## Introduction

In the artist film *Erysichton* (2013) by the Canadian artist Jon Rafman a striking resemblance to the famous sentient computer of Stanley Kubrick's *2001: A Space Odyssey* (1968) is present. Kubrick's film shows the main antagonist, a spacecraft computer named HAL 9000, as a red and yellow camera eye that is ever present in the space of the story. This non-blinking camera eye, voiced by Douglas Rain, speaks in a calm and articulated manner, and contrasts greatly with the voices his crew members who are more disfluent, especially when insecure. As the critic and composer Michel Chion has stated on the character: "HAL's speech differs entirely from all previous audiovisual representations of the voices of robots, computers, or oracles: there's no visual modulation with moving parts or on-off lights synchronized with the vocal utterances."[1] In *Erysichton* a recurring visual motif is a green color graded image of a reptilian eye in the same close-up as HAL's camera eye is represented in *2001*. The shot of the reptilian eye is accompanied by a voice, however, it is uncertain whether to connect the eye and the voice to the same source. There are two important distinctions in the voices: firstly, HAL's voice is connected to a body, albeit in the form of hardware, while the voice that speaks over the green eye's image is *disembodied*, there is no indication that the eye is the owner of that voice. Secondly, there is no mechanical noise that accompanies HAL's voice whereas the voice in Rafman's video is *synthesized*, it is artificially transformed to create a robotic voice effect.

The object of research for this thesis will be the disembodied and synthesized voice in the recent works of artist filmmakers. My main question is grounded in one of the main concepts in the theory on sound by Michel Chion, added value. According to Chion, added value is "the expressive and informative value with which a sound enriches a given image so as to create the definite impression, in the immediate or remembered experience one has of it, that this information 'naturally' comes from what is seen, and is already contained in the image itself."[2] Added value is engaged when a sound and an image occur at the same time, forging an immediate and necessary relationship between what is heard and seen. The main question of this study is: what is the added value of the disembodied and synthesized voice in the contemporary film work of Canadian artist and filmmaker Jon Rafman, British artist and filmmaker Ed Atkins, British artist and filmmaker Helen Marten and German-Japanese artist and

---

[1] Chion 2009, p. 333.
[2] Chion 1994, p. 5.

filmmaker Hito Steyerl. But what is the importance of an inquiry into this vocal instrument as it is functioning in artist film? Why does the voice demand this attention?

**Vision as the master sense**

A common thing to say when starting an inquiry into contemporary media is that we live in an age with an abundance of visual stimuli. We emphasize that the human of the twenty-first century is constantly bombarded with visual information originating from various sources such as computers, our smartphones, commercials and so forth. As a result, our attention frequently fails to notice the vibrations that more often than not accompany the barrage of moving and still images. This tendency towards the visual has triggered criticism towards approaching media as exclusively visual. Art historian W. J. T. Mitchell, for example, has proposed to question the term 'visual media' and to approach media as specific mixtures of specific media.[3] By dismissing the term he opts for a more distinct classification, one that also considers other sensory dimensions such as the aural and the tactile, and possibly the senses of smell and taste (even though Mitchell does not explicitly mention the latter two). Since antiquity the sense of sight has been the most dominant of the classic five senses (sight, hearing, smell, taste, touch). This predominance of the visual is not a result of the invention of printing and the advent of the modern era, as many would be tempted to believe. According to historian Robert Jutte, sight's position in the hierarchy is typically interpreted as a product of human evolution.[4] Sight aids best, following Aristotle and Plato, in the process of acquiring knowledge and understanding.

