



# Same Time, Same Place, Keep it Simple, Repeat: Four Rules for Establishing Causality in Interactive Audio-Visual Performances

Yago de Quay

Department of Radio-Television-Film, University of Texas at Austin, USA

yagodequay@utexas.edu

## Abstract

Recent consumer-grade technologies that extract physiological biosignals from users are being introduced into interactive live performances and innovating its practice. However, the relationship between these signals and the responsive audiovisual content is often not understood by the audience. Recent discoveries in neuroscience can address this issue by proposing perceptual cues that help us connect the things we see and hear in our environment. Drawing from the field of neuroscience, and more specifically the theory of crossmodal binding, this paper proposes four rules that govern the mechanism which attributes causality between audiovisual elements: *same time, same place, keep it simple, repeat*. Intended as a set of guidelines for artists, they will help the audience unify, and understand the underlying cause to, what they see and hear in a performance. The last section describes a brainwave-based performance called *Ad Mortuos* that applies these four rules. A video of the performance is available at <http://www.tiny.cc/admortuos>.

**Keywords:** crossmodal binding, EEG based brain computer interface, audiovisual performance

## 1. Introduction

There is an increasing interest in developing interfaces that convert physiological, neurological and motor processes like heartbeat, muscle contraction, brain activity and body motion into digital media. Powered by the growing availability of consumer-grade medical devices, artistic experiments with biosignals are enriching the performative landscape.

However, these new interfaces tend to conceal the cause and consequence relationship between the performer's actions and the audiovisual output (Barbosa, 2006; d'Escriván, 2006; Hook et al., 2012; Jensenius, 2007). Although interactive transparency is up to the discretion of the artist, some argue that there is a need for approaches that make it clear and meaningful to the audience what and how audiovisual materials are being manipulated

(Armstrong, 2006; Schloss, 2003; Wechsler, 2006).

Two popular approaches have been the application of Embodied Cognition schemas and Human Computer Interaction principles (Hook et al., 2012; Leman, 2008). Together they provide suitable analogies that link the somatic qualities of a gesture or sound to the affordances of an interface (Larson & Johnson, 2011; Tanaka, 2010). These analogies are useful for user-centered design but they box our creativity within intuitive, generic models (Downie, 2005).

An alternative approach proposed by (Callear, 2012; Cooke, 2010; Coulter, 2010; Whitelaw, 2008) and pursued here, draws from neuroscience a means to leverage the neurological process that governs the unification of

audio and visual information into coherent, meaningful chunks. This process called cross-modal binding picks up spatiotemporal cues that indicate to the brain that an object and sound in our environment are related and have a causal relationship. When crossmodal binding is created synthetically with a live performer and interactive digital media, we are tricking our brain into creating these physical-virtual objects that share characteristics with the natural environment responsible for the evolution of crossmodal binding.

This paper argues that unrelated audio and visual elements in interactive live performance can be bound into a single percept by following four rules based on crossmodal binding: *same time, same place, keep it simple, repeat*. Designed with the audience in mind, these four rules promote audiovisual unity and clarify the cause and effect relationships between the performer and interactive audiovisual content. If followed, these rules will free the artists to work with abstract content and obscure interactive technologies and still have them perceptually and cognitively bound into a unified concept in the eyes of the spectator.

**Section 2** of this paper investigates literature on crossmodal binding supporting the four rules. **Section 3** prescribes eight simple design strategies that help establish causality and coherence between the live performers, visuals and sound. **Section 4** describes a performance called *Ad Mortuos* that was presented in the MOVE! concert series at UT Austin. The goal of the piece was to create a coherent expressive medium that combines thought, movement, sound and image using the four binding rules.

