A STUDY ABOUT CONFOUNDING CAUSATION IN AUDIO-VISUAL MAPPING

ADRIANA SA

EAVI group, Computing Department, Goldsmiths
University of London
United Kingdom

adrianasa@goldw.ac.uk

BAPTISTE CARAMIEUX

EAVI group, Computing Department, Goldsmiths
University of London
United Kingdom

b.caramiaux@gold.ac.uk

ATAU TANAKA

EAVI group, Computing Department, Goldsmiths
University of London
United Kingdom

atau@goldsmithsdigital.com

Keywords: Audio-Visual Mapping, Perception, Causation

The text reports a study, which draws upon methods from experimental psychology to inform audio-visual instrument design. The study aims at gleaning how an audio-visual mapping can produce a sense of causation, and simultaneously confound the actual cause and effect relationships. We call this a *fungible audio-visual mapping*. The participants in this study are shown a few audio-visual mapping prototypes. We collect quantitative and qualitative data on 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. The sense of causation embraces the whole when the specific cause-effect concepts are inconclusive.

xCoAx 2014

2014.xCoAx.ORG PORTO, PORTUGAL COMPUTATION, COMUNICATION, AESTHETICS AND X

1. INTRODUCTION

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 1994:5) He noted that audition supports vision:

> *Why don't the myriad rapid visual movements in kung fu or special effect movies create a confusing impression? The answer is that they are "spotted" by rapid auditory punctuation. (Chion 1994:11)*

Musicians and audio-visual performers may find it problematic if perception biases sonic events that facilitate visual apprehension, subordinating timbres, textures, vibrations and nuances of musical expression, which form the wealth of multilayered relations between the sounds themselves. From Pierre Schaeffer (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 1997:8) In fact, 3D animators often place the sound of a footstep slightly before the foot actually hits the ground.

Our investigations are driven by an urge to understand whether, and how, the audio-visual relationship can foster a sense of causation without subordinating the music. As practitioners we desire the image to create a reactive stage scene without distracting the audience from sound organisation. Meghan Stevens proposed that the music remains dominant when the audio-visual relationship is *partially congruent*, but she stressed that her theories were "created from limited evidence." (Stevens 2009:3) Her notion of partial congruency is particularly subjective with abstracting sounds and images, because it relies on the sonic and visual shapes.

In a previous publication, we resorted to cognition/attention research in order to clarify how the audience's experience can be driven through the music – modulated, but not obfuscated, by a moving image. (Sa 2013) In many aspects vision tends to dominate over audition, but attention can be manipulated so that one sense dominates over the other. (Sinnet *et al* 2007) Attention increases the perceptual resolution of the information under focus, whether it is automatically driven to stimuli, or under individual control; furthermore, attention is drawn automatically to stimuli that are infrequent in time or in space. (Knudsen 2007) We concluded that to keep the music in the foreground one must dispense with disruptive visual changes, which automatically attract attention. There can be a wealth of discontinuities at a level of detail, but the image must enable perceptual simplification in order to provide a sense of overall continuity. One should apply gestaltist principles to visual dynamics; these psychological 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).

Clearly perceivable cause-effect relationships are as problematic for the music as disruptive visual changes. The gestaltist principles are an example of how we form conclusive concepts despite many inconsistencies. Indeed the primary aim of the brain is to be efficient in detecting, perceiving and responding to the world (Calvert *et al.* 2004). 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 explain that the binding of auditory and visual information is not merely associative: the visual discounts the aural and the aural discounts the visual. (Kubovy and Schutz 2010) The process is unconscious and presided by mind-dependent concepts, which they call *audiovisual objects*. Perceptual binding is undoable when the auditory and the visual stimuli appear unequivocally related; that is, when mind-dependent concepts are conclusive.

The question is 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 a publication titled “how an audio-visual instrument can foster the sonic experience,” (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.

The study here reported aims at demonstrating the fungible mapping independently from the other two principles, and regardless of any personal explorations or technical platforms. After watching each of several audio-visual mapping prototypes, the participants are questioned on their sense of causation and their sense of understanding the cause-effect relationships. We will analyze the quantitative and qualitative data by considering gestaltist principles and Kubovy & Schutz' notion of audiovisual object.

2. THE AUDIO-VISUAL RELATIONSHIP

2.1. TYPES OF PERCEPTION

Chion described three types of listening, or modes, which we can extend into the audio-visual domain. The first is *causal listening*, which "consists of listening to a sound in order to gather information about its cause (or source)" (Chion 1994:28). 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.

