliar interface more easily.

Not only can the interface of a mobile device be improved by this developed sound visualisation simulation environment, but also can the created applications widely be used for many educational purposes. Visual and auditory sensory are closely involved with each other. On the one hand, hand drawing exercises with sound in this

developed environment provided a good opportunity to bring sound or music into the drawing practice. On the other hand, the database of 128 midi music sounds in two dimensions (frequency and pan) enhanced the traditional music teaching. The concise design integrated with drawing function was likely to be of more appeal to learners. More importantly, the environment held a record of every single experiment in the background, and was capable of providing a learning curve or analysis. Furthermore, the possibility to tailor learner-centred experiments by learner themselves gave this environment a great deal of flexibility and openness. It allowed a healthy cycle of further development in the future.

6 Conclusion

The developed sound visualisation environment was able, to a certain degree, to translate the visual stimuli into sound stimuli although we did encounter some shortcomings in the experimental implementation. Furthermore, this system is planned to be implemented on mobile devices. It is believed that this sound simulation would work better on a touch screen and could result in better learning outcomes whilst on the move. This initiative not only brings scientists and educators together, but it is also hoped that it represents a possible e-approach to a constructivist way of learning.

References

1. Butler, R.A., Humanski, R.A.: Localisation of Sound in the Vertical Plane with and without High-frequency Spectral Cues. Percept. Psychophys. 51, 182–186 (1992)
2. Carlile, S., King, A.J.: Monaural and Binaural Spectrum Level Cues in the Ferret: Acoustics and the Neural Representation of Auditory Space. Journal of Neurophysiology 71 (1994)
3. Dewey, J.: How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process (New edn.). D.C. Heath, Lexington (1933)
4. Gibson, E.J.: An Odyssey in Learning and Perception. MIT Press, Cambridge (1991)
5. Guttman, S., Gilroy, L., Blake, R.: Hearing What the Eyes See: Auditory Encoding of Visual Temporal Sequences. Psychological Science 16(3), 228–235 (2005)
6. Hartmann, W.M., Acoust, J.: Localisation of Sound in Rooms I. Soc. Am. 74, 1380–1391 (1983)
7. Martino, G., Marks, L.E.: Cross-modal interaction between vision and touch: the role of synesthetic correspondence. In: Perception 2000, vol. 29, pp. 745–754 (2000)
8. Mayer, R.E.: Memory and Information Processes. In: Weiner, I.B. (ed.) Handbook of Psychology, vol. 07. Wiley, Hoboken (2003)
9. Purves, D., Lotto, R.B.: Why We Wee What We Do: An Empirical Theory of Vision (November 2002)
10. Reuters: Global Mobile Phone Use to Pass 3 Billion (2007)
11. Rinott, M.: Sonified Interactions with Mobile Devices. In: Proceedings of the International Workshop on Interactive Sonification, Bielefeld, Germany (2004)
12. Schlottmann, A.: Is perception of causality modular? Cognitive Sciences 4 (2000)
13. Smith, D.R., Walker, B.N.: Effects of Auditory Context Cues and Training on Performance of a Point Estimation Sonification Task. Applied Cognitive Psychology 19(8) (2005)
14. Walker, B.N., Lindsay, J.: Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. Human Factors 48(2), 265–278 (2006)