http://dx.doi.org/10.14236/ewic/EVA2016.4

Approaches to Visualising the Spatial Position of 'Sound-objects'

Jamie Bullock
Birmingham Conservatoire
Birmingham, UK
jamie.bullock@bcu.ac.uk

Balandino Di Donato
Birmingham Conservatoire
Birmingham, UK
balandino.didonato@mail.bcu.ac.uk

In this paper we present the rationale and design for two systems (developed by the Integra Lab research group at Birmingham Conservatoire) implementing a common approach to interactive visualisation of the spatial position of 'sound-objects'. The first system forms part of the AHRC-funded project 'Transforming Transformation: 3D Models for Interactive Sound Design', which entails the development of a new interaction model for audio processing whereby sound can be manipulated through grasp as if it were an invisible 3D object. The second system concerns the spatial manipulation of 'beatboxer' vocal sound using handheld mobile devices through already-learned physical movement. In both cases a means to visualise the spatial position of multiple sound sources within a 3D 'stereo image' is central to the system design, so a common model for this task was therefore developed. This paper describes the ways in which sound and spatial information are implemented to meet the practical demands of these systems, whilst relating this to the wider context of extant, and potential future methods for spatial audio visualisation.

Digital art. Mobile applications. Music. Performing arts. Technologies.

1. INTRODUCTION

This research relates to the electronic manipulation of sound following its conversion to an electrical signal, and the means by which sound(s) should be visualized in this context. We limit our study here to the representation of *digitized* sound, specifically transformations on digital signals through interaction with graphical user interfaces (GUIs). This kind of digital signal processing makes possible the synthetic generation of any conceivable sound either by processing complex waveforms or combining simple ones. Such overwhelming sonic capability presents significant design challenges around *how* signal processing should be presented to end users, and how they should interact with this presentation. Many digital audio processing tools now exist and are widely available in a range of contexts including mobile devices, laptop and desktop computers and web browsers. Use cases include sound design for film, TV, radio and games industries; music composition, performance and production. A fundamental question for us as researchers is therefore: how can we visually represent sound in a way that is optimized for human interaction?

2. RESEARCH CONTEXT

2.1 Existing approaches to sound visualization

Sound visualization for the purposes of digital audio transformation generally operates at one of a number of possible levels of abstraction. These are shown in Table 1.A problem with these approaches is that they separate the means of visually representing sound from the visual representation of sound transformation inputs. For example 'sound' is typically represented in digital audio software using a 'waveform' visualization, whilst controls for user input to transformations are represented as UI widgets such as sliders, knobs and buttons. Second order transformations over time are represented using linear breakpoint functions (or envelopes). A further fragmentation of visual representation is introduced in the case of 3D spatial audio positioning (e.g. with a 5.1 surround sound system), because a single perceptual quality (spatial location) corresponds to multiple underlying control parameters (e.g. *x*, *y*, and possibly *z* co-ordinates). Navigating between these levels of representation (usually presented in separate 'views') introduces complexity into software design and introduces significant cognitive overhead for users (Storey *et al*. 1999, Wang Baldonado *et al*. 2000). In order to address these issues, an alternative approach to sound

Table 1: *Identified perceptual dimensions of musical timbre and visual texture (Giannakis and Smith 2000)*

Abstraction level	Visualisation method	Representation Type	Exemplars
Low	Audio waveform	Physical (x = time, y = amplitude)	Common audio editors, DAW software
Mid	Sonogram	Physical (x = time, y = frequency, colour = amplitude)	Specialist spectral transformation software
High	Icon	Pictographic (e.g. a loudspeaker as a sound source)	Spatialisation software, configuration dialogs
High	Notation	Symbolic (e.g. musical notation, visual "envelopes")	Audio plugins, notation software, sequencers

visualization is required, which better aligns with human cognition by providing a unified visual representation.