## 2. Crossmodal binding

A cat accidentally pushes a vase off a table and it smashes on the floor. We are able to rapidly orient our eyes to the scene, catch a glimpse of the culprit, and deduce what happened partially because we treated what we heard and saw as referring to a single spatiotemporal event (Holmes & Spence, 2005). The process of grouping vision and sound into a single percept, referred to here as crossmodal binding, is

a crucial, perceptual task that allows us to recognize objects in the real world that represent the common cause to the things we sense (Whitelaw, 2008). The way we combine sound and image depends on the causal relationship one infers through top-down decisions (the cat pushed the vase) (Körding et al., 2007), as well as automatic bottom-up combinations (the loud smash came from the vase) (Whitelaw, 2008). The ecological goal of crossmodal binding is to find a common cause of what we see and hear to an object or event in the environment.

Sounds and images that share similar qualities such as timing, localization, and semantic compatibility, are more likely to be treated as referring to a single multi-sensory event or object rather than separate unisensory events (Spence, 2007). These qualities are referred to as amodal properties since they are not specific to any particular sensory modality (Welch, 1999). Furthermore, belief in a causal relationship between auditory and visual information, also known as unity assumption, can also serve as a cue for binding (Schutz & Kubovy, 2009). Experiments with simple auditory and visual stimuli suggest that arbitrary sounds and images can be bound together simply through spatiotemporal coincidence (Effenberg, 2007; Körding et al., 2007; Schutz & Kubovy, 2009), and that the recognition of spatiotemporal coincidence is enhanced when the observer believes that the auditory and visual stimuli should go together (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Molholm, Ritter, Javitt, & Foxe, 2004).

### 2.1 Same time

When an auditory phenomenon and a visual phenomenon are presented at the same time an inevitable and irresistible weld is produced between them (Chion, Gorbman, & Murch, 1994). This weld allows seemingly unrelated pairs of audiovisual material to come together. For example, using temporal disparity as the independent variable, (Lewald & Guski, 2003) conducted three experiments that asked participants to judge the impression of the likelihood of a common cause, spatial alignment, and synchrony, of simple audio and

visual stimulus. In all three experiments temporal disparity had a significant negative effect on participants' judgment. Optimal results were found when the audio stimulus was presented 50-100 ms after the visual stimulus.

Consecutive stimuli tend to "pull" the perceived time of occurrence towards each other in an effect called "temporal ventriloquism" (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In their studies, Morein-Zamir, Soto-Faraco and Kingstone found that two consecutive visual events were perceived as happening in a shorter time frame when sound markers intervene between them compared to when the sounds were not present. A longer time frame was experienced when the sound markers happen before and after both visual events. The authors claim that the audio markers were attracting the perceived time of occurrence of each visual event towards them.

Stimuli with slow attacks make asynchronies harder to detect (Jaskowski, 1993). (Vatakis & Spence, 2006) found that subjects watching a reverse video clip of a guitarist playing classical music had difficulty in identifying the starting order of the asynchronous string sound and string pluck. The authors claim that the reverse video made it hard to detect when the string were being plucked. Asynchronies in the normal and reverse video clip were detected at around 30 ms and 150 ms, respectively, using just noticeable difference (JND) and a test for the perceived point of subjective simultaneity (PSS).

## 2.2 Same place

If sound and image are presented close to each other in space, observers tend to perceive a single underlying cause, if they are far apart from each other observers tend to infer two independent causes (Körding et al., 2007). Using as an example a ventriloquist and a puppet, if the voice and puppet appear together in the same place we ascribe one source to the auditory and visual stimuli. Increasing the distance between the voice and puppet makes us infer that each originates separately from an auditory source and visual source.

Spatial distance between the interaction of two visual object is negatively correlated with causality (Yela, 1952). Using the classic Michotte launching display experiment, Yela's study demonstrates that increasing the spatial gap between the stopping point of the first moving object and the starting point of the second moving object decreases the chances of causal impression, that is, we are less likely to believe that the first object caused the second one to move.

Spatially moving events with the same trajectory may produce stronger audiovisual binding than static events (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002) Furthermore, visual motion tends to influence the auditory motion but not vice versa (Körding et al., 2007). This sensory dominance asymmetry can be explained by modality precision hypothesis which predicts that the sensory channel that provides the most reliable information will dominate crossmodal integration, which in the case of localization is vision (Soto-Faraco, Spence, & Kingstone, 2004).