Sight's privileged position in the hierarchy of the senses has only weakened since the last quarter of the twentieth century. Film and media scholar Nora M. Alter writes that two groundbreaking studies investigating the acoustics of media were only published in 1992, Rick Altman's *Sound Theory, Sound Practice* and Douglas Kahn's and Gregory Whitehead's *Wireless Imagination: Sound, Radio, and the Avant-Garde*. Alter accounts this growing fascination towards sound to technological developments in sound production from the 1960s and onwards that have enabled sound to become ''more easily produced, distilled, manipulated, and controlled than ever before.''[5] Another reason for sound to be recognized as an object of

---

[3] Mitchell 2011, p.76.
[4] Jutte 2004, p. 64.
[5] Alter 2004, p. 1-3.

research is the growing distrust for vision as the master sense. From the twentieth century onwards ocularcentrism aroused a distrust among intellectuals resulting in a "palpable loss of confidence in the hitherto 'noblest of the senses'."[6] As much as various authors pose that sound has been undertheorized, the last decade has seen a renewed interest in sound. For my argument I approach the study of sound interdisciplinary and in line with as it is theorized in two disciplines: the theory and history of art (with an emphasis on video art) and film studies.[7] The aforementioned publications by Alter can be seen as critical points in the study of sound, with Kahn and Whitehead covering the arts and Altman film studies. In the following I shortly outline how sound has been theorized in both the visual arts and film studies.

**Sound in theory**

When speaking of sound within the theory of art one mostly discusses the artistic discipline of sound art or, as Douglas Kahn and Gregory Whitehead term it, audio art. Only in the second half of the twentieth century there was a rise of academic interest around sound matters in the arts. One of the reasons for the lack of historical and theoretical writing on aurality is that, except for the practice of music, there never existed an artistic occupation solely devoted to acoustics. According to Kahn and Whitehead "there is no history of a self-described and autonomous art in the way that one might think of the history of sculpture."[8] In fact, the term 'sound art' was only conceived in 1983 when composer William Hellermann curated an exhibition at the Sculpture Center in New York which was called 'Sound/Art'.[9] Another reason for the absence of writing has been the privileging of music as the art of sound. This persistent privileging has led to the strange situation that all sound practices had to be categorized within music. Noise and sounds not originating from traditional instruments were considered extra-musical (in Latin 'extra' literally means 'outside of/beyond'), resulting in the slow appropriation of any sound to the musical domain. Even though studies such as the publications by Kahn and Whitehead have catalysed a resurgence of academic interest in the properties of

---

[6] Jay 1993, p. 588.
[7] In accordance with Mitchell's statement that there are no visual media, by 'theory and history of art' I refer to the line of study what is more commonly identified as the visual arts. However, as sound-based art, with this reasoning, has to be positioned within the visual arts I have to revise my formulation of the discipline to a less visually-oriented term.
[8] Kahn and Whitehead 1992, p. 2.
[9] Licht 2009, p.3.

sound-based or sound oriented art, the sonic dimensions of contemporary films by artists are still largely overlooked. In this study I have chosen to approach the case studies as artist films. The term combines the disciplines of art history and film studies, and as the case studies owe much to art history, more specifically video art, and the research tools are mostly inherited from film studies I see the 'artist film' as the most suitable interdisciplinary term.

The history of video art has been important for the attention towards sound. As Holly Rogers, as quoted in media scholar Paul Hegarty's publication on video and sound art, points out in her book *Sounding the Gallery* (2013) video emerged "from avant-garde music and sound use (as well as performance and conceptual art)."[10] More importantly, as artist Bill Viola explains, in a technological sense, "video has evolved out of sound (the electromagnetic) and its close association with cinema is misleading since film and its grandparent the photographic process are members of a completely different branch of the genealogical tree (the mechanical/chemical). The video camera, being an electronic transducer of physical energy into electrical impulses, bears a closer original relation to the microphone than it does to the film camera."[11] Another artist, Steina Vasulka, has noted that "video always came with an audio track, and you had to explicitly ignore it not to have it."[12] In the order of the history of videotape it is the audiotape that was developed first, and consequently artist who picked up the medium paid attention to this perception that video contains a fundamental audiovisuality.[13] It is strange then that some publications on video art do not always mention or acknowledge this relation to sound and consequently discuss the works mostly for their visual quality. However, as many video works nowadays are created digitally (such as the case studies in this study) this debt to sound is not as self-evident and this could be a reason for the lack of attention to sound in these more contemporary digitally-based films. Moreover, even though sound has not always been overlooked in art history, a research object as specific as the disembodied and synthesized voice has certainly not been heard and will receive its share of attention in this study.