2.2 Relating auditory and visual perception

Chromesthesia, the cross-modal perception of sound and colour, is a neurological phenomenon that has been explored widely by philosophers, psychologists, and artists (Marks 1975). Despite significant progress in this area, for example through quantitative studies (Calkins 1893, Polzella & Biers1987), and recent work by Natilus (2015), Partesotti & Tavares (2014) and Cassidy (2015), there is still significant uncertainty surrounding the communality between sound and its visual correlates. Related to this, Bregman (1994) highlights one of the primary differences between auditory and visual perception:when we listen to a sound, we focus on the sound emitting *source* rather than its 'reflected' *image* as with visual perception.Furthermore, in certain situations human perception of audio-visual spatial phenomena can differ from physical reality, for instance when experiencing the ventriloquist effect (Frassinetti *et al.* 2002, Spence & Driver 2000), where we 'incorrectly' associate sounds with a particular source in space through visual correlation, rather than audio alone.

Additional audio-visual factors need to be considered when sound sources are moving in physical space. We refer to the transformation path of a sound source's spatial location from one set of co-ordinates to another as its 'spatial trajectory'. In the physical world in both auditory and visual domains, such trajectories can only be directly observed retrospectively and under specific conditions. For example the condensation trail generated by an aeroplane and the Doppler effect generated by an emergency vehicle's siren. In these cases our brain enables us to perceive more than is manifested in the real world. In fact, we are able to reconstruct a trajectory through our perception even though in some cases information is missing to us (Bertin&Berthoz2004, Boulinguez *et al.* 2009, Kawato 1999). This also enables the brain to predict future trajectories of moving objects. Numerous approaches have been employed to visually represent such trajectories in software systems (Sega Amusements 2011, WB Kids 2015). This method of representation extends to 'predictive' trajectories, which are displayed in order to inform the user of future trajectories travelled by an object (EA Sports 2016).

2.3 Visual representation of sound

In our work, we focus on visual display of sound, specifically timbral audio qualities and spatial propagation features of the sound source within a stereo image. An example of such a visual representation of sound in musical practice can be found in the score of Stockhausen's Studie II (Stockhausen 1991), where the single blocks on the top part of the score represent sound's spectral content and the bottom part represents amplitude over time.

Cymatics, is a method providing enhanced visual access to acoustic phenomena that are typically only experienced through our senses of hearing and touch (Lewis 2010). It has been defined by Dr Hans Jenny who was the first person to accurately record this visual representation following a scientific method(Lewis 2010). Another method for representing sound is the Schlieren imaging system, a non-intrusive method for studying transparent and optical media, which was first adopted in 1800s to study fluctuations in optical density (Mazumdar 2013, Settles 2012, Taylor&Waldram1933). This system allows users to visually capture the movement of air molecules generated by a sound source.

2.4Representing sound-objects

Unlike certain visual forms (for example a painting or photograph), sound always has a temporal dimension. As mentioned in section 2.1 this dimension is typically represented visually by assigning the graphical ‘x axis ’to ‘time’ in audio software. Such an approach is problematic for a direct manipulation interface for two reasons:

1. It presupposes that the audio being manipulated is ‘fixed’ i.e. an existing sound file. This model therefore does not work for ‘live’ audio transformations, e.g. from a microphone
2. It means the graphical x-axis cannot be used to also represent ‘space’, e.g. in the case of left-right stereo panning

Our proposed solution to this is instead to represent an audio *source* and transformations upon it using a combination of symbolic and pictographic notation within a virtual 3D space. This allows the dimensions of the virtual space to map to the dimensions of a 3D acoustic space (physical or perceptual). Since it is cumbersome to interact with a 3D visual representation via a 2D input (such as a mouse or trackpad), our proposed model also assumes 3-dimensional input. Thus, our model maps absolute position of user’s input in 3D physical space to a *relative* position within a 3D virtual space, which corresponds to a point within a 3D sound image. This could potentially be rendered on a small scale (e.g. through binaural headphones) or large scale (e.g. ambisonic auditorium) system.

We use the term ‘sound-object’ here to describe the visual representation of a discrete sound source within the virtual space. This differs from the Schaefferian definition of a object sonore (Schaeffer 1966), which refers to a sound event over time (i.e. that has fixed duration) that is perceptually separated from its source (e.g. the sound of door slamming played through a loudspeaker). In our proposed system sound-objects can represent single digital sound sources such as continuous or looped audio file playback, or a live audio input from a microphone or sound card’s line input.

