Creating tangible spatial-musical images from physical performance gestures

Natasha Barrett
Institute for Musicology, University of Oslo.
Postboks 1017 Blindern
0315 OSLO
nlb@natashabarrett.org

ABSTRACT
Electroacoustic music has a longstanding relationship with gesture and space. This paper marks the start of a project investigating acousmatic spatial imagery, real gestural behaviour and ultimately the formation of tangible acousmatic images. These concepts are explored experimentally using motion tracking in a source-sound recording context, interactive parameter-mapping sonification in three-dimensional high-order ambisonics, composition and performance. The spatio-musical role of physical actions in relation to instrument excitation is used as a point of departure for embodying physical spatial gestures in the creative process. The work draws on how imagery for music is closely linked with imagery for music-related actions.

Keywords
Electroacoustic music; ambisonics; motion gesture analysis; sonification; tangible spatial images, embodied cognition.

1. INTRODUCTION
Historically, electroacoustic music has established close ties between source-sounds and spatial composition: ties between the experiential source-sound recording session in terms of both the kinetic motion of the sound-making object and the performer's excitation of the instrument, audio recordings that feature spectral and dynamic changes *implying* spatial gesture, experiential and recorded spatial relationships between individual sound elements in a 'scene', and the way spatial ideas are developed compositionally. This paper examines some of the relationships between gestural behaviour, sound objects and tangible sound images in composition and performance using motion capture, sonification and artistic creation. The work is based on a premise linking imagery for music with imagery for music-related actions, where the acousmatic spatial image can be regarded and perceived as an embodiment of the human agent, conveyed via sonification as an auditory display (to some extent proposed by Roddy and Furlong [14]). I take an experimental approach to motion defined spatial sound images beginning in the electroacoustic source-sound recording session: a close-microphone recording, an untrained performer intuitively exploring an object for its potential as an acousmatic source and 3D motion tracking of both performer and instrument creating spatial-motion datasets. The recordings are analysed for correlations between performed action, instrument motion and changes in the spectrum over time. This analysis was then used to indicate how motion data and sound recording can be decomposed into spatial gestural units as sources for 'composed' datasets. These new datasets are used in interactive parameter-mapping 3D sonification. The results suggest implications for human agency and embodied cognition in the creation of tangible acousmatic images.

2. BACKGROUND
2.1 The 'object' in focus
Defining a sound object in gestural terms is an interesting challenge. As an extension of Schaeffer's sonorous object, Godøy [8] proposes the 'gestural-sonorous object' where we, 'recode musical sound into multimodal gestural-sonorous images based on biomechanical constraints… hence into images that also have visual (kinematic) and motor (effort, proprioceptive, etc.) components.' Although it is outside the scope of this paper to pursue a phenomenological discussion, we can ally this assumption with some practical observations from the acousmatic and electroacoustic creative process: whether a duration of sound is perceived as an object depends on many parameters such as changes in the spectrum over time and any extrinsic semantic inferences within the current musical frame of reference. These in turn vary depending on micro-, meso- or macro-temporal listening strategies and composed structures. We already see a problem: there is no guarantee that listening and compositional strategies overlap; and this is an experience that becomes particularly fickle in a larger work. However, if we assess how compositional source-sounds are acquired, we find that the gestural-sonorous object is embedded in the recording and performance technique: (1) a biomechanically defined gesture excites the physical-kinetic resonant properties of a sound-making object, (2) a single gestural articulation may be prolonged through an improvised gestural extension, after which the sound is left to naturally decay, else is dampened or interrupted, (3) the performer then allows the sound and body to pause, at which point, (4) a clear gestural unit has been recorded and is available for compositional selection and development. This specific technique used in traditional electroacoustic source-sound recording creates natural gestural sonorous objects that can be used as units in the analysis.

If we turn from composition towards listening and experience, the role of gesture is clear: Cox [3] presents the 'mimetic hypothesis', and Godøy [10] the motor-mimetic component in music perception and cognition, both involving the mental simulation of sound-producing gestures when we perceive and/or imagine music. Furthering this line of enquiry Godøy draws on motor equivalence, where gestural categories are transferable from one setting to another. As we shall see later, motor-mimesis may bond perceptual theories of embodied cognition with the material qualities of sound in Smalley's idea of 'gestural surrogacy' [15] and its origins in composition and listening, which assumes that we assess the gestural process starting in the material of the sound, after which we identify sources and causations that may detach from experienced physical gesture and sounding sources. Our degree of understanding falls along a continuum from the 'primal gesture', which is an individual's concept of gesture outside the sounding flow of the music, through sounding domains, to 'remote surrogacy', where human actions behind the sound disappear leaving only remnants of gesture implied by tensile and proprioceptive properties. We shall see in section 4.1.1 how 3D gesture injects tangible experiences into third order and remote surrogacy.