The second mode is *semantic listening*, which "refers to a code or a language to interpret a message" (Chion 1994:28). 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 1994:29). Chion provides perspective by stating that hiding the source of sounds "intensifies causal listening in taking away the aid of sight" (Chion 1994:32). In applying to the audio-visual domain, we consider how "reduced" might refer to stripping the perceptual experience of conclusive causes and meanings.

2.2. CONCLUSIVENESS AND INCONCLUSIVENESS IN AUDIO-VISUAL BINDING

Bob Snyder describes how relating new sensory information to previous experience enables perception to operate based on assumptions. (Snyder 2001) Events activate memories that have been previously activated by similar events. Among these memories, very few become highly activated and conscious; Snyder coins the term *semiactivated memories* to describe those memories which remain unconscious, forming expectations.

Expectations condition the process of multisensory synthesis. As we process 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* (McGurk 1976), in which non-matching speech sounds and lip movements are perceived as a new phoneme that diverges from both.

Schutz and Kubovy conducted a study about the perception of a (video-recorded) percussive action (Schutz and Kubovy 2009). 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. For example, the visual impact over a marimba fits naturally with a percussive sound, but not with a piano sound. Thus, when the sound and the image are slightly desynchronised (up to 700ms), the former combination leads to automatic interactions, and the later does not. The ecological fit depends on previous concepts, called audiovisual objects (Kubovy and Schutz 2010). The process of multisensory integration is undoable when audiovisual objects are conclusive: conscious awareness does then not "recover" discounted sensory information.

Chion 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 while knowing that they have a common origin, meaning a common cause: the film. People can draw upon implausible relationships in film, as it frames attention. We can say that binding sounds and images while finding the combination implausible is forming inconclusive audiovisual objects. And we can say that sensory interactions can be undone when audiovisual objects remain inconclusive.

So how do we form inconclusive audiovisual objects with abstracting sounds and images, which do not refer to anything but themselves? One-to-one synchronization would connote a common cause, but also make cause and effect relationships unequivocal. The challenge 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 the audience desists trying to understand the instrument, and focuses on the perceptual experience itself.

3. 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.

3.1. 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.

People are generally familiar with audio-visual software and VJ culture, which means that synchrony produces a high sense of causation, and consistent synchrony makes cause-effect relationships clearly perceivable. We wanted to see if these relationships remained clear once synchrony was occasionally interrupted.

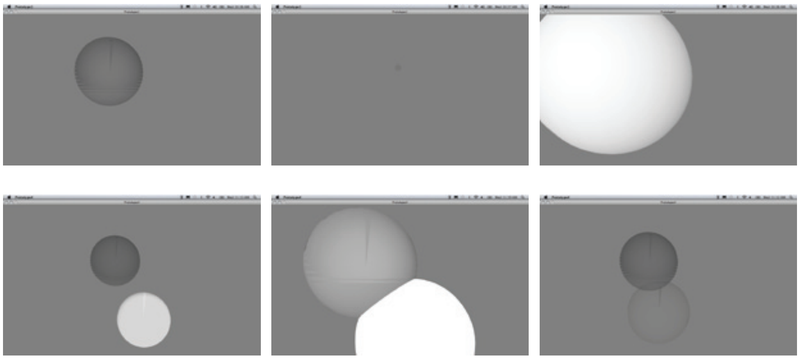
Additionally, we decided to see how latency, i.e. the delay between the audio and visual stimulus, affects perceived causation. Using Schutz' and Kubovy's marimba experiment as a guide (Schutz and Kubovy 2009), we compounded the effect of latency by 1.) randomizing latency, 2.) randomly interrupting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

Given that synchrony conveys a sense of causation, does the feeling persist when complexity obfuscates 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).

3.2. STIMULI

The study employed four audio-visual mapping prototypes, programmed in Processing Java-based procedural graphics environment,[1] and shown on a computer.

Fig. 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*)



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: we wanted to ensure that perceptual binding was due to the mapping itself, 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 smoothened otherwise frantic visual changes.

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).

Prototype 2 / Random latency

In Prototype 2 the sphere is drawn with random delay upon amplitude detection. There are occasional points of synchronization, and maximum delay is 1 sec (automatic

1 *http://processing.org/*

multisensory interactions due to plausible cause-effect relationships may occur even when the effect is delayed up to 700msec (Kubovy and Schutz 2010)).

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.

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.

3.3. PROCEDURE

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, nobody was acquainted with our investigations about cognition/ attention; they were exposed later, in an article which had not been published at the time (Sa 2013).