3. TWO IMPLEMENTATIONS OF THE MODEL

3.1 System 1

The first system we describe seeks to address the problems defined in section 2.1 by investigating the application of a ‘direct manipulation’ paradigm in the context of audio transformation. Direct manipulation is a human-computer interaction concept with multiple related and emergent meanings originating from the early 1980s. For the purposes of our research, we take the three principles originally developed by Shneiderman:

1. Continuous representation of the objects and actions of interest;
2. Physical actions or presses of labelled buttons instead of complex syntax;
3. Rapid incremental reversible operations whose effect on the object of interest is immediately visible (Shneiderman, 1999).

Thus, we define a direct manipulation interface for sound transformation as one in which the user interacts directly with an object representing the sound, and whereby sound can be transformed by manipulating that object. Since in the physical world sound can be neither seen nor touched, we propose a virtual *proxy* representation, which users manipulate through free-hand interaction (Baudel and Beaudouin-Lafon 1993). Our hypothesis is that by modifying such a visual representation with real-time free-hand input, accompanied by continuous audio feedback (and possibly haptic feedback), an illusion can be created that users are directly touching and manipulating sound itself.

System 1 implements this concept, initially using only one form of audio transformation: sound spatialisation, specifically binaural 3D sound positioning. This audio processing technique allows sound to be positioned in *perceptual* 3D space using stereo headphones. Commonly used in computer games, binaural processing enables a sound to appear to be coming from in front of, behind, to the left, right, or above the listener. Exploiting more advanced psychoacoustic effects, coupled with visual cues can also enable sounds to appear to be more proximal, distant, or occluded.

3.1.1Prototype System Design

As a starting point, the simplest and most obvious interaction and visualisation model was chosen. Since prior work (Gelineck & Korsgaard 2015, Jankowski et al 2013, Vorländer 2007) has successfully used spheres within a 3D virtual space (usually a hollow rectangular prism) to represent sound-objects, we decided to use this as a starting point for our research.

The initial iteration of System 1 was therefore implemented in Unity, a cross-platform game engine suited to 3D graphics and interaction. A Microsoft Kinect 2 is used as a motion capture device in order to detect the position of the user’s hands as well their hand pose. The Kinect 2 was chosen as an initial input device due to the hand pose recognition built into the SDK. Poses initially used are ‘hand closed’, ‘hand open’ and ‘index finger’. The user’s hand centre position is captured by the Kinect and translated into a point within the co-ordinate system of a virtual 3D space so it can be visualized within a virtual environment (VE). The initial design for this is shown in Figure 1.

This visual display represents a virtual space into which sound-objects (represented as spheres) can be positioned. The centre point of each sphere represents the corresponding sound-object's positional audio location, which is rendered in real-time using binaural playback through headphones. For example, if the user moves a sphere from left to right in the virtual space, the sound-object will simultaneously be localized from left to right in the headphones. In order to provide 3D binaural audio rendering, the 3Dception Unity plugin by TwoBigEars is used (3Dception 2016).

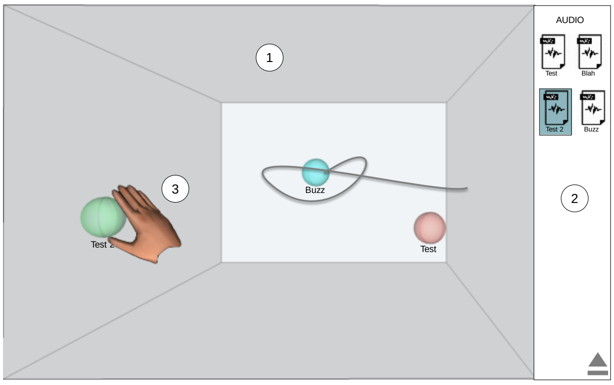

Figure 1: *Initial design showing virtual space (1), sound palette (2) and avatar hand position in the virtual space (3)*

The visual display also provides a hand avatar for the user enabling them to accurately position their hand within the virtual space. A palette of sound sources is displayed in a 2D panel on the right-hand side of the screen. When the user's physical hand position is placed such that their hand position within the VE is in front of a sound source (e.g. an audio file e.g. microphone input), a 'hand closed' pose will cause the corresponding source to be 'picked up'. As the user moves their hand across into the virtual space, the source icon will change to a sphere and a 'hand open' pose will cause the source to be 'released' into the space. At this point audio will become active for the source, it is now a valid sound-object, and its visual representation changes to a sphere within the virtual space. Once a source becomes a sound-object within the VE, it can be picked up and moved around through a grab (hand closed) → move → drop (hand open) interaction. Consistent with the 'object' metaphor, sound-objects can be de-activated (stopping audio playback) by removing it from the VE. This is achieved by grabbing a sphere and dragging it onto the 'eject' icon in the bottom-right corner of the sound source palette.