2.2 Sonification

For the experiments described in this paper, the interactive parameter-mapping 3D spatial sonification program 'Cheddar' was used. Cheddar has been discussed at length in [1]. In brief, Cheddar sonifies multiple 3D spatial datasets in up to 5th order 3D and 11th order 2D higher order ambisonics. The user can freely explore the spatial world: the virtual listening position is freely controlled and the sound-field recalculated in real-time to maintain accurate spatial relationships between data points and the listening location. Sound is transformed by the data with a flexible, user-defined mapping system.

Other approaches to gestural sonification include Polotti's EGG [13], Kapur et al. [12] and Grond et al. [11] both using the VICON motion tracking system, and Diniz et al. [4] using the Optitrack motion capture system. There are many examples where motion tracking and sonification are used in a dance context. However, such gestures are defined for their visual effect, whereas in music, the physical relationship between body, instrument and sound take precedence over visual results.

3. EXPERIMENTS IN SPATIAL IMAGERY

This section presents the method to capture the 3D sound object and express motion defined spatial imagery with sonification. It goes on to postulate how the method can be used as an acousmatic composition and performance tool. The study is intended as a pilot project for future investigations into 3D sound imagery.

3.1 Sonification

3.1.1 Motion and sound capture

In a recording session, one person explored the sonorous properties of a cymbal using the method described in section 2.1. Seven microphones were used: four cardioids placed 90 degrees around the edge of the cymbal, one omni-directional above and one below the centre. A seventh reference microphone was synchronised to the motion recording. 3D motion data was captured using the Qualisys optical motion-capture system and eight Oqus 300 cameras. Seven passive markers were placed on the cymbal and 21 on the body. Motion data was recorded at a rate of 250 Hz.

3.1.2 Preliminary data selection and data analysis

With the available resources, Cheddar could sonify four data streams simultaneously in real-time, meaning four markers could be addressed in real-time ([1]). As sound images are composed from more than a single point in space or from more than one motion trajectory, four trajectories created a reasonable preliminary image. In future work we will use data analysis to detect correlations between body, instrument and sound motion as a way of identifying the most relevant markers. This experiment used an ad hoc approach where it was apparent that the thumbs were the most significant cymbal excitation points and an obvious choice for sonification. The four cymbal perimeter makers were also selected. Each of these six datasets was formatted appropriately for Cheddar (time, 3D spatial coordinates, velocity and acceleration). As a preliminary multi-modal assessment of the cymbal's 'audio spatial image' in relation to the thumbs 'visual spatial image', the microphone signals were encoded in ambisonics for a variety of virtual listening positions. Motion data from the four cymbal markers was used to spatialise the corresponding microphone signals in ambisonics, creating realistic a cymbal motion. Audio images and motion visualisations were then synchronized. This proved useful in anticipating multi-modal versus purely audio archetypes later in the analysis. Example 1 is one such video. The virtual listening location is at ear height above the cymbal.

3.1.3 Sonification and mapping

Two sets of data were used in the sonification: (1) four perimeter cymbal markers and (2) two opposite cymbal markers and two thumb markers. These sets allowed comparison between solo cymbal images and images of the cymbal correlated to physical excitation. In the scope of this paper only the second dataset is discussed.

A shortlist of sound inputs were chosen for the sonification: an extract from the cymbal recording, an abstract source with vague pitch implications, filtered noise synthesized in real-time, a noisy cracking sound and a bumblebee recording. These sounds were aurally chosen based on their variety of intrinsic and extrinsic features, some supporting, (e.g. the cymbal sounds) others in contrast to the context of the data (e.g. the bumblebee sound). For a detailed discussion on sonification source sound inputs see [1]. Vertical space was mapped to pitch shift and to noise filter cut-off. In tests, velocity was found to be a more useful aural gestural index than acceleration and was thus mapped to amplitude and to other event enhancement parameters. Distance-amplitude scaling was applied in the spatial synthesis with an understanding that when we change the location of the virtual listening position we must be careful that proximity effects are not confused with other changes in the mapped variables.

