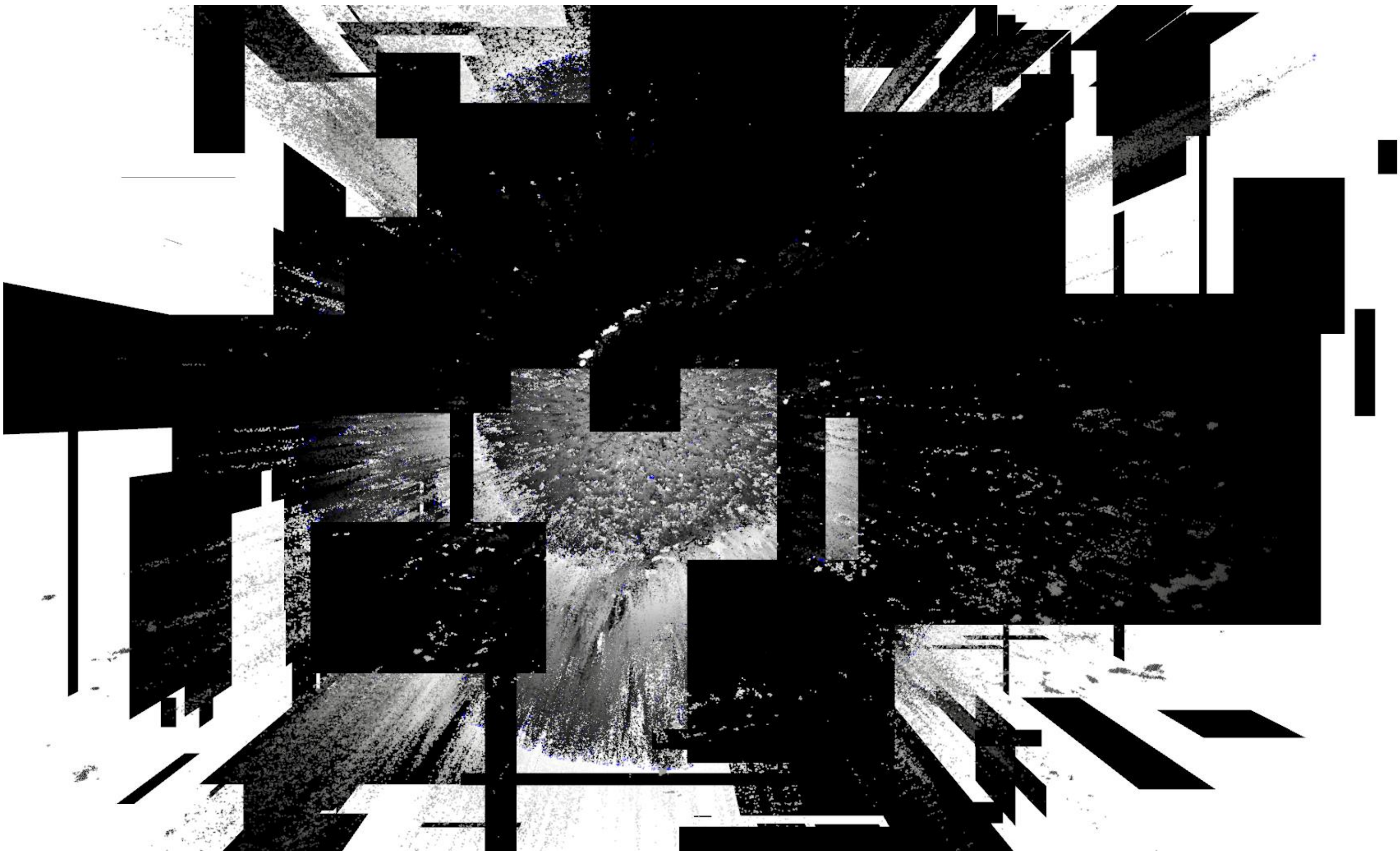
Cross – Modal Parameters and Approaches

Considering possibilities of cross-modal interactions in the audio-visual environment, with perception as point of departure



Bachelor thesis Miro Bollen Institute of Sonology, 2020

Abstract

In this paper I investigate **cross-modal interactions** between perceptual modalities, primarily between sound and image, and to a lesser extent tactile and vestibular involvement.
How these interactions can be manipulated in artistic practice is a central question in my research. In order to answer this question and as point of departure, texts relating to perceptual cross-modal phenomena are examined. Derived from this academic research and other personal experiences, cross-modal **parameters** and compositional **approaches** to these parameters are considered and experimented with. Finally, **interfaces** and their influence on the creation of the interactions and other audio-visual aspects are investigated.

Acknowledgements

I would like to thank:
all the teachers in the Institute of Sonology for years of inspiration and substantial insights,
Bjarni Gunnarsson, my supervisor, for encouraging me during this research,
my dear parents for reading and helping with the finishing touches of the paper.

Contents

Introduction

In this paper I investigate **cross-modal interactions** between perceptual modalities, primarily between sound and image, and to a lesser extent tactile and vestibular involvement. I understand cross-modal interactions to be interactions on a neurological level between signals from different senses before and during processing by the brain. How these interactions can be manipulated in artistic practice is a central question in my research. In order to answer this question and as point of departure, in the first chapter, texts relating to perceptual cross-modal phenomena are examined. Derived from this academic research and other personal experiences, cross-modal **parameters** and compositional **approaches** to these parameters are considered and experimented with in the second chapter. Finally, **interfaces** and their influence on the creation of the interactions and other audio-visual aspects are investigated. The research in this paper is both academic and practice based, which I hope is mutually stimulating.

Developing an audio-visual language which has the *interaction* between sound and image as an integral part of its vocabulary is one of my long-term goals. After having created mostly audio applications in code, I started experimenting with programming visuals which sparked my interest in image and, in a way, can be seen as another method of exploring the programming environment. Since then, I have attempted to combine code-generated material from both modalities into works in which I aim to give both a comparable amount of importance. Coming from a primarily electronic music background however, my knowledge within these modalities is not necessarily equally developed. After having created several audio-visual pieces, I became aware that when working with two modalities, there are three forms of expression: modality 1 (sound), modality 2 (image) and the relationship between them, and that this relationship need not be static. This paper investigates some possible methods of expression within these relationships.

Audio – Visual Unity

In psychology, the neurological process of looking for common properties across modalities when receiving multi-sensory impulses, is called *cross-modal verification* (Taberham 2018, 169-182). The common properties, are properties that do not belong to one specific sense, such as rhythm, smoothness or intensity (Taberham 2018, 158-159). In this paper, I will call them *multi-modal properties*. According to psychologist J.J Gibson (et al. 1969, 113–115) humans perform cross-modal verification with information from the different senses and when the patterns and rhythms match, we generally perceive the modalities as a single event and thus as unified. Throughout the text the term *audio-visual unity* will be used to describe the perception of sound and image as a single event.

It is this search for common properties which has appealed to many artists over the centuries and has been omnipresent in both commercial and non-commercial practices since the 1920's (Taberham 2018, 174). In figurative contexts, such as cartoons, use of this phenomenon is often termed 'mickey mousing' (Taberham 2018, 170-171). In more abstract works, however, it can also be found. While researching sources for this paper I occasionally stumbled upon a text which described historical techniques of correlating sound and image (Ribas, Luísa, 2014). Due to technical limitations, the correlation was not as much a given in the past as it is now. Nevertheless, many compelling multi-sensory artworks and creative devices were developed that attempted to unify the two media. The desire to correlate the two has continued in modern times where both advances in analogue and digital technology have spurred numerous works which explore rich fusions of sound and image. Two Japanese artists, who have been particularly influential on my personal work are Ryoichi Kurokawa and Ryoji Ikeda. Remarkably, both audio-visual artists have a background as composers in electronic music. Their approach to image appears to me to have a musical foundation, which, in combination with the audio-visual unity in their works, is a source of inspiration.

[CONTAIN] and *neuro.flux*

There are two audio-visual works which are mentioned throughout this paper and are related to my research. While the topics discussed in this paper are not necessarily the only points of focus in these works, they play an important role. Both compositions are based on similar abstract sound and image material and encompass influences from Ikeda and Kurokawa, drum and bass and science-fiction (books, films and video games). *[CONTAIN]* (2019) is a sequence of three short works in which I tried to achieve the perceptual unity mentioned above between sound and image. This work raised questions regarding the static relationship between the two media and encouraged me to think about strategies that could be implemented to challenge this unity. It is important for me, however, that the *interconnectedness*, or as Michel Chion (1994, 1-141), calls it 'the audiovisual contract', between the modalities remains intact. With the term *interconnectedness*, I am not necessarily suggesting that the media are perceived in unity, but that a large part of the common properties of the modalities are still in harmony. In *neuro.flux* (2020), I attempted to implement this interconnectedness and test the limits of unity by looking for the boundaries and working with them artistically. I did this by first unifying the modalities and then focusing on creating specific deviations in particular audio-visual properties and thus creating perceptual disturbances. Many of the parameters, approaches and interfaces explored in this paper are experimented with in *neuro.flux,* which can be seen as the practical component of this research. This work also serves as my final project for the Sonology Bachelor degree. In a way, for me, *[CONTAIN]* was a study for *neuro.flux*: in order to challenge unity, I first had to be able to create it. There are four specific fragments from *neuro.flux* which accompany this paper and are referred to.

Cross-Modal Perception

In this chapter, ideas, experiments and research, mostly derived from textual sources are discussed. These papers contain concepts which form the basis of my experimentation and primarily originate from branches within psychology and neuroscience. The focus is on multi-modal properties and the potential for these properties to generate cross-modal interactions. Possible applications of these characteristics as parameters is discussed in the second chapter.

