

The Fungible Audio-Visual Mapping and its Experience

Adriana Sa

EAVI group, Department of Computing,
Goldsmiths University of London, London,
UK

adrianasa@clix.pt

Baptiste Caramiaux

EAVI group, Department of Computing,
Goldsmiths University of London, London,
UK

b.caramiaux@gold.ac.uk

Atau Tanaka

EAVI group, Department of Computing,
Goldsmiths University of London, London,
UK

atau@goldsmithsdigital.com

ABSTRACT

This article draws a perceptual approach to audio-visual mapping. Clearly perceivable cause and effect relationships can be problematic if one desires the audience to experience the music. Indeed perception would bias those sonic qualities that fit previous concepts of causation, subordinating other sonic qualities, which may form the relations between the sounds themselves. The question is, how can an audio-visual mapping produce a sense of causation, and simultaneously confound the actual cause-effect relationships. We call this a fungible audio-visual mapping. Our aim here is to glean its constitution and aspect. We will report a study, which draws upon methods from experimental psychology to inform audio-visual instrument design and composition. The participants are shown several audio-visual mapping prototypes, after which we pose quantitative and qualitative questions regarding their sense of causation, and their sense of understanding the cause-effect relationships. The study shows that a fungible mapping requires both synchronized and seemingly non-related components – sufficient complexity to be confusing. As the specific cause-effect concepts remain inconclusive, the sense of causation embraces the whole.

KEYWORDS

Audio-visual Mapping; Perception; Causation; Multisensory Integration; User Studies.

1 | RESEARCH CONTEXT

Our research is driven by an urge to understand whether, and how, an audio-visual instrument can foster a sense of causation without distracting the audience from the music. From Pierre Schaeffer (1966) to recent acousmatic composers, people have argued that sounds must be detached from their originating cause to be fully experienced. Jeff Pressing also investigated the audio-visual relationship in digital 3D environments, noting that perception operates from vision to audition whenever a direction of causation is discernible. (Pressing, 1997) Not coincidentally, 3D animators often place the sound of a footstep slightly before the foot actually hits the ground.

As practitioners we desire the image to create a reactive stage scene: sensing an audio-visual relationship brings additional meaning to the audience's experience. Michel Chion coined the term added value to describe the expressive and informative value with which a sound affects an image, creating "the definite impression that this information or expression "naturally" comes from what is seen, and is already contained in the image itself." (Chion, 1994) The problem is, "What we hear is what we haven't had time to see." (Chion, 1994, p. 61) When perception biases those sonic qualities that help visual apprehension, it subordinates other qualities such as timbre, texture, vibration, and the nuances of the performer's expression. The wealth of

multilayered relations between the sounds themselves becomes lost for conscious awareness.

Meghan Stevens assessed that the music remains dominant when the audio-visual relationship is partially congruent, but she stresses that her theories were “created from limited evidence.” (Stevens, 2009) Those theories are particularly subjective with abstracting sounds and images, because her notion of congruency relies on the intrinsic qualities of the sounds and the visual shapes.

In a previous publication titled “How an Audio-Visual Instrument Can Foster the Sonic Experience,” we resorted to psychology and neuroscience in order to clarify the problem. (Sa, 2013) At any time, we are presented with an enormous amount of stimuli, which compete to reach conscious awareness. Our perceptual mechanics are driven to chunk the information, thus we can handle large amounts simultaneously, through cues (Snyder, 2001; Knudsen 2007). The information arriving from the individual senses interacts and influences processing in other sensory modalities, conveying a unified experience of multisensory events. The process is known as multisensory integration. In multisensory integration, vision tends to subordinate audition; yet attention can be manipulated so that audition dominates over vision (Sinnott et al., 2007).

Given our research question, this is a crucial finding. It also brings an artistic challenge, seemingly a paradox. On the one hand, the audience’s attention should be manipulated so that vision does not automatically subordinate audition. On the other, attention should be left under the audience’s individual control. Importantly, attention is drawn automatically to stimuli that are infrequent in time or in space, but deliberate attention equally makes for us to improve the quality of related information in all domains: sensory, motor, internal state, and memory (Knudsen, 2007). Whether attention is automatic or deliberate, it makes for us to optimize the resolution of the information under focus. This happens in two ways: by directing sensory organs toward a target, and/or modulating the sensitivity of neural circuits accordingly. Thus the question is: how can we make for perceptual resolution to be optimized for the music, if the audience directs the eyes toward the screen?

We needed a theoretical tool to compare the strength of auditory and visual stimuli. It became clear that the sense of intensity derives from the combined effects of automatic and deliberate attention; as such it depends on the event, the stimuli panorama, and individual predispositions. Thus we created a taxonomy relating intensity and attention with continuities and discontinuities (Sa, 2013). The taxonomy distinguishes those apprehensions driven through stimuli, from those under individual control. It considers that infrequent stimuli attract automatic attention, and deliberate attention equally makes us optimize perceptual resolution. It shows that to keep the music in the foreground one must dispense with disruptive visual changes, i.e. radical discontinuities. One should apply gestaltist principles to visual dynamics, so as to enable perceptual simplification; these principles describe how we organize the perceptual field in the simplest and clearest way possible, deriving the meaning of the components from the meaning of the whole (Rubin, 1921; Koffka, 1935; Wertheimer, 1938; Bregman, 1990; Palmer, 1999; Snyder, 2001). But there can be a wealth of visual discontinuities: the foreseeable logic can shift, yet not disruptively. The visual dynamics can exhibit progressive continuities and ambivalent discontinuities. With progressive continuities, successive events display a similar interval of motion in certain direction. That is not the case with ambivalent discontinuities, which entail irregular motion; they appear continuous at low perceptual resolution, but if one deliberately pays attention, discontinuities become more intense.