In the preceding paragraph I briefly outlined how the study of sound has been treated within theories of art and the special role of video in raising attention to audio. Within the discipline of film studies sound has had much more attention from academics and critics alike,

---

[10] Hegarty 2015, p. 2.
[11] Viola 1990, in Viola 1995, p. 158-159.
[12] Meigh-Andrews 2006, p. 85.
[13] Spielmann 2008, p. 7.

but most widespread attention has been directed to the study of film music.[14] This does not mean that within sound the voice remains undertheorized, as studies have been dedicated to voice-over narration. The difference, however, is that when one studies 'voice-over narration', such as film scholar Sarah Kozloff in *Invisible Storytellers: Voice-Over Narration in American Fiction Film* (1988), one is occupied with studying the act of communicating a narrative over the images of the screen, and this diverges from my study because I am not specifically focused on content uttered by a voice.[15] Rather I am also concerned with the voice its demeanor. How does it sound and how does it approach the audience?

Even the most well-known and arguably most-used introduction to the art and analysis of film, David Bordwell's and Kristin Thompson's *Film Art: An Introduction* (originally published in 1979, with editions reprinted until 2017), omits a segment on the voice as a medium in film. Film scholar Rick Altman criticizes *Film Art* for applying a musical model for the analysis of film sound, arguing that musical criteria in film sound seriously limit the study. According to Altman "musical notation assumes that each sound is single, discrete, uniform and unidimensional," however, sound is a heterogeneous phenomenon and therefore "when we listen to recorded sound we are therefore always listening to a particular account of a specific event."[16] This musical perspective on film may account for the large amount of publications on film music and lack of it on other film sounds. Among these overlooked sounds stands the voice, a sonic component often confused with speech. The voice encompasses any sound made my the human vocal tract, while speech is limited to uttering products from a lexicon. As Michel Chion has explained, the act of speech is mostly heard for its meaning or signification, forgetting the medium of the voice itself.[17]

---

[14] For a short overview of publications dedicated to the study of film music I refer to studies like Theodor Adorno's and Hanns Eisler's *Composing for the Films* (1947), Claudia Gorbman's *Unheard Melodies: Narrative Film Music* (1987), Caryl Flinn's *Strains of Utopia: Gender, Nostalgia and Hollywood Film Music* (1992), Arthur Knight's and Pamela Robertson's *Soundtrack Available: Essays on Film and Popular Music* (2001) and Mervyn Cooke's *A History of Film Music* (2008). In the book series *Very Short Introductions* by the Oxford University Press there is no edition engaged with film sound. Relevant to this study, only the editions *Film, Film Music* and the technically oriented edition on *Sound* have been published, suggesting a dominance of the study of film music within the discipline.
[15] A key text on the cinematic voice is Mary Ann Doane's 'The Voice in the Cinema: The Articulation of Body and Space' (1980), other articles focus on the specific qualities of actors/characters such as Claudia Gorbman's 'The Master's Voice' (2014) and Katherine Kinney's 'The Resonance of Brando's Voice' (2014), and on characters whose words go unheard as in Justin Horton's 'The Unheard Voice in the Sound Film' (2013).
[16] Altman 1992, p. 16.
[17] Chion 1999, p. 1.