Cross-Modal Phenomena

<u>"Synaesthesia"</u>

Cross-modal phenomena were already of interest to me before I started this research, and I have applied them to a number of previous works in greater or lesser degrees. Usually this has taken the form of flashes and pulses in fixed media, and stroboscopes in installations. Why flashes and pulses? Flashes temporarily overwhelm the senses and create a window in which the boundaries between modalities may become blurred. The pulse in sound may *become* the flash in image or the other way around. An example of this is when two auditory pulses occur in close chronological sequence and the first is accompanied by a bright flash; an illusionary second flash can sometimes be perceived together with the second pulse (Salter 2011, 205-211). What interests me is the kind of entanglement that occurs at this very moment between the modalities, an interaction which can bring about a kind of perceptual ambiguity which in its turn can create tension in the perceiver. This illusionary flash could be seen as a kind of temporary induced synaesthesia or 'cross-modal leakage' (Salter 2011, 205-211) – for lack of a better term. The modal phenomenon I am referring to is in fact not true synaesthesia, however, which is a rare neurological condition (Wikipedia 2020), but more like a universal form of synaesthesia that every brain may experience in varying degrees. To address this experience, a distinction can be made between **sensation** and **perception**. According to John A. Waterworth (1997, 327-330), sensation is what the sense organs register, and perception is the experience the brain has directly afterwards. What one experiences with synaesthesia is that a sensation of one sensory organ leads to a perception in another (Salter 2011, 208). This phenomenon can occasionally be experienced by non-synaesthetes under the right conditions, such as the illusionary second flash previously mentioned. Unfortunately, the conditions in which cross-modal leakage occur cannot easily be reproduced and thus are not reliable enough to be considered a compositional parameter.

The adjective "synaesthetic" is often loosely used to describe art in which two modalities are registered by the brain as one. As mentioned, I prefer to use the term *audio-visual unity* to describe this phenomenon, as "synaesthetic" contains misleading references to the neurological condition.

<u>Cross-modal Modulation</u>

Cross-modal modulation is a phenomenon which is an important focus for this paper and differs from synaesthesia (or cross-modal leakage) in that it implies exposure to two media, whereas the latter only concerns one, which "leaks" into the other. **Modulation** of one perception by another sensation when impulses from both modalities are being registered, is a reasonably frequent occurrence. One example of this perceptual phenomenon is what is called the *McGurk effect*, which happens when the visual information from a moving mouth changes the way we perceive speech coming from that mouth. In a video released by the BBC (2010), this modulation is demonstrated quite effectively. A mouth is shown which appears to produce the sound "Bah", however after a number of repetitions the movement of the mouth changes and appears to produce the sound "Fah". The change is also perceived in the audio even though the sound does not alter. This effect often occurs in real life when someone is getting poor audio information but a clear image (Wikipedia 2020). The McGurk effect specifically refers to speech, but the fact that visual information changes our aural perception could be imagined with non-semantic sounds and visuals. The reverse could then also be imagined where non-semantic sounds influence our vision. Indeed, an experiment by Berger & Ehrsson (2017) where participants are presented with two discs which overlap briefly in a diagonal movement is a good demonstration of this. Due to cross-modal modulation, these discs are perceived as either passing through each other or bouncing off each other depending on the sound (or silence) that is accompanying them. If a pulse is played synchronous to the intersecting discs, they seem to bounce away from each other, whereas if a different sound is played (or none), often the discs appear to cross. Furthermore, information given to the participant before the experiment also affected the results.

During the remainder of this chapter other cross-modal modulations within multi-modal properties will be discussed in which one modality modifies the perception of another. Some of these modulations could potentially be purposefully brought about to create a certain tension. How this may be used expressively will be considered in the second chapter. It may be interesting to note that interactions do not only happen between sensory modalities, but also between aspects *within* these modalities. The intensity of a tone for example affects the way we hear the pitch of that tone (Nuehoff 2004, 249-269), this is, however a subject that will not be covered in this paper.

Synchronicity

The following definition of the term synchronicity will be applied: 'the simultaneous occurrence of events which appear significantly related but have no discernible causal connection' (Lexico 2020). The perception of synchronicity between modalities is a widely discussed topic within psychology, neuroscience and the avant-garde cinema and many interesting experiments have been carried out.

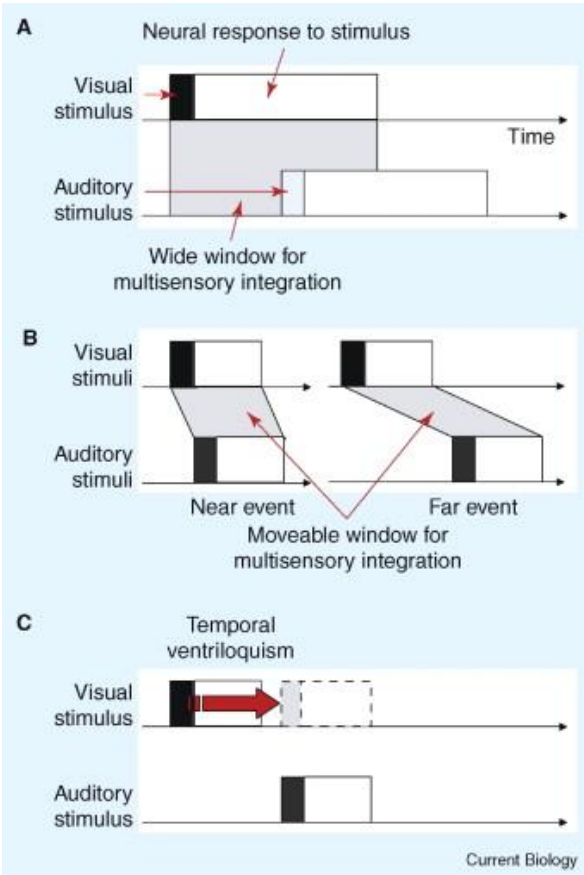
<u>Synchronous Multi-Sensory Impulses</u>

Chion (1994, 63-65) proposes the term 'synchresis', which he explains is the mental joining of sound and image when they happen at the same time. Chion gives examples of short visual and sound events, which, when combined, 'irresistibly' *weld* together. I would argue that this might often be the case with audio-visual events with a sharp attack and short duration (less than 5 seconds) but not necessarily with events which have a soft attack and a longer duration. Think of the image of a glass object being moved over a wooden surface, not *any* sound could be placed accompanying it to make the whole be perceived as one. A background murmur of voices for instance played simultaneously with the glass would not give the impression that the glass is causing the murmuring, even when started simultaneously. Other audio-visual properties also play an important role. There is, however, no doubt that humans have a strong tendency to link sound and image when they happen synchronously.

In an article by J. Neuhoff (2004, 249-269), specific brain cells that handle multi-sensory impulses are discussed. These cells do not respond significantly to either purely visual or purely auditory stimulation, only to the combination. This cellular activity could partly explain the tendency mentioned above to connect modalities. Neuhoff continues by presenting an experiment carried out by Wallace and Stein (1997) with kittens which shows that these brain cells were not present at birth and even at 12 days old the neurons did not respond in the same way as in adult cats. This suggests that the cells may develop with perceptual experience. In humans these cells might also be present which could imply that our perception of multi-sensory impulses is in part learnt over time (Neuhoff 2004, 249-269).

Window of Multisensory Integration

The time frame in which the perception of synchronicity occurs is the window in which multisensory integration happens. This time frame varies in duration, but is small: around several hundred milliseconds (Spence & Squire 2003, 519-521). In reality there is almost always a time difference between the arrival of auditory and visual stimuli from a particular event. There are two reasons for this; firstly, light travels faster than sound, which means that if an event happens far away, the sound is considerably delayed. Secondly, the chemical transduction of image by the eye is slower than the mechanical transduction of sound by the ear (Spence & Squire 2003, 519-521). This is the reason that over a short distance we can react faster to a sound than to an image (Ghuntla et al. 2014, 35-38). In general, if an event occurs at approximately 10 meters, the modal delays are equal and impulses from both modalities arrive at the same time. Most everyday events are, however, experienced closer or further than 10 meters and are still perceived as synchronous (Spence & Squire 2003, 519-521). If an event happens further away from the spectator, the window of multisensory integration shifts later in time and expands slightly: apparently this window can move in relation to the distance of an event and thus adapts to the additional time which it takes both modalities to arrive (Silva et al. 2013) (Spence & Squire 2003, 519-521). It is important to note that this window has different beginning and ending points in time for the auditory and visual stimuli. Due to the temporal inequality of the window between sound and image, where an audio-visual event occurs in which the image is slightly delayed in comparison with the sound, asynchrony will be perceived more often than when the sound is delayed (Spence & Squire 2003, 519-521). Apparently, delayed image is more intrusive than delayed sound. This makes sense, because in the physical world, when you see an event far away, the sound arrives later; it rarely occurs that the sound of an event reaches you before the image does.



Window for multisensory integration (Spence & Squire 2003, 519

Material Properties

'…vision is a useful modality for understanding the surface properties of an object, and audition is a useful modality for understanding its internal properties' (Waka et al. 2014). In this quote, the surface properties of an object could be seen as *texture*, and the internal properties as *material*. The situation that one can *feel* the material or texture in audio-visual stimuli just by looking and listening, is to many a familiar experience.

Although I found articles discussing possible tactile-image and tactile-sound interactions, I did not discover any texts relating to the audio-visual integration in texture perception (Klatzky & Lederman 2010, 210-227). Audio-visual integration in material perception is, however, discussed in a study by Waka Fujisaki (et al. 2014) in which it is demonstrated that a sound can alter the perception of the material properties of an image and generally has a stronger influence than a image of the object. A number of sources have written about this phenomenon in film, but most of these are focussed on either *vision* (Jennifer M. Barker's *The Tactile Eye* and Vivian Shoback's *The Address of the Eye*) or *audition* (Michel Chion's *Audio – Vision)*. Sound has often been used in cinema to emphasize or exaggerate material characteristics. A useful side-note is that recognising the material properties of an object involves complex sensory computation and that the modalities interact on a higher cognitive level during processing in the brain than some of the other phenomena mentioned in this chapter (Fujisaki et al. 2014).

