Motion as the Connection Between Audio and Visuals

Niall Moody, Dr Nick Fells and Dr Nicholas Bailey
Centre for Music Technology, University of Glasgow
niallmoody@yahoo.co.uk, n.fells@music.gla.ac.uk, N.Bailey@elec.gla.ac.uk

Abstract
This paper explores the relationship between sound and (moving) image in the context of designing a computer-based musical instrument (or, strictly, an audiovisual instrument). Based on Michel Chion’s notion of synchresis, a preliminary framework is developed to guide the design of new audiovisual instruments. This framework uses motion to form connections between the two elements.

1 Introduction
Over the years there have been numerous works of art which explore audiovisual relationships. This paper looks at the relationship with a view to developing an audiovisual instrument. Primarily concerned with how audio and visuals may be perceived as connected, the paper presents the idea that motion may be used to form this connection, based on Michel Chion’s notion of synchresis. The paper finishes with a discussion of audiovisual mappings and a brief description of the instrument currently being designed based on these principals.

2 Background
A full contextual discussion of the extensive history of audio-visual artworks is beyond the scope of the current paper; however there are two works in particular that highlight the main ideas espoused in it1. Firstly, looking at Oskar Fischinger’s ‘Studie Nr. 7’ (Fischinger 1931), the visuals consist of a number of white polygons move and change shape on a black background in a way that seems to mirror the motion of the accompanying music (Brahms’ ‘Hungarian Dance No. 5’). Here there is a very tight connection between audio and visuals, a connection that seems to be entirely based on motion (certainly, the connection does not seem to exist when viewing stills of the visuals). Secondly, a related piece with a similarly strong audiovisual connection is Alex Rutterford’s video for Autechre’s Gantz Graf (Rutterford and Autechre 2002), where the visuals consist of an abstract rendered object which dynamically changes shape. Again, the motion of this object seems to reflect the motion of the music, creating a very tight perceptual connection between the two elements.

3 Synchresis
Much of the content of this paper is based upon Michel Chion’s notion of synchresis. According to Chion, synchresis (created out of a combination of the words synchronism and synthesis) is “the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time”2. What he is referring to is an effect that has been commonplace in film for a number of years now, and (to borrow one of Chion’s examples) can be particularly noticeable when looking at the way punches are often represented in film. In real life, punches rarely make much sound - whether they inflict pain or not - yet in film it is relatively rare that punches are portrayed in a naturalistic way (where the sound heard is exactly what would be heard in real life). Instead, we are accustomed to hearing assorted exaggerated whacks and thumps when a punch connects, the punch almost seeming unreal, and somehow false, if such sound is absent. The point is that the sound, though not actually related to the images we’re seeing in any physical way, somehow enhances the image (Chion terms this enhancement “added value”3), and makes it seem more real. Our brain recognises the synchrony between image and sound and intuitively creates a connection.

While it could perhaps be argued that, in the case of the punch, the exaggerated sound is necessary to convey the physical nature of the action (in that it may not necessarily be con-

1For a comprehensive overview of the subject, see ‘Visual Music: Synaesthesia in Art and Music Since 1900’(Brougher, Strick, Wiseman, and Zilczer 2005).
2(Chion 1994), p.63
3Ibid, p.5
veyed entirely successfully via image alone), Chion claims that it is nevertheless possible for the visuals and sound to be entirely unconnected (i.e. the only thing that links them is their synchresis). Indeed, there are numerous examples of synchresis in films where there is no connection between audio and visual other than their synchrony. For example, a relatively common instance is that of a visual shot of someone walking and, instead of hearing footsteps, orchestral hits are played in sync with each step (this is particularly common in, for example, looney tunes cartoons). While the sound and visual are entirely unrelated if viewed separately, the tight synchrony encourages the viewer’s brain to make a connection, to the extent that these two unrelated sequences come to be viewed as a single object.

Chion notes that not all sounds and images may be connected as simply as this however, and states that synchresis “is also a function of meaning, and is organised according to gestaltist laws and contextual determinations”⁴. The result is that certain sounds will ‘adhere’ to a particular visual better than others, and that this will often rely significantly on the context within which the connection is made. To return to the footsteps example previously, Chion views this as having an “unstoppable”⁵ synchresis, such that it would be possible to attach any kind of sound to the image without breaking the connection between the two. This is due to our experience of the world - we learn from experience that footsteps make a sound, and, in viewing a sequence of someone walking, we expect to hear a corresponding sound - what that sound actually is, is less important than the fact that the sound occurs when we’re expecting it.