Clearly perceivable cause-effect relationships are as problematic for the music as disruptive visual changes. Perception is a process of multi-sensorial synthesis, and as we bind aural and visual we also skew stimuli that do not converge (Pick, 1969; McGurk, 1976; Shams et al., 2002; Schutz and Kubovy, 2009). Kubovy and Schutz (2010) explain that when perception binds sensory information, the visual discounts the aural and the aural discounts the visual. The process is unconscious and presided by mind-dependent concepts, which they call audiovisual objects. The problem is, the perceptual binding (and discounting) is definite when the auditory and the visual stimuli appear unequivocally related; that is, when mind-dependent concepts are conclusive.

The question then is whether, and how, an audio-visual mapping can foster a sense of causation, and simultaneously confound the cause and effect relationships so that they remain inconclusive. We call this a fungible audio-visual mapping. In our mentioned publication (Sa, 2013) we substantiate the fungible mapping as a principle for instrument design and composition. It goes together with two other principles: to threshold control and unpredictability so as to potentiate sonic expression; and to dispense with disruptive visual changes, which would automatically attract attention, subordinating audition. We took other publication opportunities to elaborate on these principles (Sa, 2013, 2014a, 2014b). Accordingly, we also described several iterations of a personal instrument, which outputs acoustic sound, digital sound, and digital image (Sa, 2013, 2014b).

The study reported in this article aims at demonstrating the fungible mapping independently from the other two principles, and regardless of any personal explorations or technical platforms.

2 | THE EXPERIENCE OF COMPLEXITY

Bob Snyder describes how events activate memories that have been previously activated by similar events (Snyder, 2001). Among these memories, only a few become highly activated and conscious. They embed unconscious memories, which he calls semiactivated memories. The question is which part of the information becomes conscious, and which remains unconscious.

Usually perception prioritizes and hierarchizes the stimuli governed by pragmatic criteria, but sometimes it does not. On the one hand, we are conditioned by the rapid decay of short-term memory (Snyder, 2001; Knudsen, 2007) and by the primary aim of the brain: to be efficient in detecting, perceiving and responding to the world (Calvert et al., 2004). When we focus on a goal such as accomplishing a task, perception simplifies the information according to presuppositions; the gestaltist principles are a major example. But on the other hand, we can permeate the broad membrane of complexity formed of sensory information, when we focus on the experience itself.

This broadened way of perceiving is well known in art, and there are many related philosophies. John Cage is an obvious reference, for having explored the

compositional potential of chance in music. He was influenced by Eastern philosophies, which teach us that one needs to suspend any intention to permeate a relation between all things. Before Cage, philosopher Henri Bergson observed that the intellect shields the human mind from what he called prime reality, an evolving dynamic flux where “all events, objects, and processes are unified” (Westcott, 1968, p. 8). Bergson also stressed that intuition is a way to attain direct contact with this reality ordinarily masked from human knowledge, and that the intellect can freely interact with intuition. Furthermore, one may recall Immanuel Kant’s definition of sublime as an extraordinary experience: we fail to understand the greatness of nature by means of determinate concepts, and yet supplant this failure with a delight stemming from our ability to grasp that greatness (Kant, 1790).

Chion described three types of listening, or modes, which imply a differentiation between what we can call pragmatic and non-pragmatic modes of perception. We can extend his terminology into the audio-visual domain.

That which he calls causal listening “consists of listening to a sound in order to gather information about its cause (or source)” (Chion, 1994, p. 28). Causal audio-visual perception is equivalent; it consists of listening to the sounds and viewing the images in order to gather information about the audio-visual mapping mechanics.

Chion’s second type of listening is semantic. It “refers to a code or a language to interpret a message” (1994, p. 28). Semantic audio-visual perception consists of listening to the sounds and viewing the images while focusing on a goal beyond the perceptual experience, as happens for example in video gaming.

The third mode of listening derives from Schaeffer’s reduced listening, which “focuses on the traits of sound itself, independent of its cause and its meaning” (Chion, 1994, p. 29). Chion provides perspective by stating that hiding the source of sounds “intensifies causal listening in taking away the aid of sight” (1994, p. 32). In applying to the audio-visual domain, we consider how “reduced” might refer

to stripping the perceptual experience of conclusive causes and meanings.

Conclusive concepts require perception to segregate the information, and prioritize the converging over the diverging. In fact, every concept is an abstraction, thus a simplification. So what happens when concepts remain inconclusive? Speaking of the neural implications of artistic experience, Susan Broadhurst says that “all works of art that (...) frustrate our expectations of any clear resolution (...) are likely to activate a specific area of the frontal lobe which appears to deal with the resolution of perceptual/experiential conflict” (Broadhurst, 2007, p. 58).

If the perceived does not conform to previous concepts, perception is likely to keep seeking for a resolution. The process can possibly bring to conscious awareness semi-activated memories, which would otherwise remain unconscious. Modulating the sensitivity of neural circuits can make for the surreptitious to become intense.