Michel Chion (1994, 114-117) suggests the term *materializing sound indices* to describe degrees of materiality in sound perception in which the most important factor is the *definition* of the sound. According to Chion, the presence of more indices, thus more detail, can make a place or an object more tangible. 'The materializing indices are the sound's details that cause us to "feel" the material conditions of the sound source, and refer to the concrete process of the sound's production' (Chion 1994, 144). This is an interesting way to think about material perception, but I would suggest there are other sound properties which can change this perception. For instance, the level of the amplitude, the steepness of the attack, the length of the decay and the colour the timbre of a sound also affect our perception of material.

Embodied Properties

'Some musical gestures powerfully recall the human motor actions that produce them, revealing the tactile physicality of their source. Some musical materials directly encourage sensation and enact the body' (Mera 2016, 157-172).

The terminology of concepts surrounding bodily perceptions is often used interchangeably which makes it harder to define them. A distinction between three concepts could be made however, to start considering these properties as audio-visual parameters: the perception of touch, of movement of bodily parts, and of self-motion.

<u>Tactile Properties</u>

The term *tactile* refers to the feeling of touch, specifically to the feeling of pressure and vibration on skin (Sciencing 2017). This sense can be enacted with high volume infra-sound. In concerts, cinema and dance clubs, the low frequency vibrations generated by the speaker system often not only produce vibration in the air, but also vibrations in the ground which can be physically felt. This can be used to enhance physicality of an audio-visual experience but also in counterpoint to auditory and visual stimuli. The use of infra-sound is discussed further in the next chapter.

Proprioception

The term proprioception has varying definitions, I will use the following: **'...**the sense that lets us perceive the location, movement, and action of parts of the body' (Taylor 2008, 1143-1149). Although this term and the related experiments below refer specifically to the observation and audition of bodily movements and are thus not really applicable as a parameter in *neuro.flux*, this information may still be useful for future works.

In a study, R. Mukamel (et al. 2010, 750-756) discusses *mirror neurons*, neurons in the supplementary motor area of the brain which are believed to exist for the purpose of learning motoric actions. These neurons not only respond to *actions* by the owner of the brain, but also to *observations* of these same actions, carried out by the owner of another brain. Looking at certain (bodily) movements can apparently trigger some of the same neurons which are fired when these movements are performed and thus create a proprioceptive illusion. Another experiment suggests that a sound resulting from a certain action can also cause these neurons to fire (Kohler et al. 2002, 846–848). Both these experiments were carried out with monkeys, but similar brain activity has been found in humans.

Self-Motion Perception

The perception of self-motion in humans is determined by multiple modalities: as well as vision, the *vestibular system* is largely accountable for this awareness. 'The vestibular system is a sensory system that is responsible for providing our brain with information about motion, head position, and spatial orientation' (Neuroscientifically Challenged 2015). It is clear from an article by Dichgans and Brandt (1978, 755-804) that visual stimuli can interact with impulses from the vestibular system to create the illusion of self-motion. Dichgans and Brandt give the example of seeing clouds move or water stream, but perhaps sounds can also induce this feeling of movement. For instance, by simulating the sound of air rushing through the ears. Audio-visual *virtual* experiences may induce an illusion of self-motion. Rollercoasters in virtual reality for instance can, in my experience, give the feeling of violent movement very effectively. This sense of speed can be produced by a combination of a rapid movement in the virtual camera and the sound of rushing air. The use of these multi-modal movements is discussed in more detail later.

Spatial characteristics

'...the usual distinction that vision gives us objects and audition gives us events is a trap. It misleads us into thinking about vision as a spatial sense and about audition as a temporal sense.' (Handel 2006, 5).

The representation of external space, like temporal, material, and embodied properties, is built by multiple senses in our brain: spatial information is collected by the modalities and if one is missing the other completes the "picture" (Neuhoff 2004, 249-264). Audition for instance can "reveal" visually hidden objects and vision can "expose" static or silent objects. A difference in perception of space between seeing and hearing which should be noted, is that the eyes are directional while the ears are omni-directional (with a frontal bias). Three concepts which may be relevant to the audio-visual perception and representation of space are **position**, **depth** and **movement** in Euclidian space.

Position

In static situations or situations with gradual changes, the visual representation appears to consistently determine the perception of the location of an object more than the auditory (Knudsen & Brainard 1995, 19-43). Neuhoff (2004, 249-264) gives the example of a common situation in movie theatres: the location of the voice often does not exactly match the location of the lips on the screen, but moviegoers rarely perceive this discrepancy and would locate the sound at the actor's mouths. This modulated representation of location by vision is called the "ventriloquism effect" (Oxford Reference, 2020) and only happens if the changes in modal information are within certain boundaries. Another experiment relating to the perception of position done by G.H. Fisher (1964, 2-14) which included blindfolded sighted people and blind people, showed that when blindfolded, naturally sighted people are better at localising than blind people. This implies that the modality of audition is attuned to be more precise in perceiving positions by having had visual experience and would support the earlier statement that multi-sensory impulses are partly learned over time (Neuhoff 2004, 249-264).

Depth

'Depth perception is the visual ability to perceive the world in three dimensions (3D) and the distance of an object' (Wikipedia 2020).

As well as initial temporal delay, perceiving the distance of an object in audition is also based on volume and possible reflections from surfaces around that object. Optical depth perception is a complicated process and can be subdivided into two different kinds of information: information which is derived from a single eye, such as absolute size or perspective, and information which is derived from a combination of both eyes, such as stereopsis, which is the judgment of distance by comparing the two different images from the eyes (Wikipedia 2020). The volumetric size of a space could also be considered to be represented by perceptions of distance, namely the distance to the walls, ceiling and floor or the absence of them. The auditory and visual information which gives volumetric size (and thus distance) has been shown to interact in a study by Hans-Joachim Maempel and Matthias Jentsch (2013) in which they concluded that when the acoustic room size does not match the optical room size, the participant depends more on vision than on audition. Again, the representation of space seems to be biased towards visual impulses.

Temporal shifts in the perception of synchrony in relation to the distance of an object have been previously mentioned in the subchapter *Window of Multi-Sensory Integration*. It appears that the distance of an object influences the perception of synchronicity. A question which could be asked is whether it is possible to make an object appear further away by delaying the sound. Jaekl, Soto-Faraco and Harris (2012, 457-462) answer this question in an experiment which shows that 'audio-visual asynchronies can produce a shift in the apparent size of an object and attribute this shift to a change in perceived distance'. Considering the possibilities for this as an audio-visual parameter will be discussed in the second chapter.

Movement

In the movement of an object it also appears (surprise) that visual stimuli have the upper hand. Perception of auditory motion is modulated by the motion of visual stimuli and when the two modalities give opposing motion information, the visual element dominates the auditory (Sato-Faraco et al. 2002, 139-146). Under certain circumstances the location of sound can, however, change the perception of the location of a static (blinking) visual element. (Hidaka et al 2009).

Repetitive exposure

During the creation of an audio-visual fixed media piece, it often occurs that the maker is exposed to the same small section (or sequence) repeatedly which has an impact on how he/she experiences this section. In a more or less finalized version of *[CONTAIN]*, for example, I realized that slight discontinuities in time which I perceived in the concert setting, were not perceived by members of the audience. For them, slight (unintended) timing differences between image and sound were experienced as synchronous, because they were seeing and hearing the work for the first time. The repetitive exposure to this section apparently made my eyes and ears more sensitive to timing of that particular sequence of sound and image. The time frame in which synchronicity is perceived seemed to become shorter. In reality there is often a miniscule latency in either image or sound (or both). Unfortunately, there is not a way to experience that specific section for the first time again. The altered experience could, however, be acted upon during the making process, maybe by exaggerating any deviations between sound and image slightly.

Immersion

In this text I will use the term immersion to signify '… giving the viewer the strongest impression possible of being at the location where the sounds and images are' (Grau 2003, 14).

Since Classical Greek theatre there has been the tradition of a static audience presented with moving sound and image (Salter 2011, 200-222). This tradition continued in the West with artists such as Richard Wagner experimenting with lighting and staging to direct the spectators gaze towards the stage while also enveloping them with a sea of sound through acoustic techniques (Salter 2011, 200-222). Immersion is to this day an important factor in different media formats. Video games, cinema and dance clubs are all examples of situations where attempts are made to create immersive environments.

Presentation Situation

When working with perceptual entanglements, it is crucial that the circumstances in which a work is presented can immerse the spectator to a certain degree to allow the perception of cross-modal interactions. If a particular audio-visual sequence is experienced as credible, the interactions have a stronger impact. A dark, silent space with fairly neutral acoustic properties is therefore desirable. Intensifying the sound and image, with bright flashes for instance, can help to emphasize cross-modal phenomena by overwhelming the senses and thus reducing the perception of other stimuli present in the room. As previously stated, the infrasound of low tones gives a tactile experience and can be used to accent the physicality of a sound. The use of a multichannel speaker setup can also help to immerse the audience. If the goal is to give the spectator the impression of being in a credible (fictional) space, it makes sense to also project sounds from behind since in any real-life situation this is the case. Whether the rear speakers have an active function or if they just project sound which is derived from what is happening on screen is a point of consideration. Chion (1994, 85-86) makes the distinction between 'active' and 'passive' offscreen sound in cinema where 'active' refers to sounds that engage with the spectators anticipation (such as a character you do not see) and where 'passive' refers to sounds that stabilise the image (which can also be called atmospheric sounds).

Distracting Density

Another aspect that can be taken into account in the making process of an immersive audio-visual piece is the balance between the complexity of the two media. I will refer to complexity as changes in density of information over time of a particular modality. If one of the two media contains too many changes in information over a small amount of time, it might cognitively dominate, which may distract the viewer from clear cross-modal interactions between the media.