3.2 System 2

System 2 builds upon the approach employed in System 1 by allowing for a wider range of audio processing types (including filtering and delays), as well as visualisation of spatial inertia and timbral change. System 2 is intended for practical usage in Human Beatboxing (beatboxing) performance, and is therefore designed for use on standard mobile devices such as the Apple iPhone.

Synesthetic and multi-sensorial studies confirmed the influence of the vision on auditory perception (Marks 1975, Goller*et al*. 2009), and reported the extent to which cross-modal events impact on the experience of spatial perception and level of immersiveness (Bolognini *et al*. 2005, Eimer 2004, Storms 1998).System 2 builds upon this research by using visual feedback toenhance the performer and audience's level of 'immersiveness' during live musical performance (Platz&Kopiez2012), specifically beatboxing performances. In this system, the sound source's visual representation does not only represent the acoustic qualities of the sound, but also the gestural user interaction which affects transformations upon it. In the case of System 2, we define gestures as movements enacted by the performer in order to represent visually the auditory quality of the vocal sound itself and the musical structure of the vocalised musical idea. System 2 allows the user to spatialise a sound source within a VE through gestural control using the*Myo* (Thalmic 2016) armband*.* As with System 1, a sound-object can be moved within the VE by dragging and dropping a visual sphere (Figure 2, A). In addition to creating sound trajectories (as defined in section 2.2) by moving the sound source's icon, it is also possible to 'draw' trajectories (Figure 2, B), which are subsequently traversed by the sound source at a later point. A third way of creating trajectories is by 'throwing' the sound source away (Figure 5), thus giving it 'inertia'.

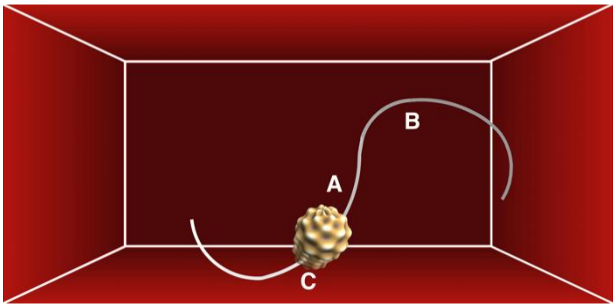

Figure 2: *The System 2's TM2 view, where A is the sound source's icon, B is the sound trajectory, and C is the sound source tail.*

3.2.1. User interaction

In System 2, the sound spatialisation is controlled using the Myo armband. A sound source can be moved by orientating the user's arm around the *x*, *y*, and *z* axes (Figure 6) to affect respectively the sound source's vertical, horizontal and depth position within the VE. Two main control gestures, which are the fist (Figure 3, a) and the 'finger spread' gesture (Figure 3, b)are used differently, depending upon the chosen modality between *Moving Mode* (MM) and *Drawing Mode* (DM). These two gestures have been chosen to allow the user to interact with the virtual world in a seamless manner by enhancing learnability (Omata & Imamiya 2000, Nymoen *et al.* 2015).

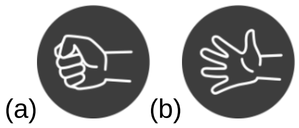

Figure 3: *fist (a) and the finger spread (b) gestures (Source: Thalmic's Myo brand guideline).*

Moving Mode (MM) is selectable by waving the hand inwards (Figure 4, a) for two seconds. In MM it is possible to grab and drop (throw) the sound-object by respectively performing a fist and a finger spread gesture. In MM, the sound-object can also be 'thrown away' by emulating the gesture to throw a ball away. In the *Drawing Mode* (DM), selectable by waving the hand outwards (Figure 4, b) for two seconds, the user can use the fist and 'finger spread' gestures respectively to initiate and stop drawing of a trajectory. After a trajectory has been drawn, selecting again MM it is possible to restart interaction with the sound-object, and by placing it over the drawn trajectory, the sound-object will start to travel along the trajectory (Figure 2, B).