Art invites us to shift our usual ways of perceiving the world. We make use of assumptions, and simultaneously we draw upon their ambivalences. The work can be accessed via multiple perspectives. It can transcend any of these perspectives, indicating their subjectivity.

3 | AUDIO-VISUAL BINDING AND SYNCHRONY

As we process sensory information, divergences across the sensory modalities can produce phenomena known as multisensory illusions, derived from automatic interactions between the senses in multisensory integration. Well-known examples are the ventriloquist effect (Pick et al., 1969), in which a sound source is dislocated towards a seemingly related visual stimulus; the sound-induced double-flash illusion (Shams et al., 2002), in which a visual stimulus is doubled when juxtaposed with a set of tones; and the McGurk effect (1976), in which non-matching speech sounds and lip movements are perceived as a new phoneme that diverges from both.

Perception seems anxious to bind sensory information, and the role of synchrony is multifaceted. In film theory, Chion (1994) states that synchrony leads inevitably to perceptual binding. In experimental psychology, Kubovy and Schutz (2010, 0. 57) state

that it does not. This seems in contradiction, but it may be not. Whereas Kubovy and Schutz investigated binding at low level in information processing, Chion may refer to binding at a higher level of perceptual organization.

Schutz and Kubovy (2009) conducted a study about the perception of a percussive action, recorded on video. By manipulating the synchronization between sound and image, they observed that synchrony does not inevitably lead to automatic sensory interactions. Sensory interactions depend on the strength of perceptual binding, which in turn depends on what they call the ecological fit between auditory and visual information. The ecological fit corresponds to how the stimuli conform to previous mind-dependent concepts, called audiovisual objects (Kubovy and Schutz, 2010). In the study, the visual impact over a marimba fitted naturally with a percussive sound, but not with a piano sound. Thus, when sound and image were slightly desynchronised (up to 700ms), the perceived marimba sound was shortened so as to coincide with the visible impact, but the piano sound was perceived in full length.

Chion (1994, p. 23) wrote that synchrony is more determining in audio-visual binding than verisimilitude: “To the spectator, it is not the acoustic realism so much as synchrony above all, and secondarily the factor of verisimilitude (verisimilitude arising not from truth but from convention), that will lead him or her to connect a sound with an event or detail.” Chion also coined the term *synchresis* to describe “the forging of an immediate and necessary relationship” between synchronized sounds and images, whether their combination is plausible or implausible (Chion, 1994). We can say that a person binds sound and image because it is implicit that they have a common origin: the film. As film frames attention, synchrony invites us to discern an audio-visual linkage, even if the link seems implausible in terms of meaning.

We can say that binding sounds and images while finding the combination implausible is forming inconclusive audiovisual objects. Perception is impeded to automatically discount information according to presuppositions, because the perceived does not conform to previous concepts. Since audiovisual objects remain inconclusive, the stimuli competition remains unsolved. Thus, sensory

interactions can be undone, and discounted information can potentially reach conscious awareness.

So how do we form inconclusive audiovisual objects with abstracting sounds and images, which do not refer to anything but themselves? One can easily assess whether the linkage between representative sounds and images is plausible or implausible. For example, the visible impact over a marimba and the marimba sound combine in a plausible way, whilst the combination with a piano sound seems implausible. But, what makes for the relation between abstract sounds and images appear plausible or implausible? And moreover, is there an intermediary state of plausibility?

The challenge here is to create an audio-visual mapping that forms causal percepts, but also throttles the fit between the sonic and the visual events, so that these percepts remain inconclusive. As the audience senses causation, they may seek to understand the instrument. As cause-effect relationships become confusing, they are invited to focus on perception itself.

4 | STUDY

In this section we report a study on how an audio-visual mapping can produce a sense of causation, and simultaneously confound the cause-effect relationships. The study aims at quantitative and qualitative data from the participants' subjective experience.

4.1 HYPOTHESIS

A fungible audio-visual mapping may combine mappings that convey a sense of causation, and mappings that do not. We wanted to see how complexity affects the clarity of perceived cause and effect relationships.

Would cause-effect relationships remain clear if they were inconsistently interrupted? We could start with a few presuppositions. Synchrony conveys a gestaltist principle called common fate, which manifests when we group simultaneously changing elements. We expect that when an object moves, all its parts move together; thus we bind sound and image if they change at the same time. In addition to common fate, consistent synchrony conveys gestalts of good

continuation: it fulfils the expectation that when something begins to move in certain direction, it will continue to move in that direction. Furthermore, people are generally familiar with audio-visual software and VJ culture, which means that consistent synchrony produces an impression of clearly perceivable cause-effect relationships. We wanted to see if these relationships remained clear once synchrony was inconsistently interrupted; the interruptions should be clearly noticeable.

Would random latency produce a sense of cause and effect relationship? We decided to see how latency, i.e. the delay between the audio and visual stimulus, affects perceived causation. And what if we added more perturbation? Using Schutz' and Kubovy's (2009) marimba experiment as a guide, we compounded perturbations by 1) randomizing latency, 2) randomly interrupting the cause-effect relationship, and 3) adding the perturbation of a synchronized, not interpolated visual parameter.