**Theoretical framework**

In film studies few publications have been written on the cinematic voice. However, a number that does will be outlined in the following. Firstly, in the study of documentaries the voice has been theorized as a crucial factor in a film's meaning. For instance, according to film scholar Jeffrey Ruoff, voice-over "has long been one of the stylistics signatures of documentary sound", directing the audience's attention to specific components of the film.[18] The power of a voice is not be underestimated and it should not be strange that a frequently used term for the disembodied voice-over in documentary film is named the 'Voice of God' commentator. According to film scholar Bill Nichols this anonymous and invisible, yet ever present, voice "arose in the 1930s as a convenient way to describe a situation or problem, present an argument, propose a solution, and sometimes to evoke a poetic tone or mood."[19] Nichols explains that the Voice of God is a voice of authority, more traditionally used as a way to transmit a certain meaning to the audience, suggesting what to make from the documentary. However, the voice, when not under the guise of the Voice of God, can achieve the opposite. Instead of convincing the audience of a particular way of thinking, the voice can address problems of subjective representation. Whether originating from a filmmaker, an offscreen character or onscreen character or not, a voice-over can emphasize that something subjective is being presented toward the audience, something to be argued with. The use of the subjective voice in documentary is reminiscent of its use in essay films by the likes of Chris Marker and certain avant-garde and experimental film and video (such as the ones made by the artists in this study). Given this resemblance of documentary form and films that border around the art world, in chapter 1 I will make an attempt to compare and apply the theory of documentary on the voice in the case-studies.

The cinematic voice has also been a topic of study in narrative film. Next to a handful of articles, Michel Chion's *The Voice in Cinema* (1999), a publication solely devoted the instrument of the voice, both embodied and disembodied, both from the perspective of the first-person and the third-person, functions as an important text in my theoretical framework. On the importance of conducting research on the voice Chion argues that, as cinema is vococentric or verbocentric phenomenon, "it almost always privileges the voice, highlighting and setting the latter off from

---

[18] Ruoff 1992, p. 222.
[19] Nichols 2009, p. 14.

other sounds."[20] By vococentric Chion refers to all sounds produced by the voice while verbocentric focuses on the production of speech and thus words. When someone is surrounded in a sound environment, his or her attention will most likely be captured by the voices, and not to any other sound. Chion compares the voice with its visual counterpart: the face. Transposing a quote by Alfred Hitchcock, who states that "no matter what the framing [...] the first thing that people will look at - faces," the French scholar argues that the "first thing people hear is the voice."[21] But in contrast to a face, a voice is not necessarily attached to something on screen. Chion devotes his publication *The Voice in Cinema* to the disembodied voice, but this is largely a voice that can still be connected to an on screen source. My research diverges from this object in two ways, by exclusively discussing disembodied voices with no face or body to be attached to, and by investigating voices that are subject to synthesization.

There are several ways of indicating that a filmic voice is not connected to a body within the film. Terms such as voice-over, voice-off and narrating voice seem to have a very general use in a literary, radio, television and film context, however these terms do not necessarily imply that the source of audible voice is not perceivable in the film. It also can't be called asynchronous, since, even though the voice and its source do not exist at the same time, its source can very well be visible when the voice is not speaking. A term very specific to the analysis of film could be applied here, the *nondiegetic* voice, not only lacking from the image but also from the story world. But in spite of that, for our current case studies, the term *nondiegetic* does not suffice because it is specifically used to discuss a narrative or plot, which typically means a feature film (Both Chion and Bordwell use the term, and both of their publications are primarily concerned with features). The artist films featured in this study do not always adhere to a clear narrative, making it more difficult to speak of a story world. In *Audio-Vision* Chion introduces the term 'acousmatic', borrowed from the theory of the composer Pierre Schaeffer (who adopted it to describe the mode of listening to radio, telephone and records), referring to "sounds one hears without seeing their originating cause."[22] A filmmaker can choose whether to visualize something that was previously acousmatic, erase its visual qualities and make it acousmatic or let it remain acousmatic altogether. In *The Voice in Cinema* Chion remodels the term to specify an acousmatic voice to an *acousmêtre*, and specifically to a voice never or not yet to be visualized, a *complete acousmêtre*.[23] This term comes closest to what I have been