The experiment included ten individual sessions. Participants were previously asked to read through a questionnaire, which they would fill after viewing each prototype. Firstly they were played the audio recording alone. Subsequently, they were shown the four audio-visual mapping prototypes in random order. After viewing each prototype they were asked to respond to a same set of questions: two quantitative and one qualitative question. They were asked to rate Causation and Transparency on a Likert scale (between 1 and 7). Additionally, they were asked to explain their rates for Transparency.

The questions were formulated as follows:

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)
B. *Can you distinguish which input factors affect which output parameters, and how?*
(1 = never; 7 = always)
C. *Please explain the latter rating.*

The average ratings for each mapping prototype were calculated, generated statistics, and compared. Our analysis and discussion make use of the answers to the qualitative question (C).

3.4. RESULTS

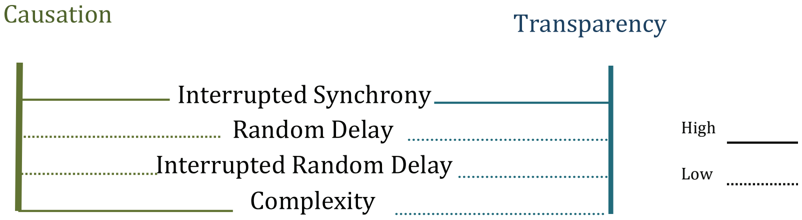
Table 1 shows the average over the participants' ratings for Causation (question A) and Transparency (question B).

According to the statistics test (T-test),

- The subjective rating of Causation is significantly higher with Prototypes 1 and 4 ($p < 0.05$) than with Prototypes 2 and 3 ($p < 0.05$).
- The subjective rating of Transparency is significantly higher with Prototype 1 than with Prototypes 2 ($p < 0.05$), 3 ($p < 0.05$) and 4 ($p < 0.05$).
- The subjective rating between Causation and Transparency is not significantly different with Prototypes 1, 2 and 3 ($p > 0.05$), but it differs significantly with Prototype 4 ($p < 0.05$).

Figure 2 represents these results.

Fig. 2 Relation between Causation and Transparency for each mapping prototype



Both Causation and Transparency were rated high with Prototype 1/ interrupted synchrony (continuous line). Both Causation and Transparency were rated low with Prototype 2/ random latency and Prototype 3/ interrupted random latency (dashed line). With Prototype 4/ complexity, Causation was rated high (continuous line) and Transparency was rated low (dashed line).

The aspect of Prototype 4 can be inferred from the answers to question C, which aimed the participants to explain their Transparency rates.

Several participants specified that they found a relation between sonic and visual events, and that this relation was confusing. For example, a participant who rated Causation with 7 and Transparency with 1 wrote "no idea" when asked to explain the latter rate. [Participant #5]

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

A participant rated both Causation and Transparency with 1 for Prototype 3, explaining: "*sound and image do not synch.*" The same person rated Causation 7 and Transparency 5 for Prototype 4, which included both synchronised and non-synchronised components: "*all visual changes are produced by sound*". [Participant #7]

A participant who rated C2 and T1 for P3, "*I can't identify anything*", rated C7 and T4 for P4: "*the instrument type effects position; amplitude effects size; some delay, too?*" [Participant #4] This shows that the delay between cause and effect led to different percepts in P3 and P4: whilst no cause-effect relation was detected in P3, the mechanical delay was detected in P4 (and this person viewed P4 prior to P3).

4. DISCUSSION

Prototype 1 aimed at confirming that audio-visual synchrony conveys causation, and that the cause and effect relationships seem clear even when those relationships are not absolutely consistent. Prototype 2 tested whether random delay disrupts the sense of causation, breaking any audio-visual relationship. Prototype 3 compounded this effect by randomly interrupting the cause-effect relationship. Finally, Prototype 4 tested how complexity effects perceived causation.

Participants found Causation high in Prototypes 1 and 4, which indicates that synchrony was taken to reveal a cause and effect relationship. The low ratings for Causation with Prototypes 2 and 3 indicate that the participants did not sense any cause and effect relationship when points of synchrony were sparse. Combining the mappings of P1 and P3, Prototype 4 produced a global sense of causation, and the cause-effect relationships were unclear. This shows that sensing causation does not depend on perceiving how specific changes in the sound may relate to specific changes in the image.

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. In the aural domain, we group simultaneously changing sounds (Bregman 1990). In a static visual image, where the movement is suggested by the relative orientation of the elements, we group elements that display a same orientation (Wertheimer 1938). Synchrony also conveys the principle of *good continuation*: it fulfils the expectation of a consistent audio-visual relationship. In the visual domain, the principle manifests when we group elements on basis of their following a consistent, lawful direction (Wertheimer 1938); or when we group visual objects which are arranged in straight lines or smooth curves. In the auditory domain, the principle manifests when we group consecutive sounds that change consistently, for example in pitch or loudness (Bregman 1990, Snyder 2001).