Given synchrony conveys a sense of causation, would the feeling persist when complexity obfuscated the base cause-effect relationships? That should manifest as a quantifiable gap between the participants' sense of causation and their sense of understanding the cause and effect relationships. To grasp underlying perceptual dynamics, we decided to ask for a description of the perceived cause and effect relationships. We could analyse the qualitative answers by considering gestaltist principles (Wertheimer 1938; Bregman 1990; Snyder 2001) and the notion of audiovisual object as mind-dependent concept (Schutz and Kubovy, 2009).

4.2 STIMULI

The study employed four audio-visual mapping prototypes, programmed in Processing Java-based procedural graphics environment (<http://processing.org/>) and shown on a computer.

The same audio recording was used in all prototypes: a short orchestration of string instruments (37 seconds), with amplitude ranging between 0 and 43 (arbitrary values). We dispensed with computer-generated sounds, which would potentially fit with computer-generated images. In this way, we ensured that perceptual binding was due to the mapping,

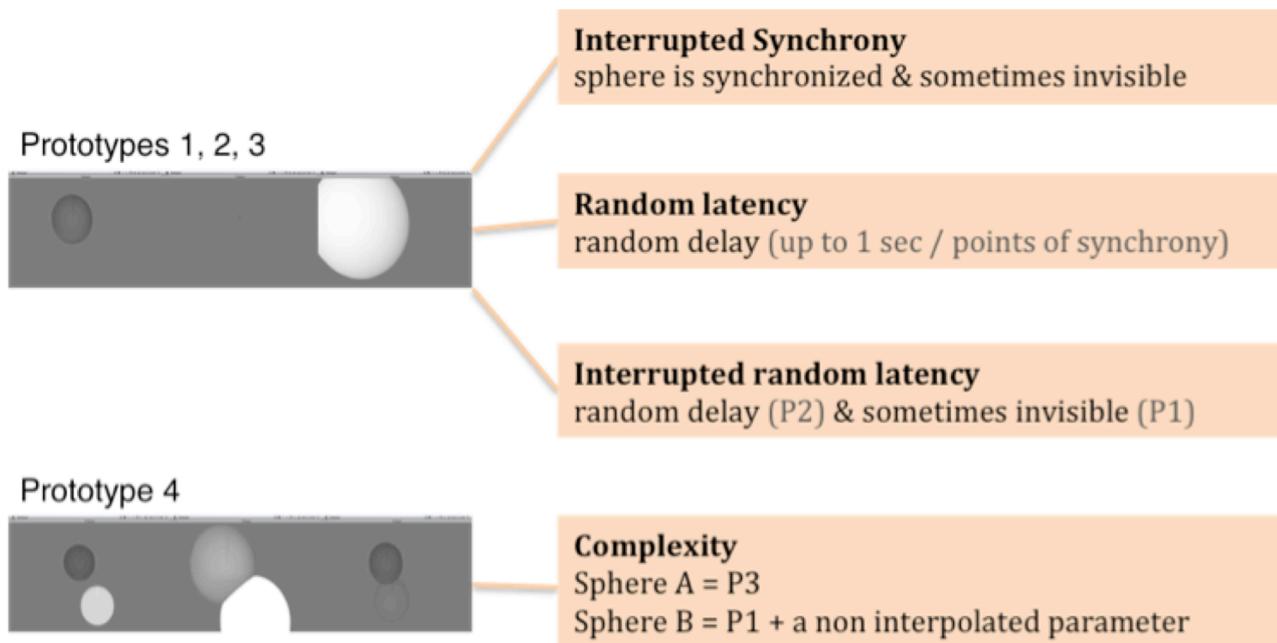


Figure 1 | Prototypes 1, 2 and 3, which exhibit one sphere; and Prototype 4, which exhibits two spheres (<http://doc.gold.ac.uk/~map01apv/StudyMappingPrototypes.mp4>).

independently from the specific qualities of the sounds and the images.

The prototypes were black and white. They exhibited a sphere (two, in Prototype 4) drawn in a digital 3D space (Figure 1). Audio amplitude was mapped to the spheres' size, colour/ transparency, and position. All parameters except position in Prototype 4 were slightly interpolated, which smoothed otherwise frantic visual changes.

4.2.1 Prototype 1: Interrupted synchrony

In Prototype 1 the sphere is synchronized with amplitude detection of the audio stimulus. It is invisible between amplitude 7 and 18 (this interval is in the range of average amplitude values).

4.2.2 Prototype 2: Random latency

In Prototype 2 the sphere is drawn with random delay upon amplitude detection. There are occasional points of synchrony, and maximum delay is 1 sec (automatic multisensory interactions due to plausible cause-effect relationships may occur even when the effect is delayed up to 700msec (Kubovy and Schutz, 2010)).

4.2.3 Prototype 3: Interrupted random latency

In Prototype 3 the sphere is drawn with random delay upon amplitude detection, as in Prototype 2. In addition it is invisible between amplitude 7 and 18.

4.2.4 Prototype 4 / Complexity

Prototype 4 displays two spheres that merge and split. Sphere A is drawn with random delay upon amplitude detection, as in Prototype 3. Sphere B is synchronized and invisible between amplitude 7 and 18, as in Prototype 1. Because the position parameter is not interpolated in sphere B, the sphere moves frantically through the X and the Z-axis in the digital 3D space.