## 2.3 Keep it simple

"The austerity of the experimental situation may in-and-of-itself lead the participants into an 'assumption of unity' concerning the likely relation between the two signals" (Spence, 2007, p. 4). The presence of events that have been associated with each other strengthen audiovisual binding (Jack & Thurlow, 1973) while the presence of distractor events that do not share amodal properties weaken it (Welch, 1999). Even consciously, distractors are hard to ignore because the brain automatically detects discordance between stimuli (Spence, Ranson, & Driver, 2000).

Complex unimodal (either visual or auditory) stimuli can reduce crossmodal binding in favor of grouping elements within one sensor modality, a phenomena referred to as intramodal binding (Spence, 2007). Experiments by (Sanabria, Soto-Faraco, Chan, & Spence, 2005) with intramodal binding investigate whether the number of equally moving visual distractors has an influence on the perceived movement of an auditory stimulus (left-to-

right or vice versa). Results are counterintuitive. They report that increasing the number of visual distractors from two to six made it easier for participants to judge the trajectory of the auditory stimulus. Since previous research by (Morein-Zamir et al., 2003; Schutz & Lipscomb, 2007; Slutsky & Recanzone, 2001; Watanabe & Shimojo, 2001) found contamination of amodal properties between two sensory input as evidence for crossmodal binding, a reduction of contamination with six distractors suggests a reduced magnitude of crossmodal binding. Taken together, these studies show that simple audiovisual conditions enhance crossmodal grouping while complex audiovisual conditions segregate both senses and can encourage intramodal grouping.

Bimodal stimuli are registered faster and more accurately than unimodal stimuli (Teder-Sälejärvi, Russo, McDonald, & Hillyard, 2005). In their study, Teder-Sälejärvi et. al. presented subjects with randomized sequences of unimodal and bimodal light/sound stimuli and tested them for speeded responses to infrequency targets of higher intensity. When presented with only light or sound, subjects in Teder-Sälejärvi's study took longer and made less correct detections than when both light and sound were present.

## 2.4 Repeat

Repeated, simultaneous exposure to artificial audio and visual pairs increases their association as a congruent crossmodal object (van der Linden, van Turenout, & Fernandez, 2011) and produces integration-related cortical plasticity in the frontal lobe (Naumer et al., 2009).

Instructional and/or situational context favoring the assumption that two audio and visual stimuli should go together (unity assumption) can lead to changes in our perceptual mechanism that facilitate crossmodal binding (Welch, 1999). (Vroomen, Keetels, de Gelder, & Bertelson, 2004) noticed that when presented repeatedly with slightly asynchronous tone burst and flashing light (~± 200ms difference), observers tended to recalibrate their tolerance window towards the direction of the lag (up to 32 ms; ~10% of the lag). Aside from

demonstrating our brains' dynamic approach to spatiotemporal congruency, these studies suggest that the unity assumption gives strong impetus to the emergence of a crossmodal representation of an object.

Unimodal familiarization of an audio or visual stimulus increases identification speed and accuracy within and across modalities in congruent audiovisual pairs (Ballesteros, Gonzalez, Mayas, Garcia-Rodriguez, & Reales, 2009). In Ballesteros's study a group of participants was exposed to 30 ecological sounds or pictures and then told to identify a similar set of sounds or pictures in a speeded object naming implicit task. Results show that familiarization reduces identification time (1860 ms) and errors (2.51%) compared to no-study event (2339 ms and 3.26%). Furthermore, unisensory familiarization decreases response time in within-modal as well as crossmodal conditions, supporting the argument that crossmodal priming is not modality specific.

## 3. Design guidelines

This section presents a preliminary effort to explain how the principles *same time, same place, keep it simple, repeat* can be applied in the composition and design of an interactive audiovisual piece. They increase feature congruency (same time, same place), reduce anti-binding distractors (keep it simple), and suggest relatedness (repeat). These guidelines provide an approach to enhancing the binding of the various creative elements such as music, sound, scenery, performers, light and video. The goals are twofold: enable artists to bind abstract material, and help the audience understand the cause-and-effect between interactive elements in cases where the technology hides the mechanism of causality. The limitations are also twofold: artistic consideration supersede the need to establish causality, and the four principles govern only those elements that are implicated in a cause-and-effect relationship. Each rule breaks down into two guidelines that strengthen crossmodal binding.