Regarding P1 (interrupted synchrony), the participants' high rates for Transparency indicate that the cause and effect concepts were conclusive. Six participants spoke of a single sphere, in spite of the gaps when the sphere was invisible. This is typical of the *closure* principle, one where we group a series of intervals to interpret the whole as a single form (Wertheimer 1938). In the auditory domain, closure manifests when we complete a linear sound sequence (e.g. a sound repeating at equal intervals, or a note scale) that lacks a sound (Snyder 2001). The six participants who spoke of a single sphere grouped the stimuli within a single cause and effect concept. The other four participants spoke of a "small sphere" and a "big sphere", assigning different cause-effect relationships to each; the stimuli for them comprised two distinct audiovisual objects. This shows that the sense of causation admits inconsistency and multiple audiovisual objects. It also shows that inconsistency does not impede the formation of conclusive cause and effect concepts.

The sense of causation continued in P4, where cause and effect relationships were further perturbated. The mapping used in P4 is one that does not conform to specific concepts. While synchronised audio-visual components convey causation, non-synchronised components counteract conclusiveness.

In P2 (random delay) and P3 (interrupted random delay), sparse synchrony points may have led the participants to momentarily bind sound and image. With P2, four participants did not find any audio-visual relationship, but six affirmed there was some sort of relationship; two of them actually mentioned latency. With P3, six participants did not find any audio-visual relationship, yet four assigned a part of the visual changes to sound. The low rates for both Causation and Transparency show that perceptual binding was too weak to be convincing.

Causation rates were high with P4, meaning that perceptual binding was convincing. Yet, three participants did not distinguish any specific cause-effect relationship. The other seven participants, among whom four stated uncertainty, assigned multiple visual parameters to multiple sonic parameters. This suggests that multiple audiovisual objects were formed at once. Since transparency was rated low, these audiovisual objects remained inconclusive.

Interestingly, several participants assigned a number of sonic parameters to a visual object (sphere A) which, when viewed independently, had been assessed to exhibit no relationship with the audio recording. Rather than segregating the audio-visual components that produced a sense of causation (sphere B) from the components that did not (sphere A), the participants sought for a global ecological fit. Since Transparency was rated low, they were aware of non-fitting information.

There was a sense of causation, given certain amount of audio-visual synchrony. As perception forms concepts of causation, the aural discounts the visual and the visual discounts the aural. But complexity generated confusion, counteracting mechanisms of perceptual simplification. Thus perception kept 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.

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 visual elements changing independently from sound do not. The point is, we do not tell them apart conclusively when multiple components changing independently from each other create complexity.

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.

6. COMPLEMENTARY NOTES

Art invites us to shift our usual ways of perceiving the world; there are many related philosophies. For example, Eastern philosophies teach us that one needs to suspend pragmatic thinking in order to permeate a relation between all things. And philosopher Henri Bergson wrote that the intellect shields the human mind from what he called *prime reality*, an evolving dynamic flux that proceeds along a definite but unpredictable course, where "all events, objects, and processes are unified" (Westcott 1968:8). He stated that intuition is a way to attain direct contact with this *prime reality* ordinarily masked from human knowledge, and that the intellect can freely interact with intuition. One may also 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).

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 at *http://adrianasa.planetaclix.pt/research/AG1.htm* and *http://adrianasa.planetaclix.pt/research/AG2.htm*.

ACKNOWLEDGEMENTS

This work was developed with a research grant awarded to Adriana Sa by FCT – Foundation of Science and Technology. 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 the anonymous reviewers, whose comments helped to improve the text.

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): pp. 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): pp. 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): pp. 41-46.
McGurk, Harry and John MacDonald. "Hearing Lips and Seeing Voices." *Nature* 264 (1976): pp. 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): pp. 203–205.
Pressing, Jeff. "Some Perspectives on Performed Sound and Music in Virtual Environments." *Presence* 6, nº. 4 (1997): pp. 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): pp. 284-305.
———. "Repurposing Video Game Software for Musical Expression: a perceptual approach." *Proceedings of NIME 2014 - Conference on New Interfaces for Musical Expression*. London (forthcoming).
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): pp. 1791-1810.
Shams, L., Y. Kamitani and S. Shimojo. "Visual Illusion Induced by Sound." *Cognitive Brain Research* 14, nº. 1 (2002): pp. 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): pp. 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, pp. 71-88. London: Routledge & Kegan Paul, 1938.