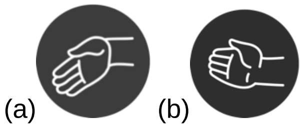

Figure 4: *wave in (a) and wave out (b) gestures (Source: Thalmic's Myo brand guideline).*

Table 2: *Identified perceptual dimensions of musical timbre and visual texture (K. Giannakis and M. Smith 2000)*

Modality	Control	Gesture
MM (wave inwards for 2 seconds)	*Drag*	*Fist*
	Drop	*Finger spread*
DM (wave outwards for 2 seconds)	*Start drawing*	*Fist*
	Stop drawing	*Finger spread*

3.2.2. The avatar

As in System 1, in System 2, the user is facilitated in grasping a sound-object through an avatar of her hand. The avatar is represented differently depending on which mode the user is in. In MM the avatar is represented by an open hand (Figure 5, a) if no fist gesture is performed, otherwise it will become a fist (Figure 5, b). In DM, the avatar has the open hand icon if no fist is performed, otherwise it appears as a pen (Figure 5, c) to indicate that 'drawing' is now possible.

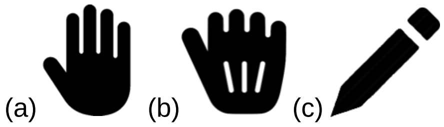

Figure 5: *Avatars.*

3.2.3. Sound-objects

As with System 1, the sound-object (figure 2, A) is initially represented as a sphere. However, System 2 further develops this visualisation method: starting from a monochromatic and plain sphere, its shape and texture over time are determined by the qualities of the post-processed sound. Specifically, the shape of the sound-object in the VE corresponds to the post-processed sound source's *spatial radiation pattern* at a given frequency. The spatial radiation pattern frequency reference is calculated by extracting the sound source's spectral centroid.

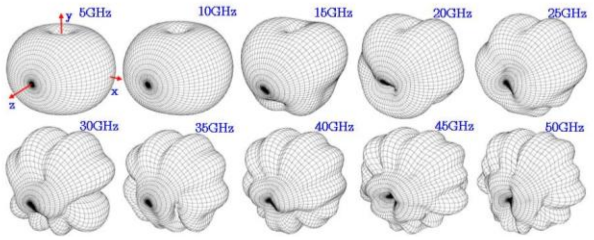

Figure 6: *An example of 3D representation of a antenna signal's spatial radiation pattern at different frequency references (Tran D., et al. 2012)*

Furthermore, the visual 'texture' of a sound-object is determined by the timbral features of the source sound. The adopted method to map timbral features into visual features derives from prior work by Giannakis and Smith (2000), which establishes a relationship between timbre and texture as summarised in the table below. A blur of the sound-object's visual representation (Figure 7) is used to represent the sound source's scattered field as it is 'thrown away'.

Table 3: *Identified perceptual dimensions of musical timbre and visual texture (K. Giannakis and M. Smith 2000)*

Timbre	Texture
Sharpness	*Repetiveness*
Compactness	*Contrast, Directionality*
Spectral smoothness	*Granularity, coarseness and complexity*
Roughness	

3.2.4. Trajectories

In both MM and DM, trajectories can be defined by (a) grabbing the sound and dropping it in a different spatial position within the VE, (b) drawing a trajectory, which is then followed by the sound-object or (c) throwing the sound-object away. Subsequently, the sound source can be grabbed and dropped in correspondence with the trajectory's path in order to make the sound source travel through it, considering both the position and speed recorded with the trajectory. In this case the trajectory is shown as a monochromatic line (Figure 2, B), where brightness is affected by depth within the VE and opacity by the gesture velocity.