*Same time:* Synchronous events and asynchronous events that have slow attacks;

*Same place:* Superposition of elements in the physical space and correlated spatial displacement;

*Keep it simple:* Absence of elements or events that are not related and audiovisual effects as opposed to uniquely auditory or visual;

*Repeat:* Repeated exposure to each cause-and-effect relationship and explicit information about the interactive relationships.

#### 4. Application of the rules in a performance

*Ad Mortuos* is a collaborative work that features live drawing, voice, brainwaves and dance. The theme was inspired by a poem that tells the story of the immortal soldier Athanatoi (Greek –“without death”) who sings on “bright love,” weaving a voice of hope through the first tableau. Then Athanatoi falls into deep séance, conjuring thoughts of the after-life and immortality. A video of the performance can be watched at the webpage <http://www.tiny.cc/admortuos>. Athanatoi is played by performer and vocalist Yago de Quay using a brain-computer interface (BCI). He wears a headset device (*Emotiv EPOC*, detecting EEG) that enables deliberate mental tasks and facial expressions to manipulate visual and musical parameters. The device connects wirelessly to the control software in a Windows 8.1 tablet that he wears on the chest, attached to a hoodie. A gyroscope sensor on the performer’s upper chest, attached to an under shirt, reports body orientation to the tablet as well. The control software is trained to recognize Quay’s facial expressions and mental commands. Together with the orientation data, these are sent by a datagram protocol (UDP) through a wireless router above the stage to the sound and visuals computers. Sound is produced using a Max/MSP patch (*Cycling ’74*) that layers voice and sound effects over a base track and then plays them through a stereo system. The visuals are manipulated in a VDMX software (*Vidvox*) and projected on the floor from overhead.

**Table 1** illustrates the mapping between the BCI performer and the digital audiovisual output. The interactive elements consist of the

BCI performer, six dancers, projections and sound effects. The BCI performer serves as the input while the six dancers, projections and sound effects as outputs. The three input types—body orientation, facial expressions and mental commands—demonstrate different degrees of perceived cause-and-effect, from obvious to invisible, respectively. Since mental commands are imperceptible to the audience, the performer uses his right hand as a visual cue to help connect the gesture to the size of the circles.

Table 1. Input and output mapping

Input	Digital Output
Body orientation	Slant of concentric circles Play position looped vocals
Facial Expressions	Distortion of circles Bleeps and noises
Mental Commands	Size of circles Right hand height

To help the audience understand the cause-and-effect mechanism in the BCI section of *Ad Mortuos*, the beginning strictly adhered to the four rules *same time*, *same place*, *keep it simple*, *repeat* before allowing independent, non-interactive events to occur such as solos by each one of the dancers. **Photos 1-4** on the next page illustrate how these principles impacted the overall visual aesthetic of the piece. The next paragraphs describe how the rules were implemented.

*Same Time:* Changes initiated by the BCI performer are immediately reflected on all interactive audiovisual elements. There are no long-term, delayed interactions. The performer’s facial expressions and body orientation are imitated by the dancers and produce an audiovisual effect. Mental commands on the other hand, take time to formulate. In order to keep them synchronous with the outputs we settled for slow, continuous audiovisual changes.



From top to bottom. Photo 1. *Same Time*. Body orientation rotating the visuals. Photo by Rodrigo Guedes. Photo 2. *Same Place*. Superimposed elements moving in the same direction. Photo by Rodrigo Guedes. Photo 3. *Keep it Simple*. Absence of non-interactive elements. Photo by Chian-ann Lu. Photo 4. *Repeat*. Repeated facial expressions deform the visuals. Photo by João Beira.

*Same Place*: All the interactive, visible elements are co-located. Body orientation, facial expressions, and hand gestures produce changes on the adjacent dancers and projections. Displacement of the different body parts and slant of the projections follow the same trajectory.