3.2 Results

3.2.1 Human and instrument gesture

In performance it is generally clear that the sound-producing gesture starts before and continues after the sound. Figure 1a shows the relationship between the motion velocity (top section), recorded audio waveform RMS (middle section) and vertical motion component of one thumb marker and its closest cymbal marker (bottom section). As expected, we see a clear synchronisation between cymbal motion and cymbal sound. The performance gesture precedes the sound gesture with a pattern that is similar but accentuated. We also see that some type of interruption occurred at 0'12 where markers are suddenly stable yet the cymbal is offset. When this cymbal motion is sonified, these moments are clearly audible (example 2a), yet also raises the question of excitation. When we introduce the thumbs motion and rescale the spatial scene and the transformation parameters in the sonification to include the new data ranges, the totality of the performance image is more clearly heard (example 2b). The start and end points of this combined sound-gesture object are clear, as are the spatial interactive dynamics within the acousmatic image.

3.2.2 Gesture decomposition and re-composition

Henbing and Leman [5], showed how sonic gestures reflect sound-producing gestures, and that these gestures can be understood as concatenations of more elementary gestural components, where the rules for recombining these elements are defined by natural and cultural constraints. As suggested by Godøy [9], we can segment motion data by identifying changes in speed, amplitude or other phase transitions. By comparing the four motion trajectories in my experiment with their corresponding audio analysis, some of these elementary gestural components were identified. The corresponding data segments were isolated and subsumed into new higher-level

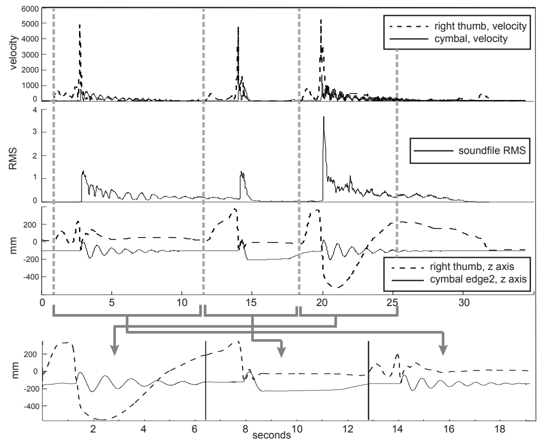

Figure 1a (top): motion and audio from one recorded take. Figure 1b (bottom): a fictional sequence from three units.

chunks to create fictional yet plausible datasets composed from many recordings. Figure 1b is a short example: three gestural sound-space objects are identified from figure 1a and then assembled into an extended fictional gestural sound-space object with an expanded timeline, capturing motion archetypes from the original. Example 3a is a sonification of this new data set using the same mapping as example 2b. This approach to composition is guided by heuristic knowledge as to what is intuitively plausible, informed by sonification tests as the composed dataset grows. The method has implications for extended spatial morphologies discussed in section 4. Although in this experiment the spatial gestural source units are not systematically decomposed into smaller atoms, the results suggest coarticulation (contextual smearing) analysis may be a fruitful line of further study.

3.2.3 Results from interaction

Real-time interaction in the sonification process revealed interesting spatial-gestural behaviours. Time scaling was particularly interesting: gestures played at original speed were inherently clear, composed of consecutive points in space, while temporal modulations resulted in degrees of spatial-gestural ambiguity. This suggests that space-time relationships are a key factor in gestural expression. Example 3b illustrates this by first halving and then quartering the tempo of the sonification in example 3a.

Scattering the events in a trajectory created a cloud that traced the original gesture and created an image from a single moving point. Rather than a random spatial scattering, this effect was achieved by data decimation and time- and event-duration-scaling in the sonification. A coherent continuum then resulted from mapping longer sounds to each data point creating a temporal-spatial overlap. Example 4a sonifies the left thumb trajectory, with the listening location this time central to the thumb motion, without scattering. Example 4b adds scattering and still displays the original gestural behaviour while appearing more as a moving image rather than a moving point.

Manipulating spatial scaling and the virtual listening location revealed relationships between gestural zones and between local and global behaviour. A small scaling factor would pull events into a close spatial context revealing global behaviour, while a large scaling factor enhanced motion speed and revealed differences in local behaviour where events far away would vanish into silence and dynamically reappear as they approached the virtual listening location.

Moving the virtual listening location highlighted motion characteristics. For example, by moving the virtual listening location to the centre of the left thumb motion, as in example 4, its behaviour is highlighted. This is in contrast to example 2 where the listening location is central to the four-marker image. The result confirms the intuitive assumption that the listening location is important in our perception of spatial imagery.