---

[20] Chion 1994, p. 5.
[21] Chion 1999, p. 6.
[22] Chion 1994, p. 71.
[23] Chion 1999, p. 21.

describing as *disembodied*, something without a physical body, and I will employ the terms interchangeably. In contrast to the study of the disembodied voice, the synthesization of the voice remains undertheorized. With synthesization I refer to the artificial transformation (which in the context of this study is mostly digital, but can very well be analogue) of recorded voices in order to sound different. Only writings on robotic or inhuman sounding voices function as relevant theory to synthesized voices. In the chapters I will refer to various scholars such as Dave Tompkins, Allain Boillat, David Sonnenschein and Andy Birtwistle. Also, due to the lack of academic publications on the subject I refer to non-scholarly sources to strengthen my argument. Can new sonic characteristics be accounted for a different value of a disembodied voice?

**Structuring the voice**

The thesis is divided in three main chapters, the first chapter focuses on common notions of power attributed to the disembodied voice. This chapter closely examines the characteristics of the disembodied vocal instrument as it is being used in Jon Rafman's *Erysichthon* and Ed Atkins's *Delivery to the Following Recipient Failed Permanently*. Firstly, in light of Rafman's film, I discuss theories of the power of the disembodied voice by film scholars Michel Chion and Mary Ann Doane, but also the authority of the voice as it is theorized in the study of documentary filmmaking by documentary scholars Bill Nichols and Michael Renov, in relation to the Rafman film. Consequently, the specifics of male and female voices that are heard will be investigated through theories by film scholars Kaja Silverman, Sarah Kozloff and Amy Lawrence. Afterwards, the focus shifts towards a mode of vocal communication that does not employ a lexicon, since, as literary scholar Steven Connor posits, the voice transmits information not only through the use of words but also through sobs, hums and coughs. Here, Atkins's *Delivery to the Following Recipient Failed Permanently* will be put under investigation in conjunction with writings by film scholars Robert Stam and Germain Lacasse. The last part of the chapter is devoted to the synthesized aspect of the disembodied voice. I start with a short introduction to the technology that is the foundation of synthesized voices. Afterwards I interrogate whether the synthesized aspect of the voice influences notions of power. What are the consequences of altering the sound of the vocal medium to its position within film?

The second chapter is concerned with the questions of unity and disunity. I open with a paragraph on what film scholar Rick Altman defines as the sound film's fundamental lie and relate it to theories by film scholars Jay Beck, Michel Chion and Mark Kerins. Subsequently, the film *Dust and Piranhas* by Helen Marten serves as a case study for an investigation into Michel Chion's concept of 'the real and the rendered'. Afterwards I discuss whether *Woods of Arcady* by Jon Rafman can be considered as 'counterpoint', a term coined by Soviet director Sergei Eisenstein, in the context of a more updated text by film scholar Kristin Thompson on 'counterpoint'. Lastly, the emphasis shifts to the concept of 'noise', as defined by economic and social theorist Jacques Attali, and related texts by curator Caleb Kelly, art historian Andy Birtwistle and the artist Hito Steyerl, and will be applied to the film *How Not To Be Seen: A Fucking Didactic Educational .MOV File* by the latter artist in order to study whether the voice is in harmony with other sounds, and, if not, what does it disrupt? What are the disunifying qualities of these films and what are its consequences?

The final chapter recedes from the analysis of the film works as texts in the Barthesian sense of the words and instead treats them as events. Following Rick Altman, this part does not consider the case studies as autonomous aesthetic entities. Instead, approaching the film as event means also taking into account the spatial conditions of the screenings and situations, considering the auditorium, the gallery and installation. The events for the films *Factory of the Sun* by Hito Steyerl and, returning to the first case study, *Erysichthon* by Jon Rafman will be moments of focus. For the third chapter I consult writings on the theory of art to substantiate my argument, such as texts by art historians Andrew V. Uroskie and Paul Hegarty, but also sources such as non-scholarly online articles and videos.[24] In the analysis of the works four attributes to the film event, as summarized by Altman, are discussed: *three-dimensionality*, *materiality*, *heterogeneity* and *multi-discursivity*. What are the consequences for the interpretation of the voice in the context of a film event and is it relevant to study sound in a more spatial understanding of film? After the third chapter I conclude with the results and a reflection on my research.