*Keep it Simple*: Aside from the lights and base track, which hardly change, all elements on stage are interactive. The three input types are never performed simultaneously and always produce a predictable effect on all the outputs (dancers, projection and visuals).

*Repeat*: It was at the discretion of the BCI performer when and how to perform the inputs. Nevertheless, they were repeated sequentially multiple times to help establish causality. The concert program informed the audience, without getting into much detail, that the BCI performer was controlling the projections and sound effects.

## 5. Conclusions

Introducing new interactive technologies in performance can foster innovative artistic practice. However, when technological mediation does not exhibit the physical causality properties present in the natural world that we are used to experiencing it runs the risk of obscuring the relationship between the performer's input and the media output. Furthermore, performances based on physiological biosignals leave the audience clueless to the nature of the input. This paper looks at empirical studies in crossmodal binding that are concerned with how our brains combine sound and image to suggest four rules *same time, same place, keep it simple, repeat* that can help re-establish the causal link between the performer, image and sound. The magnitude of synchronicity and proximity (same time, same place) of audio and visual events has a positive effect on their grouping. Audiovisual distractors and incongruity has a negative effect in grouping (keep it simple). Explanation about, and exposure to, audiovisual pairs enhances the observers' assumption that they should go together (repeat). Applicable to meditated performances, these four rules give hints to the audience's

brain that the performer, images and sounds are related and referring to an underlying cause. This paper ends by suggesting eight design guidelines and applies them to an interactive audiovisual piece entitled *Ad Mortuos* that uses brainwaves to control music and visuals.

### Credits for *Ad Mortuos*

Poet: Stephanie Pope

Spoken word: LaQuerra Carpenter

Choreographer: Yacov Sharir

Composer/Sound Design: Bruce Pennycook

Visuals: João Beira, Rodrigo Carvalho

Costume Design: Kelsey Vidic

BCI performer/Vocalist: Yago de Quay

Dancers: Emily Snouffer, Becca Bagley, Gianina Casale, Summer Fiaschetti, Katie McCarn and Allyson Morales

Camera: Pedro Miguel Resende, Filipa Rodrigues

### References

- Armstrong, N. (2006). *An Enactive Approach to Digital Musical Instrument Design*. (Ph.D.), Princeton University.
- Ballesteros, S., Gonzalez, M., Mayas, J., Garcia-Rodriguez, B., & Reales, J. M. (2009). Cross-modal repetition priming in young and old adults. *European Journal of Cognitive Psychology, 21*(2-3), 366-387.
- Barbosa, Á. M. (2006). *Displaced Soundscapes: computer supported cooperative work for music applications*. (Doctor per la Universitat Pompeu Fabra with Mention of European Doctor), Pompeu Fabra University, Barcelona. Retrieved from [http://www.abarbosa.org/docs/t\\_abarbosa.pdf](http://www.abarbosa.org/docs/t_abarbosa.pdf)
- Callear, S. (2012). *Audiovisual particles: parameter mapping as a framework for audiovisual composition*. (Ph.D.), Bath Spa University.
- Chion, M., Gorbman, C., & Murch, W. (1994). *Audio-vision: sound on screen*. New York: Columbia University Press.
- Cooke, G. (2010). Start making sense: Live audio-visual media performance. *International Journal of Performance Arts and Digital Media, 6*(2), 193-208.
- Coulter, J. (2010). Electroacoustic Music with Moving Images: the art of media pairing. *Organised Sound, 15*(1), 26-34.
- d'Esquiván, J. (2006). To sing the body electric: Instruments and effort in the performance of electronic music. *Contemporary Music Review, 25*(1-2), 183-191.
- Downie, M. (2005). *(Context) Choreographing the Extended Agent: performance graphics for dance theater*. (PhD), MIT, Boston. Retrieved from [http://openendedgroup.com/index.php/publication\\_s/thesis-downie/](http://openendedgroup.com/index.php/publication_s/thesis-downie/)
- Effenberg, A. O. (2007). *Movement Sonification: Motion perception, behavioral effects and functional data*. Paper presented at the 2nd International Workshop on Interactive Sonification, York, UK.
- Holmes, N. P., & Spence, C. (2005). Multisensory Integration: Space, Time and Superadditivity. *Current Biology, 15*(18), R762-R764.
- Hook, J., Schofield, G., Taylor, R., Bartindale, T., McCarthy, J., & Wright, P. (2012). *Exploring HCI's relationship with liveness*. Paper presented at the CHI '12 Extended Abstracts on Human Factors in Computing Systems, Austin, Texas, USA.
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept Mot Skills, 37*(3), 967-979.
- Jaskowski, P. (1993). Temporal-order judgment and reaction time to stimuli of different rise times. *Perception, 22*(8), 963-970.
- Jensenius, A. R. (2007). *Action-sound: developing methods and tools to study music-related body movement*. (Ph.D.), University of Oslo.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE, 2*(9), e943.
- Larson, S., & Johnson, M. (2011). *Something in the Way She Moves: The Metaphor of Musical Motion Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington, IN: Indiana University Press.
- Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158*(4), 405-414.
- Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*: The MIT Press.
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468-478.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: a High-density Electrical Mapping Study. *Cerebral Cortex, 14*(4), 452-465.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining

temporal ventriloquism. *Brain Res Cogn Brain Res*, 17(1), 154-163.

Naumer, M. J., Doehrmann, O., Muller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audio-visual object representations. *Cereb Cortex*, 19(7), 1641-1653.

Sanabria, D., Soto-Faraco, S., Chan, J., & Spence, C. (2005). Intramodal perceptual grouping modulates multisensory integration: evidence from the crossmodal dynamic capture task. *Neuroscience Letters*, 377(1), 59-64.

Schloss, W. A. (2003). Using Contemporary Technology in Live Performance: The Dilemma of the Performer. *Journal of New Music Research*, 32(3), 239-242.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Experimental Psychology: Human Perception and Performance*, 35(6), 1791-1810.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, 36(6), 888-897.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12(1), 7-10.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res Cogn Brain Res*, 14(1), 139-146.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-Modal Dynamic Capture: Congruency Effects in the Perception of Motion Across Sensory Modalities. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 330-345.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(2), 61-70.

Spence, C., Ranson, J., & Driver, J. (2000). Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Perception & Psychophysics*, 62(2), 410-424.

Tanaka, A. (2010). *Mapping Out Instruments, Affordances, and Mobiles*. Paper presented at the New Interfaces for Musical Expression, Sydney, Australia.

Teder-Sälejärvi, W. A., Russo, F. D., McDonald, J. J., & Hillyard, S. A. (2005). Effects of Spatial Congruity on Audio-Visual Multimodal Integration. *Journal of Cognitive Neuroscience*, 17(9), 1396-1409.

van der Linden, M., van Turenout, M., & Fernandez, G. (2011). Category Training Induces Cross-modal Object Representations in the Adult Human Brain. *Journal of Cognitive Neuroscience*, 23(6), 1315-1331.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, 1111(1), 134-142.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, 22(1), 32-35.

Watanabe, K., & Shimojo, S. (2001). When Sound Affects Vision: Effects of Auditory Grouping on Visual Motion Perception. *Psychological Science*, 12(2), 109-116.

Wechsler, R. (2006). Artistic Considerations in the Use of Motion Tracking with Live Performers: a Practical Guide. In S. Broadhurst & J. Machon (Eds.), *Performance and Technology: Practices of Virtual Embodiment and Interactivity*. Palgrave Macmillan.

Welch, R. B. (1999). Meaning, Attention, and the "Unity Assumption" in the Intersensory Bias of Spatial and Temporal Perceptions. In G. Aschersleben, T. Bakhman, & J. Müsseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events*. Amsterdam: Elsevier.

Whitelaw, M. (2008). Synesthesia and Cross-Modality in Contemporary Audiovisuals. *Senses & Society*, 3(3).

Yela, M. (1952). Phenomenal causation at a distance. *Quarterly Journal of Experimental Psychology*, 4(4), 139-154.