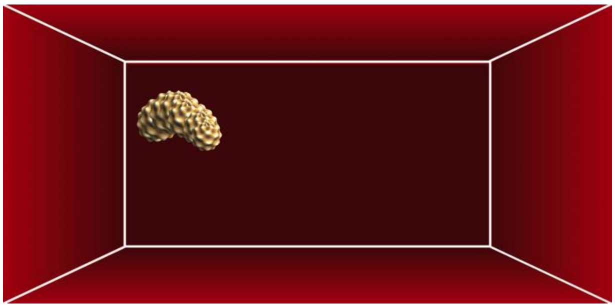

Figure 7: *The System 2's TM2 view after the sound source has been thrown away.*

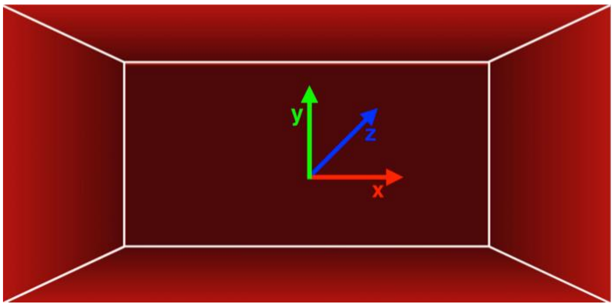

Figure 8: *Coordinates within the virtual environment.*

3.2.5. Implementation

The system has been implemented using the Myo armband as a input device, MyoMapper (MyoMapper 2016), developed using Processing (Processing 2016) and Pure Data (Pure Data 2016) for elaborating the incoming audio signal. Myo is a device that tracks muscle activity through eight medical-grade stainless steel EMG sensors, and the forearm's orientation using a sensitive nine-axis IMU containing three-axis gyroscope, three-axis accelerometer and three-axis magnetometer. In addition, it allows communication of visual feedback to the user through dual indicator LEDs, and haptic feedback through short, medium, and long PWM vibrations. It sends data to a computer or mobile device using Bluetooth® wireless technology, and is self powered through a built-in rechargeable lithium ion battery (Thalmic Labs 2016). MyoConnect (Thalmic 2016) is used for the gesture recognition of the fist, 'finger spread', 'wave in' and 'wave out' gestures. The movement of the sound-object is calculated using the IMU as established in Haque *et al.*(2015), which proposes a dedicated pointer acceleration function (Formula 1), following guidelines reported in (Nancel *et al.* 2013),to transform the arm velocity into pointer cursor velocity, which in our case is the avatar.

$$v_{cursor} = v_{arm} \times \left(58.28 + \frac{5060.14}{1 + e^{-1.45\times(vel_{arm}-1.66)}}\right)$$

Formula 1: *Pointer acceleration function to transform Varm (rad/s) to Vcursor (mm/s) (Haque et al. 2015).*

Trajectories in *3.2.4,* are defined by first mapping (i) the movement intensity and (ii) the orientation of the arm into sound-object velocity and initial angle. The former (i) is calculated to the sum of (a) the mean absolute value of the forearm's muscle activity (Formula 2) after Arief *et al.* (2015), and (b) the absolute value of the arm's movement velocity into initial velocity (Formula 3). The latter (ii) by mapping the orientation of the arm into initial angle (Formula 4) of the trajectory. Both the initial velocity and the initial angle are calculated at the moment of performing 'finger spread' gesture, which is the command to 'throw' the sound away.

$$MAV = \frac{1}{N}\sum_{k=1}^{N}|X_k|$$

Formula 2: *EMG Mean Absolute Value (MAV), where Xk is EMG data at k and N is number of samples. This function resulted the most efficient function in order to describe hand gestures activity through forearms muscles activity (Arief et al. 2015).*

$$V_i(m/s) = EMG_{MAV} + \left|\left(Acc_i(x, y, z)\right)\right|$$

Formula 3: Initial velocity of the sound source's trajectory.

$$\Theta_i(x, y, z) = O_i(x, y, z)$$

Formula 4: *Initial angle, calculated by direct mapping of the arm's values (yaw, pitch and roll) at the time i.*

Knowing the trajectory's initial velocity (v) and launch angle (Θ), the trajectory's height (*h*), time (*t*) and distance (*d*) are calculated as follows (Formula 5, a, b and c), where *g* is the gravity acceleration.

(a) $h = \frac{v_i^2 \sin^2 \Theta_i}{2g}$ (b) $t = \frac{\sqrt{2} * v}{g}$ (c) $d = \frac{v^2 \sin(2\Theta) * v}{g}$

Formula 5: *Initial angle, calculated by direct mapping of the arm's values (yaw, pitch and roll) at the time i.*

4. CONCLUSIONS

In this paper we have described a range of approaches to sound visualisation with particular reference to its role in digital audio signal processing. We have introduced a 'direct manipulation' model for sound representation where the visualisation simultaneously represents the system output and input, and we have shown how this model can serve as the basis for two distinct audio processing systems. Given that the essential components of the model have been used successfully in prior research, it is not surprising that the underlying visualization approach 'works', however we have described significant development beyond the basic model and shown that there is scope for extension and further exploration.