4.3 PROCEDURE

The participants were recruited from Goldsmiths, University of London. All had knowledge of computing. Thus, if they did not understand a mapping (low Transparency rate) and yet felt causation (high Causation rate), it could not be due to being unfamiliar with software. Importantly, at the time nobody was acquainted with our investigations about cognition/attention; they were published later.

The experiment included ten individual sessions. Firstly, the participants were asked to read a questionnaire; they should respond after viewing each of four audio-visual mapping prototypes. Then they

were played the audio recording alone, prior to viewing the mapping prototypes, in random order. The questionnaire asked them to rate their sense of causation (Causation) and their sense of understanding the cause-effect relationships (Transparency), on a Likert scale (between 1 and 7). Additionally, it asked them to explain their Transparency rates.

The questions were posed as follows, except headers:

Causation / quantitative question

A: How would you scale the sense of a cause-effect relation between sound and image?

(1 = no cause-effect relation between sound and image; 7 = all visual changes relate to changes in sound)

Transparency / quantitative question

B: Can you distinguish which input factors affect which output parameters, and how?

(1 = never; 7 = always)

Explanation / qualitative question

C: Please explain the latter rating.

4.4 RESULTS

The average ratings for each mapping prototype were calculated, generated statistics, compared, and further related with the answers to the qualitative question.

Table 1 shows the average over the participants' ratings for Causation (question A) and Transparency (question B).

We used a statistics test (T-test) to compare the ratings. It showed that:

- The subjective rating for Causation is significantly higher in Prototypes 1 and 4 ($p < 0.05$) than it is in Prototypes 2 and 3 ($p < 0.05$).
- The subjective rating for Transparency is significantly higher in Prototype 1 than it is in Prototypes 2 ($p < 0.05$), 3 ($p < 0.05$) and 4 ($p < 0.05$).

Table 1 | Average ratings for Causation and Transparency.

PROTOTYPE	A: causation	B: transparency
1	5.6	5.1
2	2.7	2.3
3	1.8	1.9
4	4.7	2.8

- The gap between the subjective ratings for Causation and Transparency is not significant in Prototypes 1, 2 and 3 ($p > 0.05$).

- The gap between the subjective ratings for Causation and Transparency is significant in Prototype 4 ($p < 0.05$).

The statistic test establishes a distinction between "high" or "low" rates. Furthermore, the aspect of the prototypes can be inferred from the answers to question C, which aimed the participants to explain their Transparency rates.

Both Causation and Transparency were rated high in Prototype 1 – Interrupted Synchrony. Six participants wrote of one changing visual shape, in the singular. For example, a person who rated Causation 7 and Transparency 7 wrote: "*sound affects the sphere's rhythm and size*" [Participant #7]. The other four participants spoke of two visual shapes. For example, a participant who rated Causation 7 and Transparency 6 wrote: "*the higher the frequency the bigger the circle; each touch on the zither makes a small circle*" [Participant #10].

In Prototype 2 – Random Delay, the global rating was low for both Causation and Transparency. Four participants wrote that they did not find any audio-visual relationship. Other participants did. For example, a person who rated Causation 2 and Transparency 2 wrote: "*I have the impression that amplitude affects the shape, but not all consistently*" [Participant #8]. Another person who rated Causation 3 and Transparency 2 wrote: "*amplitude affects nothing (or little); it seems the guitar is affecting, but it proved inconsistent; some delay effect*" [Participant #9].

Similarly, the global ratings were low in Prototype 3 – Interrupted Random Delay. Six participants did not

find any audio-visual relationship. For example, a person who rated Causation 1 and Transparency 1 wrote “no relation”. [Participant #4] The other four persons did find a relationship. For example, a person who rated C2 and T2 wrote: “at times when the sphere is small it synchs to sound, at other times it appears during pauses between musical phrases” [Participant #1].

In Prototype 4 – Complexity, causation was rated high and transparency was rated low. Several participants stated that they found a relation between sonic and visual events, and that this relation was confusing. For example, a participant who rated Causation 7 and Transparency 1 wrote “no idea” when asked to explain the transparency rate [Participant #5]. Three participants did not distinguish any specific cause-effect relationship, and the others assigned multiple visual parameters to multiple sonic parameters. Among these, four specified that they were uncertain. For example, a person who rated C6 and T3 wrote: “it is hard to concrete one input with one output. There are many possible factors, but some seemed to have gotten the sound very closely” [Participant #3].

Prototype Complexity includes the mapping from prototype Interrupted Random Latency. Thus we compared the data from these two prototypes. After viewing Interrupted Random Latency, a participant who rated both Causation and Transparency 1 explained: “sound and image do not synch”. Yet after viewing Complexity, the same person rated Causation 7 and Transparency 5, explaining: “rhythm and tempo affect the sphere; all visual changes are produced by sound” [Participant #7]. This means that the person saw a relation between the sounds and the visual shape from Interrupted Synchrony; yet when viewed independently, this shape had been assessed to exhibit no relationship with sound.

That person saw Interrupted Random Latency prior to Complexity, and one could ask if the change was due to the order in which the prototypes were shown. Yet, we found a similar change with another participant, who saw Interrupted Random Latency after Complexity [Participant #4]. In Interrupted Random Latency, the person rated Causation 2 and Transparency 1: “I can't identify anything”. In Complexity, he rated Causation 4 and Transparency 3: “the instrument type effects position; amplitude

effects size; some delay, too?” Whereas in Interrupted Random Latency he detected no cause-effect relation, in Complexity he detected mechanical delay.