4 Motion as the Connection

Using synchresis, we can see that it may be possible to create a connection between audio and visual that, though not strictly objective, will nevertheless be perceived in much the same way by anyone who experiences it, irrespective of cultural considerations. Looking more closely, however, what is actually happening when we experience the synchresis of the footsteps, for example? We see the foot moving in a particular way, and we hear sound accompanying (or from a different perspective, reacting to) that motion. Indeed, it is the author’s contention that synchresis of this form is based on the fact that the motions of the two domains (visual and aural) are related in some way. With the footsteps, we see the foot moving downward, then coming to a sudden stop, at which point a sound event is initiated. This sound event imparts various pieces of information, but looking at its motion, we can see that its amplitude envelope, for a start, is closely related to the visual motion of the foot. When the foot comes to a halt on the ground, the amplitude envelope of the sound in a sense reacts, in that there is a sudden sharp increase in the sound’s level, following which the amplitude decays, as the foot is no longer in the process of colliding with the ground. There is of course additional motion in the sound (spectral content for example), but from looking at the amplitude envelope alone we can already see a clear link between image and sound in terms of motion.

This form of ‘collision-based’ motion - where the primary stimulus is the sudden collision of two visual objects - is not the only form of motion which can act as a connection between the two domains however. Another form of motion is the unhindered motion of an object in a linear trajectory across the screen. The author is of the opinion that this kind of motion can prove just as powerful a connection as the collision-based form, provided it is accompanied by a related motion in the audio realm. To understand how this may be, we need to look at our experience of the sound made by objects thrown through the air. A filmic example of this would be in period movies with battle scenes where arrows are used (see for example Gladiator, or any other such film) - when the arrows are flying through the air, there is an accompanying ‘whoosh’ sound. Another example would be the sound of a jet plane in flight. The point is that experience tells us that objects that move through the air tend to make a sound (albeit provided they are moving relatively fast).

Indeed, the idea that motion can act as a connection between audio and visuals is based on our experience of the world. Our experience of the physical world tells us that objects which we can see are in motion will tend to emit sound, as a consequence of this motion (from this perspective, the visual stimuli telling us the object is in motion is also a consequence of the motion). This experience is what allows synchresis to work - our experience tells us to expect some kind of sound in conjunction with certain visual stimuli (and vice-versa, depending on the situation), and our brain, expecting this aural ‘event’, will connect almost any sound to the visual, assuming there is some kind of related motion between the two.

At this point a caveat should be made regarding the kind of motion which may be put to use as this kind of connection. It is important that it is perceivable motion. To elaborate, a constant sine tone in the audio realm could be considered as possessing a certain motion, in that it relies on the vibration of particles in air in order to be audible. To the human ear however, the sound produced is fundamentally static (we are assuming the amplitude is constant), as the perception is of a single tone which does not possess any motion of its own. The same applies to the visual realm - if motion is occurring too fast for the eye to perceive, it is hard to see how it

---

⁴Ibid, p.63
⁵Ibid, p.64
Table 1: Forms of Motion

<table>
<thead>
<tr>
<th>Constant Velocity</th>
<th>Collision-Based Motion</th>
<th>Periodic Motion</th>
<th>Gravity-Based Motion</th>
<th>Discontinuous Motion</th>
</tr>
</thead>
</table>

could be useful in establishing a connection with an audio stimulus (though realistically, this may be harder to achieve anyway with current monitors/projectors, since aliasing will come into play before the point where motion becomes blurred).

5 How to use the Connection

Having examined the ways in which motion may connect audio and visuals then, how can this be put to use in the design of new instruments? The first step is to define the various types of motion available to us, with a view to creating some simple audiovisual mappings as a starting point for further work. With motion, the author would make a distinction between forms of motion, and domains in which motion can occur. To elaborate, forms of motion would refer to a high level description of how something moves, where the something could be anything, whether visual or audio (for example, one form of motion would be periodic motion). Domains in which motion can occur, on the other hand, refers to parts of the audio and visual realms where motion can be perceived (the ‘something’ that’s moving: a visual example would be the position of an object of some kind).

5.1 Forms of Motion

Table 1 shows a list of some forms of motion, according to the above definition, though this is by no means intended as a complete list.

- **Constant Velocity**: This should be fairly self-explanatory. Compared to the other forms of motion, this could perhaps be seen as providing a weaker connection between audio and visuals, since there are no discrete temporal events. This does not mean it cannot prove useful in certain situations, however. An example could be based on the experience of a stationary viewpoint watching objects (e.g. cars) moving at a high speed past it - the related sounds would pan and be subjected to the doppler effect accordingly.