5. FUTURE WORK

Future objectives concern the adjustment of the trajectory formulae taking into account understanding from Saberi & Perrott (1990), Chandler& Grantham (1992), which investigate the lower threshold at which humans are sensitive to sound sources trajectories, and from Speigle & Loomis (1993), who explore the perception of sound source's distance. We also aim to find ways to generate cross-modal haptic illusions through visual and auditory feedback (Biocca, F. *et al*. 2002, Lécuyer, A. 2009) in order to further enhance the level of immersiveness within the VE, and to allow the user to interact with the sound-objects through touch feedback.

6. REFERENCES

Arief, Z., Sulistijono, I. A., and Ardiansyah, R. A. and (2015) Comparison of five time series EMG features extractions using Myo Armband.*2015 International Electronics Symposium (IES),* Surabaya, Indonesia, 29–30 September 2015, 11–14. IEEE.

Baldonado W., Michelle Q., Woodruff A., and Kuchinsky, A. (2000) Guidelines for using multiple views in information visualization. W*orking Conference on Advanced Visual Interfaces*, 110-119. ACM,

Baudel, T., and Beaudouin-Lafon, M. (1993) Charade: remote control of objects using free-hand gestures. *Communications of the ACM* 36(7) pp. 28–35.

Bertin, R.J.V., andBerthoz, A. (2004) Visuo-vestibular interaction in the reconstruction of travelled trajectories. *Experimental Brain Research*, 154(1), pp.11–21.

Biocca, F., Inoue, Y., Lee, A., Polinsky, H., and Tang, A. (2002) Visual cues and virtual touch: Role of visual stimuli and intersensory integration in cross-modal haptic illusions and the sense of presence. *Presence 2002*, Porto, Portugal, October 9-11, 2002, 410-428.

Boulinguez, P., Savazzi, S., andMarzi, C. A. (2009) Visual trajectory perception in humans: Is it lateralized? Clues from online rTMS of the middle-temporal complex (MT/V5). *Behavioral brain research*, *197*(2), pp. 481-486.

Bregman, A.S. (1994) *Auditory Scene Analysis: The Perceptual Organization of Sound*, Bradford Books, MIT Press

Calkins, M. W. (1893) A Statistical Study of Pseudo-Chromesthesia and of Mental-Forms. *The American Journal of Psychology*, *5*(4), pp. 439–464.

Cassidy, O. (2015) Hearing in Color; Seeing in Sound: Chromesthesia and Its Influences on Audio-Visual Work.

Chandler, D. W., and Grantham, D. W. (1992) Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *The Journal of the Acoustical Society of America*, *91*(3), pp. 1624-1636.

EA Sports, http://nautil.us/issue/26/color/what-color-is-this-song (retrieved 15 March 2016)

Eimer, M. (2004) Multisensory integration: how visual experience shapes spatial perception. *Current biology* 14(3), pp. R115-R117.

Frassinetti, F., Bolognini, N., and Làdavas, E. (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), pp. 332–343.

Giannakis, K., and Smith, M. (2000) Auditory-visual associations for music compositional processes: A survey. *International Computer Music Conference ICMC2000,* Berlin, Germany.

Goller, A.I.,Otten, L.J., and Ward, J. (2009) Seeing Sounds and Hearing Colors: An Event-related Potential Study of Auditory-Visual Synesthesia. *Journal of Cognitive Neuroscience*, 21(10), pp. 1869–1881.

Haque, F., Nancel, M., and Vogel, D. (2015) Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. *33rd Annual ACM Conference on Human Factors in Computing Systems*, 3653–3656, ACM.

Jankowski, J., and Martin H. (2013) A Survey of Interaction Techniques for Interactive 3D Environments. *Eurographics2013*, Girona, Spain, 65–93, Eurographics Association.

Kawato, M. (1999) Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, *9*(6), pp. 718–727.

Lécuyer, A. (2009) Simulating Haptic Feedback Using Vision: A Survey of Research and Applications of Pseudo-Haptic Feedback. *Presence*, 18(1), pp. 39–53.

Lewis, S. (2010) *Seeing Sound: Hans Jenny and the Cymatic Atlas*, University of Pittsburgh, Pennsylvania, USA.