Parameters and Approaches

As mentioned in the introduction, in *neuro.flux* I am looking for expressiveness through interaction between sound and image. In this chapter, a number of cross-modal **parameters**, derived from the academic research in the previous chapter, will be considered. I will discuss whether these parameters are appropriate to adopt in a fixed media work in relation to *neuro.flux*. Some practical factors will also be reviewed, such as the range in which a certain parameter can vary (if applicable). The parameters are divided into four categories which refer to subchapters in the previous chapter. The four categories of parameters discussed are:

- (A-)synchronicity
- Material Properties
- Embodied Properties
- Spatial Characteristics

In *neuro.flux* and in this paper, the focus is particularly on the parameter of synchronicity, in future works, I hope to experiment with some of the other parameters examined in this paper. In order to play with the interaction between sound and image expressively, I mostly explore the musical aspect of **tension and release** by the manipulation of the parameters. In the Western music tradition, tension and release is based on the resolution of dissonance into consonance. In the audio-visual situation, one example of tension and release could relate to degrees of synchronicity, with *a-synchronicity* relating to tension and *synchronicity* to release.

After considering the parameters mentioned above, I elaborate on some compositional **approaches** used to experiment with these parameters in *neuro.flux*. These approaches are examples of techniques of manipulations of one modality in relation to the other and have their origins in cinema, avant-garde film and my own practical experimentation. *neuro.flux* consists of six major sections, in a number of these I have implemented an approach that I consider suitable for that particular section. The reasoning for these choices will also be discussed.

In order to develop my audio-visual language further, I realised that the perceptual unity reached in *[CONTAIN]* had to be challenged. Adopting (mostly) the same tools and similar source material, *neuro.flux* is made in a related audio-visual language to *[CONTAIN]*. Building on the experience acquired in unifying sound and image, I first attempted to link the audio-visual materials as convincingly as possible. Starting from perceptual unity allows for a focus on discreet parameters of interaction for both the creator and spectator without too much distraction from other cross-modal phenomena. After the merge of the modalities I tried to apply the approaches discussed in this chapter to "de-unify" the modalities.

Parameters

Ultimately, all parameters mentioned only occur over time; none are static, which results in the discussion of temporal organisation in multiple sub-chapters.

(A-)Synchronicity

The parameter of synchronicity is based on temporal alteration of one modality to the other, and has been a focus of mine in *neuro.flux*. Theoretically it ranges from an infinite negative (in relation to synchronicity) to an infinite positive deviation in time, although the range in which it could effectively be used is in fact very small. As discussed in the subchapter *Window of Multi-Sensory Integration* the size of the time frame in which synchronicity is perceived is about several hundred milliseconds and varies depending on audio-visual order or distance. If the difference in time between sound and image is outside of this time frame, the modalities are not perceived as synchronous and thus not as a single event. However, I am under the impression that if the other qualities of the event, other than time, still match, the brain may still connect the two on a higher cognitive level. An example is the time difference between the flash and the thunder of a lightning strike which often well exceeds this window. Alternating between unified perception and these higher cognitive connections can be an effective way to create expressive audio-visual interactions. In *neuro.flux*, I have primarily explored the boundaries of the window of multi-sensory integration to create these alternations. Not only the amount of temporal alteration is relevant when performing these temporal shifts: as stated earlier, the chronological order of modalities is also an important factor. Delaying the image, for instance, results in a more perceptible effect than delaying the sound with the same duration. To address degrees of tension in (a-)synchronicity these considerations should be kept in mind.

An important tool of expression in time, is rhythm. Rhythm can be perceived not only by both eyes and ears, but also through other modalities, such as touch. In fact, time can be seen as the only multi-modal property which is perceived by **every** modality and is thus in creating multi-modal works possibly the most powerful tool. Practically, I have only "rhythmized" the three modalities mentioned. If an iterative rhythm is present, expectations regarding synchronicity may be present and can be addressed. In *neuro.flux* there are two sections containing a steady pulse in which rhythm plays a prominent role. In these sections *rhymical counterpoint* has been implemented to address deviations in time, which will be discussed later.

'[...] In this sense, instead of saying a 'perception of rhythm', it would be better to speak of a 'rhythmization of perception" (Abraham, Rand and Torok 1995, 16).

Material Properties

I have been cautious to deviate material characteristics from unity as I am under the impression that changes made in these properties quickly disturb the audio-visual unified experience. In *[CONTAIN]*, I tried to merge the material qualities of the sounds to the textures and materials of the visuals. Falling clouds of grains were for instance accompanied by sweeps of filtered noise. I attempted likewise to merge the two in *neuro.flux*, in which it rarely leaves the safe confines of this unity. Only by the alteration of other parameters, such as time, were the material properties affected. In future pieces I hope to address shifts in material experience in more detail and I already have some ideas on how to proceed. Delicate discontinuities in material can be created by using filtering or distortion techniques in sound and changes in texture in image, while leaving the other properties intact. Filtering or distorting the sound could also been seen as altering Michel Chion's amount of *material indices*. The attack and release of a sound and motion of the visuals (for instance how two things collide) can also be used to alter the perception of a material.

Embodied Properties

Sub-frequencies can be used to enact tactile properties. These sub-frequencies generally occur in range of 0-200 Hz and if there is enough amplification in these frequencies, sounds can literally be felt. Contrary to the material interactions discussed above, I have applied this *physical* tactile experience to a large extent. The purpose of the sub frequencies in *neuro.flux* was more to create this tactile experience than to produce an audible tone: I tried to treat infra-sound as a separate modal layer, occasionally accentuating strong points of unity or creating physical tension in between them. I mostly used a single sub-frequency, namely around 40 Hz; with such a low frequency, the perception of pitch is much less accurate and thus it can be perceived to be the fundamental of many accompanying sounds in the mid-high range. As well as infrasound from low sinoid tones, the sounds that functioned as kickdrums in the rhymical parts of *neuro.flux* also played a tactile role. The degree of tactile involvement could be seen as a parameter.

Perceptual shifts in self-motion in the individual modalities may present possibilities for expressivity. Following an earlier train of thought: in image, a sense of movement and speed can be created by shifting camera angles, camera distance and by dynamically making objects approach or move away from the camera. With sounds that imitate fast movement, like air rushing through the ears or changes in volume, a similar sense can be evoked, though not as strongly as image can. The degree in which these virtual movements induce the feeling of self-motion could be seen as the degree in which the vestibular system interacts with visual or auditory impulses and could be considered a parameter. In *neuro.flux*, I employed this parameter in a number of instances. One example is near the beginning of the work (fragment 1).

Spatial Characteristics

The induced sense of self-motion discussed above, can be created by altering spatial characteristics such as camera distance. There are, however, a number of other spatial characteristics which can be played with expressively, such as the horizontal and vertical position and movement of an object in Euclidian space. In the specific fixed media format in which *neuro.flux* resides, these might not be the most potent of parameters. There are a couple of reasons for this. First of all, relying on the research mentioned earlier, the image tends to dominate the sound when the modal information on position and movement in space approximate each other but do not match. Modest changes in sound position therefore have little impact on the perception of location. Secondly, the screen on which *neuro.flux* will be shown is of limited size. Thus, visual spatial information can only be manipulated to a certain degree. When using stereo sound, the difference between the horizontal or vertical spatial information may not be large enough to create a noticeable effect. *neuro.flux* is stereo in essence and I have thus decided not to practically adopt the parameters of (horizontal and vertical) spatial position and movement at all. With the use of multichannel sound, however, a large discrepancy between sound and image could be realised and used expressively.

Differences in depth between sound and image provide more artistic possibilities to alter the audio-visual experience than spatial positions or movements and are applied a couple of times in *neuro.flux*. As mentioned in relation to the perception of synchronicity in time, when an happens far away, the sound can be delayed and still be perceived in unity with the image. The perception of a virtual distance could therefore be enlarged by delaying the sound in relationship to the image. A compelling example of this effect is in a scene in *Star Wars: Episode II – Attack of the Clones* (Lucas 2002). In this action-packed scene a space chase is depicted in which at one point an explosion occurs from a fictional 'seismic charge'. During the first second of the explosion a visual expansion and a subsequent contraction is illustrated, accompanied with silence, after which it expands again, this time accompanied by a wall of sound. The large delay between auditory and visual stimuli creates the perception of (large) distance and size. The fact that the time before the explosion is filled with sound accentuates the silence during the expansion and contraction. In *neuro.flux* I have tried to implement a similar effect on one occasion specifically (fragment 2).

As well as distance, the size of a space can also be considered as a parameter in the single screen fixed media format. In the visual domain this can be realised by changing the virtual volumetric size or camera distance, in the audio domain reverberation characteristics can be manipulated. During the time span of *neuro.flux* the size of the virtual space changes dynamically. The modalities, however, are more or less consistently linked and I have not consciously introduced discrepancies in volumetric space.

Approaches

Counterpoint

In this paper, I will use the following definition of counterpoint: 'Use of contrast or interplay of elements in a work of art' (Merriam-Webster. 2020).

Once it became possible to make a *fixed* soundtrack to image, some Russian film-makers extended their artistic ideas about *montage* to the sound domain. In the 1928 manifesto *A Statement*, S.M Eisenstein, V. I. Pudovkin and G. V. Alexandrov (Weis & Belton 1985, 83- 85) clearly warn for the temptation of using 'natural' sounds in film and synchronised speech deeming it the 'line of least resistance' and advocating the use of 'orchestral counterpoint' between visuals and sound. Unfortunately, although there are many beautiful examples in existence of works inspired by their thinking, their plea was largely not followed and the use of 'natural' sounds quickly became the standard in both commercial and artistic practises. In *A Statement* Eisenstein et al. assert that orchestral counterpoint is created by 'nonsynchronisation' and by treating the sound as an additional montage element, however, this is not really elaborated on. Pudovkin's text *Asynchronism as a Principle of Sound Film* (Weis & Belton 1985, 86-91) discusses this in more detail with references to examples of his own work. The contrapuntal relationship between sound and image explained in this text is based on emotional themes and on figurative representations. For me, this highlights a difference between some forms of cinema and my own audio-visual practise; even though there are strong similarities, cinema is often figurative and not necessarily intentionally musical. Pudovkin's counterpoint, for instance, could be seen as an interplay of *meaning*, whereas in *neuro.flux* the audio-visual counterpoint is based on specific audio-visual qualities and during making it I attempted to treat the parameters discussed above consistently as musical layers.