---

[24] The online sources serve to provide a better impression of the specific exhibitionary situations of the case studies.

**Chapter 1:** Sourceless Voices and Otherworldly Inducement: The Powers of the Disembodied and Synthesized Voice


The following chapter investigates the powers disembodied and synthesized voices can exert over the audience, in order to define the expressive and informative value of the vocal instrument in contemporary artist film. This is not meant as a generalizing statement as the findings resulting from the analysis of my case studies are by no means directly applicable to other artist films. I start with a discussion of Canadian artist Jon Rafman's 2015 film *Erysichthon* wherein I examine theories on the voice and power by film scholars Mary Ann Doane and Michel Chion and subsequently compare the usage of the film's voices with regard to voice-over styles common in documentary filmmaking as theorized by documentary film scholars Bill Nichols and Michael Renov. In the discussion of *Erysichthon*, assisted by the theories of Doane but also by film scholars Kaja Silverman, Sarah Kozloff and Amy Lawrence, I also argue how common notions on male and female voices are subverted in the film by giving the female voice an unusual demeanor. After rising these issues in Rafman's film I turn to another work, this time by British artist Ed Atkins and his 2011 film *Delivery to the Following Recipient Failed Permanently*. In this case study I explore how the voice's juxtaposition to non-verbal sounds can be interpreted and whether this influences the power of the disembodied voice in the light of theories by semiotician Mikhail Bakhtin and film scholar Germain Lacasse. At the end of the chapter the disembodied voices of both films are under investigation for their synthesized qualities. Firstly, in *Delivery* I show how the exaggeration of a voice can make it lose strength and secondly, in *Erysichthon*, I argue that the voice's synthesization results in disturbance and alienation. This concentration on the synthesized sound of the voices in both film results from my view that synthesization of the voice has received the least academic attention. For this reason I consult theories that are not specifically written to study synthesized film voices but turn to various disciplines such as an article by cognitive scientist Katarzyna Pisanski on human voice modulation and a book by music journalist Dave Tompkins on the history of the vocoder in order to give my discussion on the sound of the voice more body. What kind of power results from the disembodiedness and synthesization of the voices in the two films?

**Power, Contemplation and Silence In *Erysichthon***

In Jon Rafman's eight minute one-channel digital film *Erysichthon* the audio-spectator is confronted with a barrage of images of objects and animals devouring themselves, a snake eating its own tail, or each other, a computer generated dragon devouring another dragon. According to art critic Mitchell Anderson, ''named for the mythological Greek king cursed with insatiable hunger, the video approaches subjects with both critique and reverence. The snake eating its own tail (...) is as mesmerizing as it is banal, referencing the film's namesake's demise and Rafman's view of cultural intake.''[25] With the auditive track of the film playing eerie soundscapes or being silent, *Erysichthon* also shows the audience images of a cube being absorbed by a black sludge, someone on a swing set in an eternal loop and never ending hallways of data centers, respectively referring to themes as self-devourment, being trapped in a loop and facing a perpetual stream of imagery. All this is slowly paced and repetitively hurled towards the audio-spectator while two disembodied voices speak (in total the voices only speak for about two of the film's eight minutes). In fact, the film adopts the form of a loop since the beginning and ending of the film are blurred by playing the same sequence at both moments, turning the film into something that is also perpetual. The images of the film seems to put forth the idea that humans are in danger of over consuming digital imagery and that it will make us 'fall into a void', as one disembodied voice adds to the visuals.[26] But is the message of this voice convincing and powerful?