The importance of human gesture became particularly apparent when different sounds inputs were used in the sonification. Setting the same sound as input for all four datasets created a coherent image, yet when cymbal and thumbs were allocated different sources, the contrast accentuated the cymbal as an instrument and body as an activator. When using the bee sound, despite the 'wrong' semantic link, human gesture in relation to the cymbal is nevertheless quite clear (example 5). We can speculate that for our perception, the 'goal' is the moment of contact, and this is more significant than semantic links evoked by the input sound.

3.2.4 Performance and tangible images

In a second experiment motion data was streamed in real-time from four makers on a performer's hands, improvising in free air while listening to the ambisonics decoded over 24 loudspeakers surrounding the space. In this improvisation we experienced an interesting feedback between motion and sound: motion actions were visually unrepresentative of the gestural sound space, highlighting that, for proprioceptive control of sound, it is misleading to treat visual physical gestures, sonic embodiment and the resulting sonification as similar. This is hardly surprising if we draw on 'sound tracing' studies [7], in which there was found to be fair agreement in the correlation of pitch contours and salient dynamic patterns, but less agreement for multidimensional timbral features. We can go on to say that visual dislocation liberates the original cause-effect coupling leading to a greater sense of tangibility.

4. DISCUSSION

4.1 Embodied cognition and tangible images

Although scientifically accurate sonification is somewhat limited in terms of sounding results, one of the main advantages of a system applying HOA is the detachment of spatial gesture from the original sounding event. In this way we can investigate spatial expression in auditory terms and assess the spatial significance of gesture without being mislead by the original multi-modal inferences. Cheddar was designed with both science and musical expression in mind: in sonifying gestural musical data it serves as an acousmatic performance instrument of spatial structures.

The results in section 3.2.3 indicate that spatial gesture may be just as important as spectral gesture in our understanding of performance action. Consequently we may adopt a theory of embodied cognition where spatial information evokes the mental simulation of the gesture action and it is this that renders sounds of remote causality tangible in an acousmatic context. Thus, tangible spatial-gestural images may be maintained even when acousmatic spectromorphology leans towards remote surrogacy.

4.1.1 Tangible imagery, gestural surrogacy, embodied cognition and motor equivalence

Example 6 is a 1'30 extract from the opening of a live performance of Topology Chamber 1 [1], created using the same method as in the experiment above, combined with original cymbal recordings. We hear that the scaling of temporal-spatial sonification creates distortions, which lead to gestural abstraction, yet the sounding result continues to bear meaning in terms of motor-equivalent spatial gesture. These

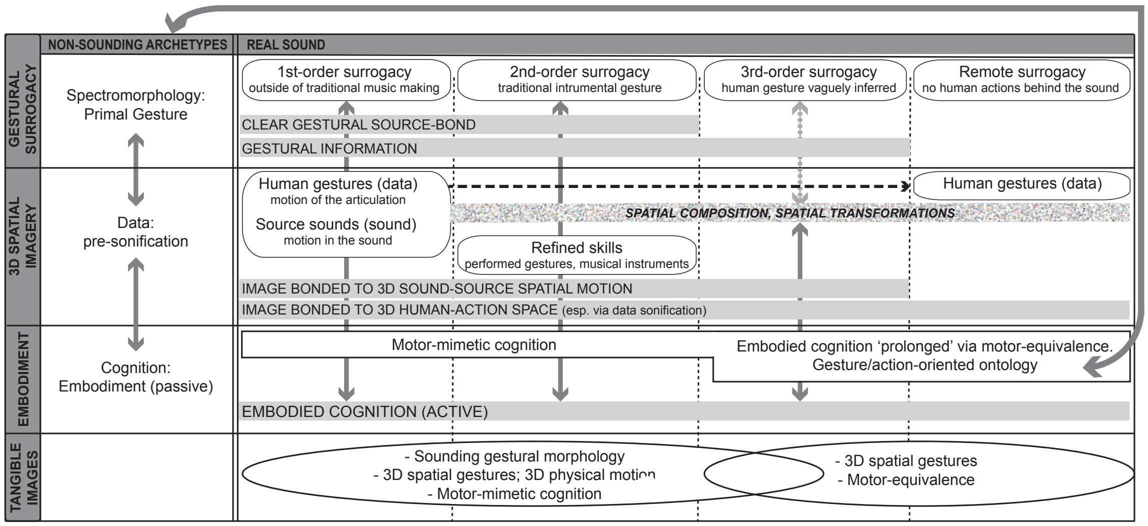

Figure 2. The relationship between gestural surrogacy, 3D spatial imagery, embodiment and tangible images.

salient spatial features of human actions serve to extend third-order and remote surrogacy into the embodied experience.