One way of achieving counterpoint, is by rhythmically "cloaking" the individual media. I use the term "cloaking" to refer to the absence of projected stimuli by a specific modality. An example of this technique is *Arnulf Rainer* (1960) by Peter Kubelka. Using only the alternation in sound between white noise and silence and in image between a black and a white screen, this piece demonstrates contrapuntal interactions between modalities effectively. There are more subtle approaches to counterpoint, for instance by rhythmically varying certain elements of each modality while other elements remain unified. In Ryoji Ikeda's work, a delicate counterpoint is often present between sound and image elements. At first glance, his work seems to be completely synchronised (even unified), however, with more careful attention to the audio-visual relationship, it becomes apparent that this relationship is not static. Certain sounds that initially appear without a visual counterpart for instance, may later be accompanied by a particular visual element, or a long visual event may appear to be triggered by a short sound. With the goal of retaining a certain credibility of the audio-visual relationship, I tried to implement this more "subtle" counterpoint in *neuro.flux*. It made sense to apply this particular technique in the rhythmical sections of the work (fragment 3). If a steady pulse is present in both modalities, certain expectations regarding the alternation between the two media can be played with, which is of special interest to me.

Points of Unity

Another approach to the audio-visual relationship related to counterpoint - from a different perspective - is using points of unity. This technique can bring about tension and release by using strategically placed points in time where sound and image meet in unity. In between these points, there can be room for cross-modal parameters to diverge and to converge again, colliding at the next point. The concept of points of unity is inspired from what Michel Chion (1994, 58) calls 'points of synchronisation' which he derived from cinema. 'A point of synchronization, or synch point, is a salient moment of an audiovisual sequence during which a sound event and a visual event meet in synchrony' (1994, 58).

These *synch points*, however, refer specifically to temporal qualities and are more like audio-visual accents in between which a certain 'temporal elasticity' (Chion 1994, 61-62) can be present. This elasticity allows for temporal deviation of the media before both collide again at the next point. As well as position (phase) in time, I would suggest points of unity also encapsulate other multi-modal properties. I will, however, focus on time deviation. Chion (1994, 61-62) points out that temporal elasticity often appears in scenes that have strong points of synchronisation and is not present in scenes that have a vague temporal quality. Fighting scenes (especially in animations), for instance, often have strong points of synchronisation; slow motion often occurs in between action points during which music and/or sound effects have a different sense of time than the image.

During the making process of *neuro.flux*, I have experimented with different time-scales in combination with different amounts of divergence in between the points of unity. To stay approximate to the borders of unity, however not strictly within, I used small scales. During this experimentation I found that with a relatively longer timescale, a smaller deviation is desired, and that with a shorter duration a larger deviation is possible while retaining the audio-visual relationship. The exact amount of deviation in between and beyond where the boundaries of unity reside is specific for each parameter. How this amount is measured for each parameter is an interesting question which needs further investigation. With the parameter of synchronicity, I have tried to explore exact measurements of time scale and (temporal) deviation. In *neuro.flux*, I did this experimentation with points of unity in a section where the image material consists of a series of sharp attacks followed by sinuous "ripples" (fragment 4). In the sound the sharp attacks are also present but are followed not by "ripples" but by slow decays. I thought this section would be suitable for implementing time differences, as the sharp attacks are strong temporal points which could act as the points of synchronisation and allow for a certain amount of temporal elasticity in between. The time frame in between these points varies from 9 to 12 seconds. Within these points of unity, I experimented freely with the velocity of the visual "ripples". In the sound, however, the decays remained more or less steady. I also experimented with applying a similar approach to other parameters than temporal change, such as material properties, but the results were, in my opinion, not compelling.

Transformations in Individual Modalities

Working on and structuring an individual modality within a fixed time frame or sequence which is already unified can bring an interesting discrepancy between sound and image. It can result in one modality deviating in timbre/time/space in a way that is natural/coherent to that specific modality. While making *neuro.flux*, I have mostly switched focus dynamically between the modalities, but occasionally I focussed on one modality for an extended temporal period. When making these uni-modal changes in a particular section, I tried to make sure I made alterations which only affected a single parameter. Attending to both modalities together, or focussing on only one, was a conscious consideration. If both are perceived, it may be more difficult to focus on modal specific changes, but if only one is perceived, alterations might be made which drastically alter the relationship. I tried to find a balance between the two in the later stages of the composition process. During the early stages, however, the sound was often transformed without observing the visuals, the opposite occurred significantly less frequently. This bias towards sound is the result of that in this work, time organisation is based largely on sound and not image. This makes sense, because, as stated before, there was an intention to treat the image in a musical way. Moreover, the environment in which *neuro.flux* was edited is designed for making music, as will become clear in the next chapter. An inequality in knowledge has also aided this bias of focus.

Interface Influences

‘Software mediates the composers' creative action of producing music but at the same time is itself moldable material for the composers’ (Bertelsen, Breinbjerg and Pold 2009, 197).

To forge cross-modal interactions digitally on the computer, an interface is necessary. Which interface is used to realise movements in cross-modal parameters and the approaches to these interactions, is an important factor and may be seen as an artistic choice. For other aspects of the digital creation process of an audio-visual work, such as source material and form, the choice of interface has an equally strong importance. The influence of using code in the programming environment, a visual GUI or a haptic controller on these different aspects is discussed.

This chapter will mostly consist of personal reflections on experiences with interfaces during the making process of fixed media pieces. I will refer specifically to the creation of *neuro.flux*, *[CONTAIN]* and an older work called *Ode To Ben Burt* (2018, abbreviated to *O.T.B.B*), which was my first audio-visual work and sparked my interest on the influence of interfaces. The term *interface* is often used relating to human-computer interactions: ‘A device or program enabling a user to communicate with a computer’ (Lexico. 2020). This definition covers most instances of the term as used in this chapter, but not all. I would therefore like to use a broader interpretation of the term: ‘a situation, way, or place where two things come together and affect each other’ (Cambridge Dictionary. 2020). In this case, the interface can be seen as a “means” in which artistic ideas are materialised.

A couple of years ago, while I was making *O.T.B.B*, which was completely developed and sequenced in code, I started to think about the influences from interfaces on the creation process; this work seemed to me to possess certain formal aspects that were lacking in my previous work, which was generally sequenced in a DAW (Digital Audio Workstation). However, *O.T.B.B* appeared to me, to have lost a certain continuity, or “flow”, which the earlier music did possess. This difference seemed to be largely caused by the difference in interface (or working environment). Aspects that play an important role in this difference are: the degree of haptic feedback, the degree and type of visual feedback and how visual or haptic controls are mapped to parameters. In subsequent works, I have attempted to find a balance between **formal ideas** and **intuition** by using different interfaces for different stages in the making process. When using a combination of interfaces for these different stages, a central issue is to determine which interface is suitable for achieving a particular goal. For instance, manually entering numbers in code might not be the most intuitive way to explore material and the controller might not be the most suitable for creating temporal organisation on a larger timescale.

Code

Code is often at the base of my audio-visual practice; for *O.T.B.B*, *[CONTAIN]* and *neuro.flux*, a large part of auditory and visual source material is generated in programming environments such as SuperCollider and openFrameworks. This material is either directly generated in code or in self-built programmes, which are controlled by a GUI or controller.

When **writing** these programmes, manually altering numbers is inevitable. While **exploring material** with the programmes, however, I generally try to avoid repetitively changing numbers in code where possible. An exception is *O.T.B.B*, where material was made by experimenting with changing numbers in code: any modification made in audio or image involved the action of removing the old number and typing in a new number. The code then had to be re-evaluated to experience the result which was not a very efficient way to experiment with the material. If a specific value is wanted for a parameter, however, which might be harder to realise with a knob or slider on a GUI or a joystick, it can be useful to enter the number manually. This can be the case for instance if one wishes multiple parameters to be *exactly* the same. For this reason, when making my own GUI, I like to include a number box displaying an editable value adjoined to certain GUI slider or knob, which gives visual feedback of that GUI element, thus keeping both methods of changing a parameter available.

In *O.T.B.B*, both the material and **form** were completely sequenced in code. Writing the sequencing part was tedious work which resulted in undesirable, redundant lines of code. This encouraged me to think more about long lines within the piece and to think ahead more to minimize time spent on rewriting code. In my opinion, this had a positive effect on the form of the final result. Another consequence of sequencing with lines of code, however, was that in order to start from a particular point in the piece, all the code preceding that point had to be disabled. These combined limitations made editing, designing transitions and layering very unintuitive.

Visual and sound events were triggered by the same code in *O.T.B.B*. The resulting synchronisation of the two media was established almost effortlessly and thus an attractive option. Images were generated in openFrameworks based on OSC (Open Sound Control) messages received from SuperCollider, where the sounds were synthesised. Not having to think about synchronisation led to a focus on other audio-visual parameters, such as correlation in *movement*. In hindsight, a parameter for temporal deviation could have easily been implemented, which could have resulted in a more interesting relationship between sound and image, but this was not a subject of focus at the time.

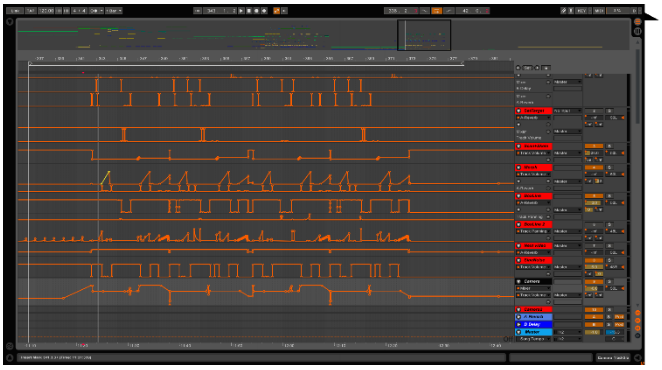
GUI

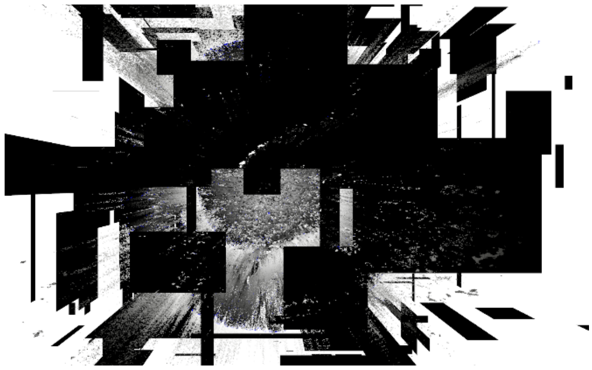
When a **material** generating programme reaches the point in development that parameters can be established, often the next step for me is to make a GUI based on sliders, knobs and buttons. Having a visual representation of the current state of parameters can help me to determine future adjustments. While this influence can help explore new combinations, it may also encourage artistic decisions based on the logic of the visual feedback of the GUI, rather than the logic of the material itself. Possibly more important for me is that using a mouse is a more haptic and efficient way of changing values than manually entering numbers. When using a GUI, processes can be changed in real-time, instead of first evaluating the code. Furthermore, the possibility to change two parameters at the same time (with a 2-dimensional slider) is also an advantage.