- **Collision-Based Motion**: This is primarily derived from the footstep example previously - in the visual realm, it refers to objects colliding with each other and then reacting. In the audio realm, however, it refers to the kind of sound associated with collisions, referring to the way that, while the visuals are in motion before and after the collision, sound will only be instigated at the point of collision (assuming it is not already in motion from a previous collision).

- **Periodic Motion**: Again this should be fairly self-explanatory, referring to motion that repeats itself in a perceivable fashion.

- **Gravity-Based Motion**: Related to collision-based motion in that it is based on physical phenomena, this essentially refers to attraction/repulsion forces such as gravity and magnetism. This is probably most easily perceived visually, though aurally it would refer to motion that gradually decreases or increases in range.

- **Discontinuous Motion**: This refers to sudden discrete jumps, as opposed to the mostly smooth motions described previously. An example would be the rapid cutting common in music videos and also seen in certain films.

5.2 Domains in Which Motion May Occur

As mentioned previously, ‘domains in which motion can occur’ refers to aspects of the visual or aural realms in which motion of the forms discussed above is perceivable. Most of the entries in Table 2 (again, this is by no means an exhaustive list) should be self-explanatory, so rather than go through each one in turn, only the less obvious entries will be discussed here.

- **Smoothness**: This refers to how smooth or coarse a particular part of the visual is. That part could the shape of an object, or a more general impression of how colours (and particularly patterns) contrast with each other.

- **Articulation (of an object)**: This refers to particular visual objects which may articulate their shape, in much the same way as humans and animals do with their arms, legs etc.
Figure 1: Some example audiovisual mappings

- **Pattern**: Refers to a visual motif which is recognisably periodic, whether it is viewed statically or in motion.
- **Brightness**: The perceptual brightness of the sound, can be obtained by calculating the spectral centroid of the sound.
- **Energy Content**: Related to Instantaneous Amplitude, refers to a more general impression of the amplitude of the sound, specifically the amount of energy imparted by the sound.
- **Noise-Pitch**: Refers to the continuum between pitch and noise.

### 5.3 Example Mappings

Figure 1 shows some simple mappings between visuals and audio, keeping to those mappings which are less subjective, and more based on common perceptions of the world, and the audiovisual properties of physical objects (there are of course many more possible combinations). The first one describes our now familiar footstep example (at least in part - strictly speaking the spectral content of the audio should be considered as well). The amplitude of the audio is controlled by the collision of objects in the visual realm. Although it is marked as a one-way process, it could be interesting to then map the audio back to visuals in some way, so that a kind of co-operative feedback can be developed. The second example is based around the idea of a throbbing, pulsating object (one could imagine a beating heart), with the size of the visual object periodically growing and shrinking, and this controlling the brightness of the audio (this could take the form of a low pass filter, where the cutoff frequency is controlled by the size of the visual object). The third example uses the amplitude envelope of the audio (specifically, its transients) to jump a visual object around the screen accordingly, the idea being to create a visual accompaniment to particular audio cues.

### 6 A Preliminary Design

Though still in the early stages, an instrument is being designed based on the aforementioned principles. The instrument is intended as a sort of ‘musical block of clay’, visually represented as a 3d, amorphous blob which responds both to the user’s input and the instrument’s audio output, which will be based on physical modelling synthesis, courtesy of the Tao physical modelling language (Pearson 2005). A physical user interface is also being designed, intended to allow for the kind of gestures possible or common with clay. Compared to Golan Levin’s ‘painterly’ interface metaphor, it could be seen as a ‘sculptorly’ interface. The aim is to connect the motions of the audio and the visuals (as well as that of the performer’s gesture) as tightly as possible, so that the instrument’s output can be viewed as an audiovisual whole, where audio and visual are not easily separated. To do this, a number of mappings based on the aforementioned clay-based interface metaphor will be used (i.e. a ‘squeezing’ gesture could reduce the size of the visual object, and perhaps alter the pitch of the audio accordingly).

### 7 Conclusions

Based on Michel Chion’s notion of synchresis, this paper has proposed a way of using motion to connect audio and visuals. This connection is derived primarily from our experience of the world (in particular, of the audiovisual properties of physical objects), and plays upon our expectations associated with that experience. With a view to demonstrating some example audiovisual mappings, various different forms of motion were examined. Further to this, the design of an instrument currently being developed based on this idea of using motion as a connection was briefly described.

### References