4.5 ANALYSIS AND DISCUSSION

Prototype 1 – Interrupted Synchrony aimed at confirming that audio-visual synchrony conveys causation, and at seeing whether the cause and effect relationships remained clear despite being inconsistently interrupted. Prototype 2 – Random Latency tested whether random delay produces a sense of audio-visual relationship. Prototype 3 – Interrupted Random Delay was meant to test the same; except it compounded inconsistencies, by randomly interrupting the cause-effect relationship. Finally, Prototype 4 – Complexity aimed at testing whether the sense of causation persists if one combines synchronised and non-synchronised audio-visual components, with a multitude of perturbations. Complexity combined the mappings of Interrupted Synchrony and Interrupted Random Delay, plus an additional mapping also exhibiting synchrony, but with lesser interpolation.

The participants found Causation high in two of the Prototypes: Interrupted Synchrony and Complexity. This indicates that *synchrony is taken to reveal a cause and effect relationship*. Conversely, the rating for Causation was low in both Random Delay and Interrupted Random Delay. In both prototypes, several participants perceived some cause and effect relationship, possibly due to the existing points of synchrony. Yet the low rating for Causation shows that perceptual binding was too weak to be convincing. In other words, *the sense of causation is weak when points of synchrony are sparse*.

In Interrupted Synchrony, the cause and effect was interrupted for a range of average amplitude values. The high rating for Transparency shows that cause and effect concepts were conclusive, despite the gaps when the sphere was invisible. This is well known in gestaltist psychology and neuroscience: inconsistency does not impede the formation of conclusive cause and effect concepts, or conclusive audiovisual objects – if we extend Kubovy and Schutz' terminology. Of course the perceived may not correspond to actual mappings, but that is irrelevant to our investigation.

After viewing Interrupted Synchrony, six participants spoke of a single sphere, in spite of the gaps when the sphere was invisible. They grouped the stimuli as a single audiovisual object. This manifests a gestaltist principle, the one of the closure. The principle describes how we “ignore” any intervals so as to infer a single object or form. An example in the visual domain is when we discern a circle in spite the line being interrupted (Wertheimer, 1938). In the auditory domain, an example is when we perceive a sound sequence – e.g. a sound repeating at equal intervals or a note scale – in spite of its actual incompleteness, e.g. a missing sound (Snyder, 2001). Closure happened with six participants, but not with the other four. Those spoke of a “small sphere” and a “big sphere”, assigning different cause-effect relationships to each. This indicates that they formed two distinct audiovisual objects. It shows that *conclusive cause and effect concepts admit inconsistency and multiple audiovisual objects*.

Combining the mappings from Interrupted Synchrony and Interrupted Random Delay, Complexity produced a global sense of causation, yet the cause-effect relationships were unclear. This shows that *sensing causation does not depend on perceiving how specific changes in the sound relate to specific changes in the image*.

The high rating for Causation means that perceptual

binding was convincing. Seven participants assigned multiple visual parameters to multiple sonic parameters, indicating that multiple audiovisual objects were formed at once. Since transparency was rated low, the audiovisual objects remained inconclusive. As a whole, the mapping of Complexity does not conform to pre-existing concepts of causation. While *the synchronised audio-visual components convey causation, the non-synchronised counteract conclusiveness*.

Interestingly, several participants assigned sonic parameters to the visual shape from Interrupted Random Latency. When viewed independently, this mapping had Causation rated low. Yet in Complexity, the participants did possibly not segregate the audio-visual components that produced a sense of causation, from the components that did not. Instead, *the participants sought for a global ecological fit*. Since Transparency was rated low, *they were aware of non-fitting information*.

In summary, *one perceives audio-visual relationships, but one feels unsure whether they correspond to actual mappings, or if they are purely perceptual*. Certain amount of audio-visual synchrony suffices to produce a global sense of causation. While we form concepts of causation, the aural discounts the visual and the visual discounts the aural. But if complexity generates confusion, that counteracts perceptual