A GUI (left) for granular synthesis parameters based on gravitational forces with a visual representation (right)

After the mixed experience of making **formal** decisions in code, a DAW was used for the temporal organisation of both sound and image for *[CONTAIN]* and *neuro.flux*. I tried to approach the works with a fresh perspective by keeping the formal aspects in mind which resulted from sequencing in code while employing the versatility of layering and drawing envelopes from the DAW. Using graphical envelopes as the main formal tool has both advantages and disadvantages. The influence from these automation lines can be a particular limiting factor on artistic choices. The use of linear or exponential envelopes can, for instance, result in predictable changes and lead to choices made based on the visual representations. However, due to the "user friendly" design and my experience in this environment, these envelopes do offer the ability to dynamically change parameters and to layer them in a reasonably intuitive way, encouraging experimentation with these layers.

During the making process of the two pieces made in this DAW, there were discrete tracks for sequencing image and sound. To bridge these two modalities, a form of communication had to be established; the tracks committed to image had an embedded Max for Live patch, called *Livegrabber* which allowed the user to send OSC messages, derived from the automation on those particular tracks. On the tracks that were dedicated to sound, this automation was used to create envelopes that control not only track-specific data such as volume, panning, master sends, but also parameters from plugins such as filter frequencies, filter gains and reverberation characteristics. On the image specific tracks, the envelopes allowed the control of visuals in a way that was similar to controlling certain aspects of sound. Due to the fact that the visuals were live-generated while editing, changes made in visuals could be seen in real time. This proved to have a significant benefit over using rendered visuals. In both pieces there was a considerable amount of envelope-manipulated visual parameters, which varied largely in function. For example, some facilitated the adjustment of particle velocity, others were merely on/off switches. Shaping the visuals in a similar way to the sound opened up many possibilities to experiment freely compared to sequencing in code. It did however, cost more effort to precisely synchronise the modalities. Even though a more complicated relationship between sound and image was encouraged through this technique, in *[CONTAIN]* the goal was to join the two modalities: in a way it was a study in unification. Using the same technique in *neuro.flux*, I tried to experiment with moving away from unity using the approaches described before. The control of individual modalities by automation provided a good platform for exploring approaches. A number of vertically aligned points on the automation lines could, for instance, be created on both image and sound-tracks. Since the modalities were already linked, these points could act as **points of unity**. In between these points, additional points can be placed and manipulated to construct modal specific deviations. **Counterpoint** was conceived by a similar technique but instead of focussing on strategically placed points, a musical counterpoint with the cross-modal elements was attempted. Also, **transformations in individual modalities** could be realised by making changes in the envelopes of individual tracks.

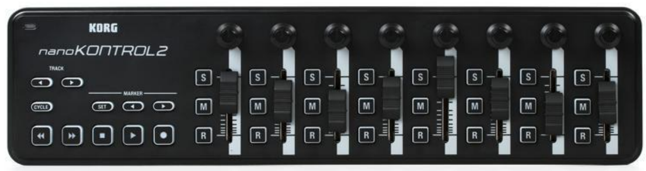




A number of envelopes dedicated to controlling the imagery of neuro.flux in the GUI of the DAW (left). The visual result could be seen in real-time (right).

Controller

Having determined the parameter ranges in the GUI, the next step for me in the development of a tool is to map physical controls to the parameters. For exploring auditory and visual **material** that is generated by a programme, using a controller has many benefits over using a GUI or making adjustments in code. Firstly, the use of a controller allows the change of more than 2 parameters at the same time, which can result in a more intuitive way of investigating different combinations. Secondly, with the use of a controller, I am under the impression that a freer exploration of the material is encouraged than when working with a GUI or in code. The absence of visual feedback may allow a certain focus on the material. For this reason, in this stage of the development a tool, the previously made GUI is no longer of relevance to me. Finally, and maybe most importantly, the controller is a more haptic way of manipulating the parameters. An important aspect of an interface is the degree in which the body is involved in the usage of it, in other words, the amount of haptic feedback. In my observation, one of the reasons for this is that the brain can unconsciously respond more precisely to sensory information when more muscles are involved in modifying parameters. Quite a few muscles are already mobilised in the operation of the mouse. The mouse is generally used to interact with a GUI, and not mapped as an expressive controller. When mapped for expressive purposes, however, it can actually be considered quite a versatile controller for exploring material in my opinion: there is no lack of experience in wielding the mouse as an input device through years of continual use. Depending on the model it may also feature a relatively high resolution. As well as the mouse, I have used a Korg NanoKontrol 2 midi controller to experiment with the *visuals* and a Logitech Extreme 3D Pro joystick for generating *sound* during the making process of *neuro.flux*.



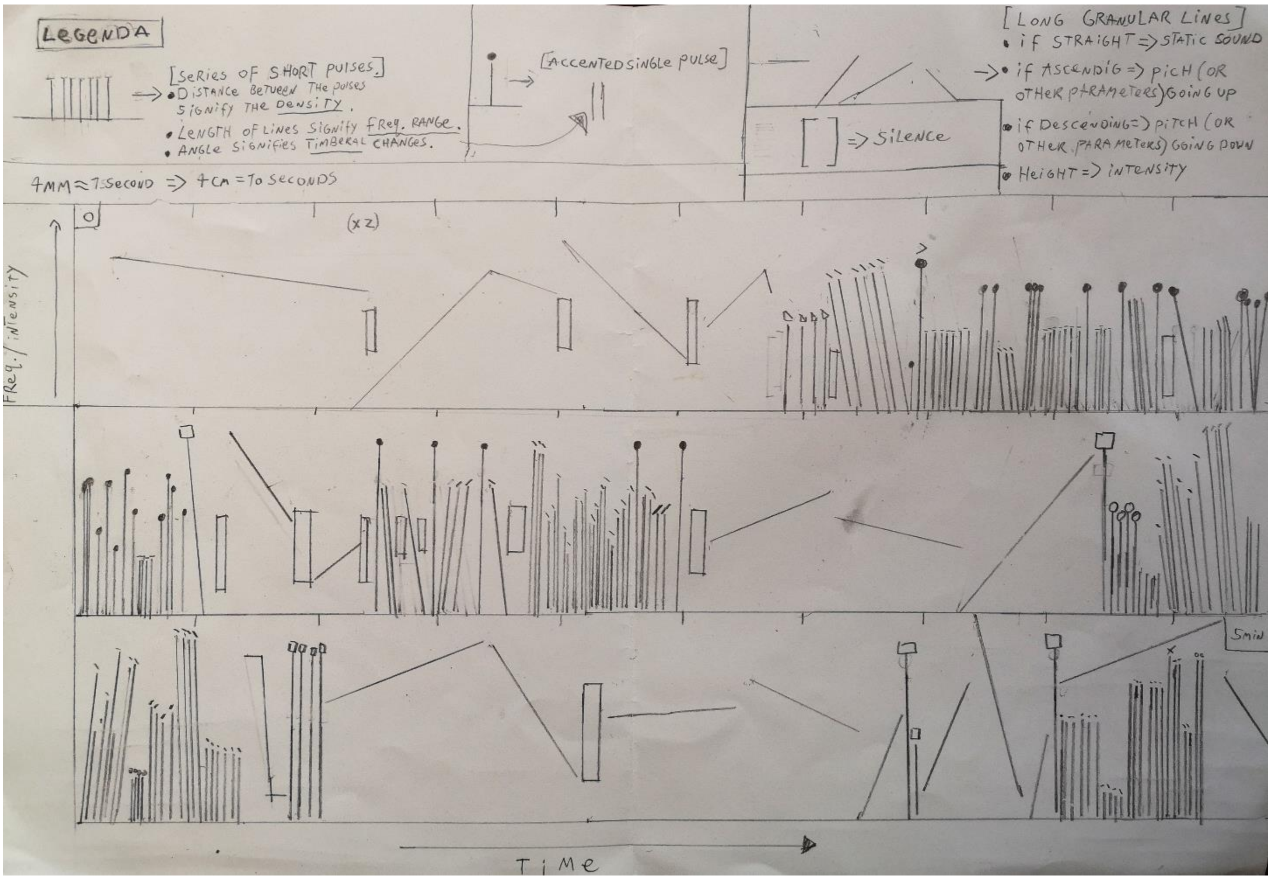
Logitech Extreme 3D Pro (left) and Korg nanoKontrol2 (right)

As well as controlling envelopes from image-specific tracks sent to openFrameworks with automation in the DAW, I also assigned most parameters to a MIDI controller to experiment with using these controls with slightly more haptic feedback. Using the controller to manipulate the visuals yielded mixed results. It was useful to try out complex combinations of visual elements, but less successful for "playing" the visuals. There were several reasons for this: I realised that I had not given the control mapping enough attention and I did not have a lot of practice with the controller, like one would have when one plays an instrument. Thus, the controller was not pushed to its potential expressivity. Where the joystick and the mouse, having a large resolution, worked as expressive controllers for me, the midi controller failed to give the necessary accuracy. The situation, for instance, where a (MIDI) knob featuring 127 values controls a virtual camera distance of 12000 pixels is undesirable. The fact that MIDI is still an industry standard is in my opinion a very limiting issue. The sheer number of buttons (35) which inhabit the Korg controller, however, can act as switches and proved to be a useful foundation for experimenting with **rhythmical counterpoint**. Being able to dynamically switch visual elements with fingers allowed for rhythmical experimentation with these elements together with sound, which was largely preconstructed. Because the switches took form as automation lines within the DAW, alterations within them could be recorded and modified by using the GUI afterwards. This way a sketch could be made while "playing" the visuals, which could then be adjusted more precisely.