In *Erysichthon* two disembodied voices alternate uttering phrases that either specifically comment on the shown images or contemplate thoughts that do not directly make a clear connection to the images. The unnatural sounding voices regularly address the audio-spectator directly in the first half or the film. According to Mary Ann Doane the power, or *authority* as she terms it, of the disembodied voice originates from its quality as ''direct address, it speaks without mediation to the audience, by-passing the 'characters' and establishing a complicity between itself and the spectator - together they understand and thus *place* the image.''[27] The voice without a localizable body interprets the image from a different space, placing it beyond criticism. The two disembodied voices in *Erysichthon* employ a similar strategy. The voices regularly address the listener in the following manner ''you were awake all night and the only

---

[25] Anderson 2015.
[26] Rafman, *Erysichthon*, 2015, transcribed voice-over.
[27] Doane 1980, p. 42.

thought in your mind was that to exist is to devour oneself.''[28] The voices in the film recall the manners of a hypnotist. They attempt to put words and thoughts into the listener's head, with ''if you look at these images enough, you begin feeling like you composed them,'' directing the audience towards a certain feeling, and another sentence like ''these images posted at random are forgotten, have ways of permeating their surroundings, and thus effecting you even though you are barely conscious of perceiving them,'' tries to convince you that many circulating images unconsciously impact the audio-spectator.[29] It has the tendency to exert a certain power on who is listening, instructing and suggesting thoughts originating from the voices he or she might not have been thinking. But does the power of this hypnotizing voice *differ* from the way it has been theorized? How do Rafman's disembodied voices relate to similar sounding voices that we usually hear in the cinema?

According to Michel Chion the disembodied voice's/acousmêtre's powers are four: ubiquity (the ability to be everywhere), panopticism (the ability to see all), omniscience (the ability to know all) and omnipotence (to have complete power).[30] Here, it is important to remind that Chion's theory is based on cinema and not specifically developed for the analysis of film works by contemporary artists.[31] The first characteristic of the acousmêtre, its ubiquity, corresponds partly with the way it acts in Rafman's film. Even though the voice in *Erysichthon* comes from a non-localized body (the way Chion describes this ubiquity) it is not transmitted through media like the telephone or radio, instruments that often serve as mediating factors that enable the voice to be everywhere and nowhere.[32] The sources of the voices in the artist's film are nowhere to be localized. The ubiquitous aspect of the voices more closely resembles the way as it is employed in documentary film, a film form wherein voices more regularly appear disembodied, as will be elaborated later in this thesis. In *Erysichthon* there is no carrier, no object that facilitates the transmission of sonic vibrations.

---

[28] Rafman, *Erysichthon*, 2015, transcribed voice-over.
[29] Rafman, *Erysichthon*, 2015, transcribed voice-over.
[30] Chion 1999, p. 24.
[31] Chion's 1999 book *The Voice in the Cinema* examines the voice in three feature films. The first being Fritz Lang's *Testament of Dr. Mabuse* (1933), the second Kenji Mizoguchi's *Sansho the Bailiff* (1954) and the third Alfred Hitchcock's *Psycho* (1960), discussing the masculine, hidden and faceless voice, the female voice of the mother and the androgynous voice respectively.
[32] Chion 1999, p. 24. Of course in feature films disembodied voices do not necessarily need an on-screen mediating factor but Chion, in his discussion of the acousmêtre, also examines disembodied voices that still have some sort of non-bodily and localizable source. This is not applicable to the voices in my case studies.