Figure 2 postulates how these phases may occur for non-sounding archetypes and real sound. The horizontal sections present elements involved in the formation of tangible acousmatic images: gestural surrogacy, 3D spatial imagery and matters concerning embodiment. The columns indicate how these elements interact. The top row of the diagram presents gestural surrogacy as conveyed by Smalley. Clear gestural source-bonds are revealed in first- and second-order surrogacy. Remote, yet nevertheless relevant gestural source-bonds are the makeup third-order surrogacy. The second row shows 3D spatial imagery and how it relates to gestural surrogacy and embodiment. In the third row, motor-mimetic cognition gives way to motor-equivalence during the transformation into third-order surrogacy for both sound and spatialisation. The fourth row suggests how tangible images are composed from different sections of the grid depending on what information is available. In the first column primal gesture is tightly allied to latent embodied cognition. The second and third columns connect the first three rows. In the forth column we encounter a dislocation of third-order gestural surrogacy from image formation, that worsens for remote surrogacy in the fifth column. However, we also encounter spatial motions that bond to human-action regardless of gestural morphology in the sound spectrum. Spatial gestural motion of the complete image evokes motor-equivalence even when, in terms of remote surrogacy, the immediate human action behind the sound appears remote.

Finally, we can identify a circular loop between primal gesture and motor-equivalence, which may be learnt through the listening process and possibly explain why acousmatic music composed of more abstract morphologies appeals to familiarised listeners more than to audiences of conventional instrumental music: familiarity with a genre conditions the perception–action cycle, where we relate what we hear to some image of the assumed sound-producing action behind what we hear, or to some other gestures that we assume correspond with the music [6].

4.2 Further work

The work presented here intends to lay a foundation for further research into listeners' experience of tangible spatial imagery. Many of the hypothesis and results will benefit from listener tests. Different approaches to creating and applying 3D gestural data in real-time sonification image formation will be explored further in future composition and performance contexts.

5. REFERENCES

[1] Barrett, N. Design and application of 'Cheddar': an interactive 3D parameter mapping sonification program. In review. www.natashabarrett.org/Cheddar.pdf 2015.

[2] Barrett, N. Topology Chamber 1. Music performance. NIME'15, Louisiana State University, May 31-June 3, 2015.

[3] Cox, A. The mimetic hypothesis and embodied musical meaning. Musicae Sclentiae, 5(2): 195-21, 2001.

[4] Diniz, N., Coussement, P., Deweppe, A., Demey, M., Leman, M. An embodied music cognition approach to multilevel interactive sonification. J Multimodal User Interfaces, 5:211–219, 2012

[5] Henbing, Li., & Leman, M. A gesture-based typology of sliding-tones in guqin music. Journal of New Music Research, 36(2): 61-82, 2007.

[6] Godøy, R. I., Leman, M. (Eds.). Musical gestures: sound, movement, and meaning. New York (N.Y.): Routledge. 2011

[7] Godøy, R. I. Images of Sonic Objects. Organised Sound, 15(1):54-62, 2010.

[8] Godøy, R. I. Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. Organised Sound, 11(2):149-157, 2006

[9] Godøy, R. I., Coarticulated Gestural-sonic Objects. In New Persecptives on Music and Gesture. Ashgate.

[10] Godøy, R. I. Motor-mimetic Music Cognition. Leonardo 36(4): 317–9, 2003.

[11] Grond, F., Hermann, T., Verfaille, V., Wanderley, M. Methods for Effective Sonification of Clarinetists' Ancillary Gestures. In Gesture in Embodied Communication and Human-Computer Interaction. Lecture Notes in Computer Science Volume 5934:171-181, 2010

[12] Kapur, A., Tzanetakis, G., Virji-Babul, N., Wang, G., Cook, P. A Framework for Sonification of Vicon Motion Capture Data. In Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05), Madrid, Spain, September 20-22, 2005

[13] Polotti, P., Goina, M. EGGS in Action. NIME'11. Oslo, 30 May - 1 June 2011.

[14] Roddy, S., Furlong, D. Embodied Aesthetics in Auditory Display. Organised Sound 19(1): 70–77, 2014.

[15] Smalley, D. Spectromorphology: explaining sound-shapes. Organised Sound 2(2):107-126, 1997.