As useful as a controller may be to explore material, it may not be the best tool to create form. A controller generally controls parameters in real-time and offers no representation of previous or future events or overview of a larger timescale. In a live performance situation, formal aspects can be created by improvisation, but when creating fixed media, I believe it is useful to have this larger overview.

Pencil on Paper (Sketches)

It is debateable if using pencil on paper can be called an interface. Following the definition stated before, however, it may be considered to be an interface to materialise formal ideas. Using pencil or pen on paper to draw shapes which represent formal ideas may be seen as a very indirect way of making sketches for electronic fixed media. Certain formal and structural concepts might be expressed more freely on paper than in code or in a DAW. A drawing, for instance, can symbolise timbral characteristics with abstracts shapes. Making sketches before making or editing material is very useful when a top-down approach is applied to a work. I personally usually prefer a bottom-up approach to making form. When a piece is well on the way, however, sketching out a representation of it on paper can help give an overview of the structure which is already present and give insight on potential larger lines. I often use this sketch as the starting point to address these longer lines and to attempt to conceive an outline of the entire piece. By doing this I hope to create a sense of totality while retaining the formal aspects which are derived from the material itself. At the end of the making process the final result in my case often does not resemble the "score" at all, but the act of outlining, observing and considering it still has a significant impact on the work.



A paper sketch drawn during the making process of O.T.B.B.

Future Experimentation

Having experienced the many benefits of working with a DAW, but also the restrictions, I would like to try sequencing in code again with a fresh perspective. I think that it can offer non-linear or generative formal aspects which may result in interesting audio-visual relationships. I am also interested in experimenting with other interfaces which I have not yet used, such as video editors and compositors. Most of all, however, in my future experiments I would like to further explore the possibilities of the use of a controller to play with cross-modal aspects. For instance, by controlling certain visual aspects with the one hand and certain sound properties with the other. This might result in a counterpoint which could be compared with playing separate musical lines with two hands on the piano.

Conclusion

In this paper possible cross-modal interactions between perceptual modalities are researched, primarily between sound and image. These interactions are entanglements that happen between the sensory modalities on a neurological level, before and during processing in the brain. Through practical experimentation and by examining related textual sources, I investigated how these interactions can be artistically and expressively manipulated.

There is no doubt that humans have a strong tendency to link sound and image when they happen simultaneously. In a previous work of mine (*[CONTAIN]*), I tried to achieve a certain audio-visual unity, which raised questions on the static audio-visual relationship. In the work which could be seen as the practical part of this research (*neuro.flux*), I attempted to challenge this unity by applying deviations in specific audio-visual aspects and thus creating perceptual disturbances. It was important for me, however, that a certain interconnectedness was retained. Texts on perceptual phenomena were taken as the point of departure to determine which aspects could potentially be altered. Information from these texts was then used to determine multi-modal properties and possible modulations of one sensory modality to another. The multi-modal properties which I surveyed in this paper are: synchronicity, material properties, embodied properties and spatial characteristics.

Based on considerations from this academic research, I established a number of *parameters*, mostly related to their potential to create cross-modal interactions. A number of relevant discoveries emerged from these considerations. Temporal deviation between sound and image seems to me to be the most compelling expressive tool. The order in which the modalities appear after being slightly shifted is crucial: if an image is delayed the shift is more noticeable than when the sound is delayed for instance. It appears that alterations made in material properties, should be subtle in order to retain a certain credibility. Embodied properties can add an extra layer of perceptual interaction. For instance, a tactile feeling can be produced by sub frequencies, and stimuli from the vestibular system can interact and be manipulated by visual and auditory impulses. Spatial characteristics can also create audio-visual interactions. Differences in the perception of distance can be induced by auditory time delays and virtual distance. The perception of volumetric size can likewise be altered by virtual space size and reverberation characteristics. Several audio-visual aspects, however, proved to be of less relevance to *neuro.flux*, for instance induced proprioception only occurs when actual human bodies are displayed. In the perception of the horizontal and vertical position and movement of an object in Euclidian space, sight seems to dominate over audition to such a degree that creating cross-modal interactions is ineffective in the stereo format: the spatial difference between information from visual and auditory stimuli is too small.

After establishing the parameters, three compositional **approaches** which can be applied to these parameters in the audio-visual environment are discussed. These approaches are examples of techniques of manipulations of one modality in relation to the other and have their origins in cinema, avant-garde film and my own practical experimentation. All three have been implemented in varying degrees in different sections of *neuro.flux*, mostly concerning temporal deviations. One of these approaches is *counterpoint*, which I have tried to realise with an interplay of multi-modal properties. Another approach is using *points of unity*¸ which has similarities with counterpoint but from a slightly different perspective. When this method of strategically placed points is applied to temporality, a certain *temporal elasticity* can be present in between the points in which the media can freely deviate in time. These points can also be applied to other multi-modal properties. The last approach discussed, concerned *transformations in a single modality*, which is frequently applied in *neuro.flux* to create shifts that are coherent to one specific modality. These uni-modal transformations are present more in audio than in image, due to the fact that the temporal organisation of the work is mostly based on sound.

The **interface** used to practically implement and experiment with the approaches above, influences the making process and artistic choices. Reflecting on personal experiences from creating three audio-visual pieces, it seems to me that certain interfaces are more suitable for certain approaches than others. This specific suitability also applies to the creation of source material and formal aspects. In the programming environment, repetitively altering numbers is inevitable when writing a programme, but can be best avoided while exploring source material or sequencing. For shaping formal aspects, a certain visual overview can be useful, which can be created with a GUI or with pencil on paper. In *neuro.flux*, all three compositional approaches were implemented using graphical tools, with linear and exponential envelopes in both image and sound to control the audio-visual parameters. This visual feedback, however, can have a limiting effect on artistic exploration and can result in predictable changes by choices made based on the visual representations. For exploring material, a controller seems to me to be the most suitable. The ability to change multiple parameters simultaneously and a certain freedom from visual feedback may allow another focus on the material. For realising rhythmical counterpoint, the controller also proved to be a useful tool. Playing visual elements with individual fingers allowed for rhythmical experimentation with these elements together with the already composed sound. In future experimentation it might be interesting to attempt to "play" both modalities with two controllers.

Communicating my thoughts in a textual environment sparked further developments in my ideas. It encouraged reflection on choices made during the creation process and influenced further artistic decisions. Through this research I have also gained knowledge of practical applications. For instance, discovering that the window of multi-sensory integration is not equal for each modality, practically helped me in the editing process. If an audio-visual event should be perceived as synchronised, the sound should be placed slightly later in time than the image. A-synchronicity on the other hand, was easier to reach by placing the image slightly later.

Reflections

The focus in *neuro.flux* has been on temporal deviations and much less on other multi-modal properties. Initially, this paper was concerned with (a-)synchronicity, on which a significant body of research in relation to art exists. Out of curiosity, however, I was tempted to dive into cross-modal phenomena in other audio-visual properties. Many of the concepts discussed in this text are connected to each other, such as temporal and spatial qualities, or immersion and audio-visual unity and I saw it as a stimulating challenge to attempt to separate these related concepts and in doing so, many of my ideas surrounding these concepts were clarified.

References

Abraham, Nicolas, Nicholas T. Rand and Maria Torok. 1995. Rhythms: *On the Work, Translation, and Psychoanalysis (Meridian: Crossing Aesthetics)*. California: Stanford University Press

Berger, Christopher and Henrik Ehrsson. 2017. "The Content of Imagined Sounds Changes Visual Motion Perception in the Cross-Bounce Illusion." *Scientific Reports* 7, 40123. Accessed 16/5/2020. https://doi.org/10.1038/srep40123.

Bertelsen, Olav W., Morten Breinbjerg and Søren Pold. 2009. "Emerging Materiality: Reflections on Creative Use of Software in Electronics Music Composition." *Leonardo* 42 (3): 197-202. Accessed 18/5/2020. https://doi.org/10.1162/leon.2009.42.3.197

Chion, Michel. 1994. Audio-Vision: *Sound on Screen*. New York: Columbia University Press.

Dichgans J. and T Brandt. 1978. "Visual-Vestibualar Interaction: Effects on Self-Motion Perception and Postural Control." In *Percetion. Handbook of Sensory Physiology Vol 8*. Edited by R. Held, H.W. Leibowitz, H.L Teuber, 755-804. Berlin: Springer

Eisenstein, S. M, V. I. Pudovkin and G.V Alexandrov. 1928. "A Statement." In *Film Sound: Theory and Practise*. Edited by Elisabeth Weis and John Belton. 1985, 83- 85. New York: Columbia University Press

Elliot, Thomas. 2015. "Hapticity in film." Accessed 18/5/2020. https://thomaselliottunsw.weebly.com/hapticity-in-film

Fisher, G.H. 1964. "Spatial localization by the blind." *American Journal of Psychology* 77 (1): 2-14. Accessed 18/5/2020. https://doi.org/10.2307/1419267

Ghuntla, Tejas, Hemant B. Mehta, Pradnya A. Gokhale, Chinmay J. Shah. 2014. "A comparison and importance of auditory and visual reaction time in basketball players." *Saudi J Sports* 14 (1): 35-38. Accessed 17/5/2020. http://www.sjosm.org/text.asp?2014/14/1/35/131616

Gibson, J.J, George A. Kaplan, Horace N. Reynolds, Kirk Wheeler. 1969. "The Change from Visible to Invisible: A Study of Optical Transitions" *Attention, Perception & Psychophysics* 5: 113-115. Accessed 16/5/2020. https://doi.org/10.3758/BF03210533

Giroux, Marion, Kulien Barra, Issam-Eddine Zrelli, Pierre-Alain Barraud, Corinne Cian, Michel Guerraz. 2018. "The respective contributions of visual and proprioceptive afferents to the mirror illusion in virtual reality." *PLoS One* 13 (8): e0203086. Accessed 18/5/2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6117048/

Goto, Suguru. 1999. 'The Aesthetics and Technological Aspects of Virtual Musical Instruments: The Case of theSuperPolm.' *Leonardo Music Journal* 9: 115-120.