Marks, L. E. (1975) On Colored-Hearing Synesthesia: Cross-Modal Translations of Sensory Dimensions, Psychological Bulletin 82(3), pp. 303–31

Mazumdar, A. (2013) Principles and Techniques of Schlieren Imaging Systems, *Columbia University Computer Science Technical Reports*.

MyoMapper, https://github.com/balandinodidonato/MyoMapper (retrieved 22 March 2016)

Nancel, M., Chapuis, O., Pietriga, E., Yang, X. D., Irani, P. P., and Beaudouin-Lafon, M. (2013) High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Paris, 27 April – 2 May 2013, 831–840, ACM.

Nymoen, K., Haugen, M. R., and Jensenius, A. R. (2015) MuMYO–Evaluating and Exploring the MYO Armband for Musical Interaction. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Baton Rouge, LA, USA, 31 May 31- 3June, 2015, NIME'15

Omata, M., Go, K., and Imamiya, A. (2000) A gesture-based interface for seamless communication between real and virtual worlds. In *6th ERCIM Workshop on User Interfaces for All*.

Partesotti, E. and Tavares, T. F. (2014) *Color and emotion caused by auditory stimuli*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library.

Platz, F. and Kopiez, R. (2012) When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception: An Interdisciplinary Journal*, *30*(1), pp.71–83.

Polzella, D. J., and Biers, D. W. (1987)Chromesthetic responses to music: Replication and extension. *Perceptual and Motor Skills*, *65*(2), pp. 439-443.

Pure Data, https://puredata.info/ (retrieved 22 March 2016)

Processing, https://processing.org/ (retrieved 22 March 2016)

Saberi, K., and Perrott, D. R. (1990) Minimum audible movement angles as a function of sound source trajectory. *The Journal of the Acoustical Society of America*, *88*(6), pp. 2639-2644.

Schaeffer, P. (1966)Traité des objets musicaux. Editions du Seuil, Paris.

Sega Amusements, https://www.youtube.com/watch?v=WjORqfFk64s (retrieved 15 March 2016)

Settles, G. S. (2012) *Schlieren and shadowgraph techniques: visualizing phenomena in transparent media*. Springer Science & Business Media.

Shneiderman, B. (1997) Direct manipulation for comprehensible, predictable and controllable user interfaces. In *Proceedings of the 2nd international conference on Intelligent user interfaces*, Orlando, FL, USA, 06–09 January 1997, 33–39. ACM.

Speigle, J. M., and Loomis, J. M. (1993) Auditory distance perception by translating observers. *Virtual Reality, IEEE Annual International Symposium*, San Jose, CA, USA, 16–23 March 1993, 92–99, IEEE Computer Society .

Spence, C., and Driver, J. (2000) Attracting attention to the illusory location of a sound: reflexive cross modal orienting and ventriloquism. *Neuroreport*, *11*(9), pp. 2057–2061.

Steven, G., and Korsgaard, D. (2015) An Exploratory Evaluation of User Interfaces for 3D Audio Mixing. *Audio Engineering Society Convention 138*.

Stockhausen, K. (1991) Studie II. *Multimediale 2*.

Storey, M.-A. Fracchia, F. D, and Müller, H. A. (1999) Cognitive design elements to support the construction of a mental model during software exploration. *Journal of Systems and Software*, 44(3), pp.171–185.

Storms, R. L. (1998) *Auditory-Visual Cross-Modal Perception Phenomena*, Naval Postgraduate School Monterey, CA, USA.

Taylor, H. G., and Waldram, J. M. (1933) Improvements in the schlieren method. *Journal of Scientific Instruments*, *10*(12), pp. 378.

Thalmic, https://www.myo.com/ (retrieved 22 March 2016)

Tran D., Haider N. N., Valavan S. E., Yarovyi O., Ligthart L. P., Lagerl. E., and A. Szilagyi A.. (2012) Architecture and Design Procedure of a Generic SWB Antenna with Superb Performances for Tactical Commands and Ubiquitous Communications. *Ultra Wideband – Current Status and Future Trends*.

Vorländer, M. (2007) *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media.

WB Kids, https://youtu.be/XqJqZKL8uEU (retrieved 15 March 2016)

3Dception, http://www.twobigears.com/3dception.php (retrieved 12 February 2016)