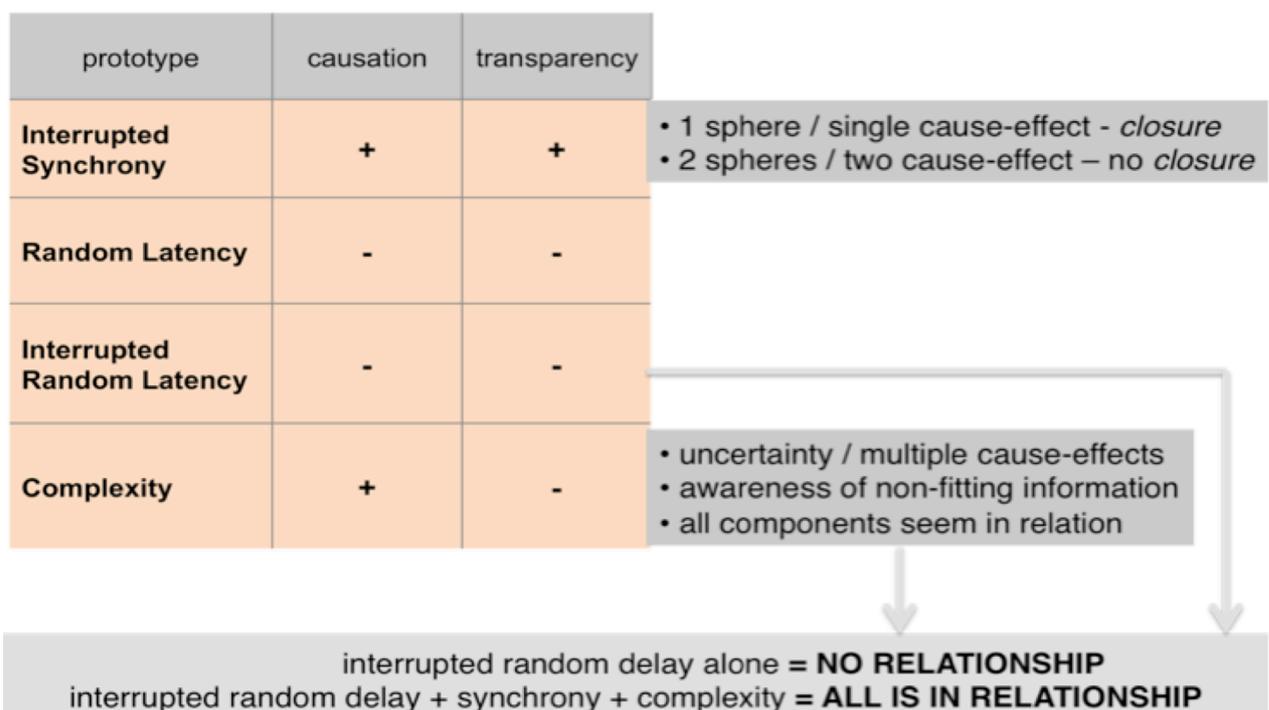


Figure 2 | Analysis of the study results.

simplification according to assumptions. Thus, perception may keep recovering sensory information that would have been automatically discarded if concepts were conclusive.

5 | CONCLUSION

The study showed that the fungible mapping includes components that convey a sense of causation and components that do not; and that the sense of causation persists when complexity confounds the actual cause and effect relationships. The study specified that: a) synchrony conveys causation even if it exhibits interruptions; one may form a single gestalt or separate gestalts, but the cause and effect relationships are conclusive; b) when sound and image are mapped with random latency or interrupted random latency, occasional points of synchrony do not suffice to produce a convincing sense of causation; and c) interruptions and diverging interpolations create complexity, confounding the actual cause and effect relationships.

Our analysis permits further conclusions. Synchrony reveals a cause and effect relationship, but not when points of synchrony are too sparse. Conclusive cause and effect concepts admit inconsistency and multiple audiovisual objects. With inconclusive cause-effect concepts, i.e. inconclusive audiovisual objects, one is aware of non-fitting information, yet perception keeps looking for a global fit.

The aspect of a fungible mapping was gleaned independently from personal creative explorations, so that it can be explored in many different ways and with any audio-visual platform. Synchrony conveys concepts of causation, and interrupted random delay does not. The point is, we do not tell them apart conclusively when the combination of mappings is complex.

We are driven to form conclusive concepts at the expense of overlooking or skewing any conflicting information. With a fungible audio-visual mapping, perception continues to acknowledge conflicting information, embracing convergences and divergences as inconclusive concepts.

Any attempt to describe perceptual dynamics in audio-visual performance will remain incomplete, but artistic motivations can be clarified with the aid of

science. Our perceptual approach to instrument design and composition frames the development of a personal audio-visual instrument, which explores three principles: audio-visual fungibility, visual continuity and sonic complexity. The instrument outputs acoustic sound, digital sound, and digital image. It includes a zither, that is an acoustic multi-string instrument with a fretboard, and 3D software which operates based on amplitude and pitch detection from the zither input. An early version is described in a special issue of Leonardo (Sa, 2013), and a later version in NIME 2014 proceedings (Sa, 2014). Further information and videos are available at:

<http://adrianasa.planetaclix.pt/research/AG1.htm>

<http://adrianasa.planetaclix.pt/research/AG2.htm>

ACKNOWLEDGEMENTS

This work was developed in research funded by the FCT - Foundation of Science and Technology and the European Union, through the operational program POPH/FSE (ref. SFRH/BD/68632/2010). We would like to thank the participants in the experiment, and Adam Parkinson who aided in the preparation of this manuscript. We are also thankful to CITAR and xCoAx, who invited us to write this extended version of the article published in the xCoAx Proceedings.

REFERENCES

- Bergson, Henri L. *Matter and Memory*. London: Allen, 1910.
- Bregman, Albert S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: MIT Press, 1990.
- Calvert, Gemma A., Charles Spence and Barry E. Stein, eds. *The Handbook of Multisensory Processes*. Cambridge: MIT Press, 2004.
- Chion, Michel. *Audio-Vision: Sound on Screen*. New York: Columbia University Press, 1994.
- Kant, Immanuel. *Kritik der Urteilskraft* (1790), ed. H.F. Klemme. Hamburg: Meiner, 2001.
- Knudsen, Eric I. "Fundamental Components of Attention." *Annual Review of Neuroscience* 30, no. 1. (2007): 57-78.