Grau, Oliver. 2003. Virtual Art: *From Illusion to Immersion*. London: The MIT Press.

Handel, Stephen. 2006. Perceptual Coherence: *Hearing and seeing*. New York: Oxford University Press.

Hertz, Paul. 1999. 'Synaesthetic Art: An Imaginary Number?' *Leonardo* 32 (5): 399-404.

Hidaka, Souta, Yuko Manaka, Wataru Teamoto, Yoichi Sugita, Ryota Miyauchi, Jiro Gyoba, Yôiti Suzuki, Yukio Iwaya. 2009. "Alternation of Sound Location Indeuces Visual Motion Perception of a Static Object.' *PLoS One* 4 (12): e8188. Accessed 18/5/2020. https://doi.org/10.1371/journal.pone.0008188

Jaekl, Philip, Salvador Soto-Faraco and Laurence R. Harris. 2012. 'Perceived size change induced by audiovisual temporal delays.' *Experimental Brain Research* 216: 457-462. https://doi.org/10.1007/s00221-011-2948-9

Klatzky, Roberta and Susan Lederman. 2010. "Multisensory Texture Perception." In *Multisensory Object Perception in the Primate Brain*, edited by Marcus Naumer & Jochem Kaiser, 210-227. New York: Springer.

Knudsen, E. I. and M. S. Brainard. 1995. "Creating a unified representation of visual and auditory space in the brain." *Annual Review of Neuroscience* 18: 19-43. Accessed 18/5/2020. https://doi.org/10.1146/annurev.ne.18.030195.000315

Kohler, Evelyne, Christian Keysers, M. Alessandra Umiltá, Leonardo Fogassi, Vittorio Gallese, Giacomo Rizzolatti. 2002. "Hearing sounds, understanding actions: Action representation in mirror neurons." *Science* 297 (5582): 846–848. Accessed 17/5/2020. https://doi.org/10.1126/science.1070311

Maempel, Hans-Joachim and Matthias Jensch. 2013. "Audio-visual interaction of size and distance perception in concert halls – a preliminary study." Accessed 18/5/2020. https://www.semanticscholar.org/paper/Audio-visual-interaction-of-size-and-distance-in-a-Maempel-Jentsch/d6cbc8803dc0855a4317d29a40476f82662fc4ff

Mera, M. 2016. "Material Film Music." In *The Cambridge Companion to Film Music*, edited by M. Cooke & F. Ford, 157-172. Cambridge: Cambridge University Press.

Mukamel, Roy, Arne D. Ekstrom, Jonas Kaplan, Marco Iacoboni, Itzhak Fried. 2010. "Single-Neuron Responses in Humans during Execution and Observation of Actions." *Current Biology* 20 (8): 750-756. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904852/

Nuehoff, John. 2004. Ecological Psychoacoustics**.** Cambridge, Massachusetts: Elsevier Academic Press.

Pallasmaa, Juhani. 2017. "Touching the world – Vision, Hearing, Hapticity and Atmospheric Perception." Accessed 23/5/2020. http://invisibleplaces.org/2017/pdf/Pallasmaa.pdf

Ribas, Luísa. 2014. "Sound and image relations: a history of convergence and divergence". *Divergence Press*. Accessed 16/5/2020. http://divergencepress.net/2014/06/02/2016-10-27-sound-and-image-relations-a-history-of-convergence-and-divergence/

Salter, Chris. 2011. "The Question of Thresholds: Immersion, Absorption, and Dissolution in the Environments of Audio-Vision". In *Audiovisuology 2*, edited by Dieter Daniels, 200-222, Köln: Verlag der Buchhandlung Walther König.

Silva, Carlos César, Catarina Medonça, Sandra Mouta, Rosa Silva, José Creissac Campos, Jorge Santos. 2013. "Depth Cues and Perceived Audiovisual Synchrony of Biological Motion". *PLoS One* 8 (11): e80096. Accessed 17/5/2020. https://doi.org/10.1371/journal.pone.0080096

Smalley, Dennis. 2007. "Space-Form and the Acoustic Image." *Organised Sound* 12 (1): 35-58.

Soto-Faraco Salvador, Jessica Lyons, Micheal Gazzaniga, Charles Spence, Alan Kinsgstone. 2002. "The ventriloquist in motion: illusory capture of dynamic information across sensory modalities." *Cognitive Brain Research* 14 (1): 139-146. Accessed 17/5/2020. https://doi.org/10.1016/S0926-6410(02)00068-X

Spence, Charles & Sarah Squire. 2003. "Multisensory Integration: Maintaining the perception of Synchrony". *Current Biology*. 13: 519-521. Accessed 17/5/2020. https://doi.org/10.1016/S0960-9822(03)00445-7

Taberham, Paul. 2018. Lessons in Perception: *The Avant-Garde Filmmaker as Practical Psychologist*. New York: Berghahn Books.

Taylor, J.L. 2009. "Proprioception." In *Encyclopedia of Neuroscience*. Edited by Larry R. Squire, 1143-1149. Sydney: Elsevier. Accessed 18/5/2020. https://doi.org/10.1016/B978-008045046-9.01907-0

Fujisaki, Waka, Naokazu Goda, Isamu Motoyoshi, Hidehiko Komatsu, Shin'ya Nishida. 2014. "Audiovisual integration in human perception of materials".
Journal of Vision 14, 4. Accessed 17/5/2020. https://doi.org/10.1167/14.4.12

Wallace, Mark & Barry Stein. 1997. "Development of Multisensory Neurons and Multisensory Integration in Cat Superior Colliculus". Journal of Neuroscience 17, 7: 22429-2444. Accessed 17/5/2020. https://doi.org/10.1523/JNEUROSCI.17-07-02429.1997

Waterworth, John A. 1997. "Creativity and Sensation: The Case for Synaesthetic Media", *Leonardo* 30, 4: 327-330. Accessed 16/5/2020. https://doi.org/10.2307/1576481.

Software

Kneppers, Mathijs. *Livegrabber*. Version 2.7. Windows 10. https://cycling74.com/tools/livegrabber/

Lieberman, Zach *openFrameworks*. Version 0.9.8. Windows 10. https://openframeworks.cc/

McCartney, James *SuperCollider*. Version 3.10.2. Windows 10. https://supercollider.github.io/

Websites

Bhat, Manjunath. 2014. "Tactile vs Haptic feedback". Accessed 17/5/2020. https://manjubhat.wordpress.com/2014/04/06/tactile-vs-haptic-feedback/

Cambridge Dictionary. 2020. "interface." Accessed 17/5/2020. https://dictionary.cambridge.org/dictionary/english/interface

Lexico. 2020. "interface." Accessed 17/5/2020. https://www.lexico.com/en/definition/interface

Lexico. 2020. "synchronicity." Accessed 17/5/2020. https://www.lexico.com/en/definition/synchronicity

Lexico. 2020. "tactile." Accessed 17/5/2020. https://www.lexico.com/en/definition/tactile

Merriam-Webster. 2020. "counterpoint." Accessed 17/5/2020. https://www.merriam-webster.com/dictionary/counterpoint

Neuroscientifically Challenged. 2015. "Know your brain: Vestibular System." Accessed 17/5/2020. https://www.neuroscientificallychallenged.com/blog/know-your-brain-vestibular-system

Oxford Reference. 2020. "ventriloquism effect." Accessed. 19/5/2020. https://www.oxfordreference.com/view/10.1093/oi/authority.20110803115439532

Psychology Today. "Synesthesia." Accessed 23/5/2020 https://www.psychologytoday.com/us/basics/synesthesia

Youtube. 2009. "Rick Ashley – Never Gonna Give You Up (Video)." Accessed 19/5/2020. https://www.youtube.com/watch?v=dQw4w9WgXcQ

Sciencing. 2017. "What Is a Tactile Sensation." Accessed 19/5/2020. https://sciencing.com/tactile-sensation-7565666.html

Wikipedia. 2020. "Depth perception." Accessed 18/5/2020. https://en.wikipedia.org/wiki/Depth_perception

Wikipedia. 2020. "McGurk effect." Accessed 17/5/2020. https://en.wikipedia.org/wiki/McGurk_effect

Wikipedia. 2020. "Synaesthesia." Accessed 17/5/2020. https://en.wikipedia.org/wiki/Synesthesia

Pictures

Excelep. 2020. "logitech-extreme-3d-pro-joystick.jpg" Accessed 19/5/2020. https://www.excelep.com.pk/131-thickbox_default/logitech-extreme-3d-pro-joystick.jpg

Sweetwater. 2020. "nanoKON2bk-large.jpg." Accessed 17/5/2020. https://media.sweetwater.com/api/i/f-webp__q-82__ha-42e5bd3ed342753f__hmac-c7115f9398bfb095f3041b7f38e19e02eca8d4b3/images/items/750/nanoKON2bk-large.jpg.auto.webp

Videos

BBC 2 Horizon. 2010. "Is Seeing Believing?" Ep. 4.

Accessed 17/5/2020

https://www.youtube.com/watch?v=2k8fHR9jKVM

Lucas, George, dir. 2002. Los Angeles: 20th Century Studios. *Star Wars: Episode II: Attack of Clones. DVD*
https://www.youtube.com/watch?v=3ME5jhsgmB4

Ikeda, Ryoji. 2019. "Ryoji Ikeda. Code-Verse. Biennale Arte 2019, Venice." Accessed 20/5/2020. https://www.youtube.com/watch?v=DH2GbM6ZqUI