- Kobayashi, M., Y. Osada and M. Kashino. "The Effect of a Flashing Visual Stimulus on the Auditory Continuity Illusion." *Perception & Psychophysics* 69, no. 3 (2007): 393-399.
- Koffka, Kurt. *Principles of Gestalt Psychology*. London: Lund Humphries, 1935.
- Kubovy, Michael and Michael Schutz. "Audiovisual Objects." *Review of Philosophy and Psychology* 1 (2010):41-46.
- McGurk, Harry and John MacDonald. "Hearing Lips and Seeing Voices." *Nature* 264 (1976): 746–748.
- Palmer, Steve. *Vision Science, Photons to Phenomenology*. Cambridge, MA: MIT Press, 1999.
- Pick, Herbert L., David H. Warren and John C. Hay. "Sensory Conflict in Judgments of Spatial Direction." *Perception and Psychophysics* 6, no. 4 (1969): 203–205.
- Pressing, Jeff. "Some Perspectives on Performed Sound and Music in Virtual Environments." *Presence* 6, no. 4 (1997): 482-503.
- Rubin, Edgar. *Visuell Wahrgenommene Figuren: Studien in Psychologischer Analyse*. Copenhagen: Gyldendalske Boghandel, 1921.
- Sa, Adriana. "How an Audio-Visual Instrument Can Foster the Sonic Experience." *Live Visuals*, eds. L. Aceti, S. Gibson, S. M. Arisona, O. Sahin, Leonardo Almanac 19, No. 3, (January 2013): 284-305.
- Sa, Adriana. "Repurposing Video Game Software for Musical Expression: a perceptual approach." *Proceedings of NIME - Conference on New Interfaces for Musical Expression* (2014).
- Sa, Adriana. "Exploring Disparities Between Acoustic and Digital Outputs". In *Leonardo Transactions*, ed. A. McLean (2014)
- Schaeffer, Pierre. *Traité des Objets Musicaux: Essai Inter-Disciplines*. Paris: Éditions du Seuil, 1966
- Schutz, Michael and Michael Kubovy. "Causality and Cross-Modal Integration." *Journal of Experimental Psychology: Human Perception and Performance* 35, no. 6 (2009): 1791– 1810.
- Shams, L., Y. Kamitani and S. Shimojo. "Visual Illusion Induced by Sound." *Cognitive Brain Research* 14, no. 1 (2002): 147-152.
- Sinnett, Scott, Charles Spence and Salvador Soto-Faraco. "Visual Dominance and Attention: The Colavita Effect Revisited." *Attention, Perception & Psychophysics* 69, no. 5 (2007): 673-686
- Snyder, Bob. *Music and Memory: An Introduction*. Cambridge: MIT Press, 2001.
- Stevens, Meghan. *Music and Image in Concert*. Music and Media, 2009.
- Wertheimer, Max. "Laws of Organization in Perceptual Forms." *A Source Book of Gestalt Psychology*, ed. W. Ellis, 71-88. London: Routledge & Kegan Paul, 1938.

BIOGRAPHICAL INFORMATION

Adriana Sá is transdisciplinary artist, musician, performer/composer. Between 1995 and 2009 she worked fulltime on her artistic projects, involving many international artists, with grants and commissions. In 1997 she started using sensor technologies to explore music connected to light, movement, architecture and weather. Currently she explores disparities between human perception and digital analysis as creative material. She has been clarifying artistic insights with the aid of cognition/attention research, and developing 3D audio-visual software that operates based on pitch detection from a custom zither input. Adriana performed and exhibited worldwide, e.g. at Calouste Gulbenkian and Serralves (Portugal), Experimental Intermedia and PS1/MoMa (US), Caixa Forum and Arteleku (Spain), ICA–Institute of Contemporary Arts (UK) or Aomori Contemporary Art Center (Japan). Since 2011 she is a researcher in Arts and Computing Technologies at Goldsmiths, with international, peer-reviewed publications in *Leonardo*, *Leonardo Transactions*, *Live Interfaces Proceedings*, *xCoAx Proceedings* and *NIME Proceedings*. <http://adrianasa.planetaclix.pt/>

Baptiste Caramiaux is a researcher in the EAVI group, at Goldsmiths, University of London, since 2012. His work focuses on Human Motion, Interaction Design, Computational Systems and Sound Perception. He received a PhD in acoustic, signal processing and computer science applied to music from University Pierre et Marie Curie (UPMC), in December 2011,

completed at IRCAM – Centre Pompidou. Before joining IRCAM, he was trained in applied mathematics and music. His work has been published in international peer-reviewed journals (ACM Transactions, Computer Music Journal, etc.), conferences (NIME, CHI, SMC, etc.) and festivals (Transmediale, etc.).
<http://www.baptistecaramiaux.com>

Atau Tanaka creates musical instruments using sensing technology to capture movements and gestures of musicians. His first inspirations came

upon meeting John Cage during his Norton Lectures at Harvard. Atau then studied at CCRMA Stanford, and conducted research in Paris at IRCAM, Centre Pompidou. He has been artistic ambassador for Apple, was the first artist to be engaged as researcher at Sony Computer Science Laboratory (CSL), and has been mentor at NESTA UK, and Artistic Co-Director of STEIM Amsterdam. His work is funded by the European Research Council (ERC). He is Professor of Media Computing at Goldsmiths.
<http://eavi.goldsmithsdigital.com>