The second characteristic is panopticism, a term derived from the institutional control system designed by philosopher Jeremy Bentham. Panopticism includes the total mastery of space by vision, and in cinema this mastery is often found in the form of a narrating voice-over, telephone voices featured in thrillers or invisible ghostlike figures whose vision can't be traced[33]. In Rafman's film it is suggested that the speaking voices are aware of what is presented on the screen. For example, when the disembodied voice utters ''if you look at these images enough…'', it knows that images are being presented to the audience. However, the voice is not always present to comment on the images that are shown and the greater part of the sentences that are spoken do not directly bear a relation to what is shown. Chion calls this acousmêtre who does not see all the exception to the rule. We are not left with the feeling that the voices in *Erysichthon* are watching with us or looking over our shoulder. At times they comment on what we see, but they also utter more poetic phrases left to linger on the images and they often keep silent to let the images speak. The voices do not display a total mastery over what the audio-spectator experiences. More specifically, it sometimes arrives as a Voice of God but also has the tendency to murmur poetic phrases or disappear. It does not completely control the space of the film.

The third characteristic of the acousmêtre, its omniscience, becomes problematic in this light, with the voice acting more or less like a part-time Voice of God. Surely, traditional gods often see and know everything (with exceptions reserved for the gods of ancient Greece), but the god resembling voice in *Erysichthon* has limitations. On the voice with limited knowledge Chion writes that it could appear much more disconcerting than an all knowing being.[34] The reason for this is that the audience does not know the limits of its knowledge. Rather, in *Erysichthon* disconcert mostly arises not from limited knowledge but from incomprehensibility. The voices speak in an open and poetic manner, making its meaning difficult to deduce. The last power, one that builds on the first three powers, is omnipotence. But as discussed, the voices' not ever-present, not all-seeing and not all-knowing characteristics do not follow towards complete control. There is probably only one form of power, hypnotic power, that I will elaborate on later in this chapter, in regard to the voice's synthesized form.

However, if *Erysichthon*'s voices do not exert complete power over the direction of the audio-spectator's thoughts, as seen in sentences like ''a moment of complicity and intimacy,

---

[33] Chion 1999, p. 24.
[34] Chion 1999, p. 26.

holding hands with death, a moment without hope and without an opening,'' where the audience is left to reflect on the meaning of this more poetic and less forceful phrasing, at the moments where the disembodied voices are addressing the audience or commenting on images, it very much resembles the use of voice in the expository mode of documentary filmmaking.[35] As theorized by Bill Nichols, the expository mode ''emphasizes verbal commentary and an argumentative logic.''[36] A commentary is directed towards the viewer, typically detached (disembodied) from the images and also presumed to be of a higher order than the accompanying images.[37] In Rafman's film this mode of filmmaking is apparent in a number of examples, firstly when the first few lines of the film are spoken. During these lines, in which a disembodied voice utters that the only thought in your, the audience's, mind is to devour oneself, a shot of a snake eating his own tail and therefore devouring itself is shown (fig.1.1.). Here, the image is clearly illustrative of the content transmitted by the audio. Another example is a sequence wherein a voice states that the images that follow each other successively in the film have ways of permeating their surroundings. In turn this is demonstrated visually through two hands taking photographs with a mobile phone. Every picture that is taken ends up, after the next one is captured, engulfing the remainder of the screen or, as the voice phrases it: permeating the surroundings (fig.1.2. and 1.3.).

But this expository style is only noticeable when the disembodied voices either turn towards the audience or the images. The voices are godlike when addressing something, they are not when they refer to themselves. An important difference with Nichols' expository mode is that even though the voices do address something in authoritarian manners, they do not always emanate an objective authority but also a subjective pondering as many open sentences are there to be interpreted. It often diverges from this documentary mode as they do not constantly pressure their views to the audience. The disembodied voice in documentary, as film scholar Michael Renov has argued, may often have been belittled, it can also be a vehicle of subjectivity.[38] When not performing in the expository mode, the voices mull over their thoughts and reveal they are not all-knowing but rather subjective and limited. The voices in *Erysichthon* have the tendency to seem all-knowing, suggest, instruct and speak *at* rather than speak *to,* but at the same time when performing more subjectively they invite the audio-spectator to contemplate. To finish, such contemplation is also made possible when the disembodied voices

---

[35] Rafman, *Erysichthon*, 2015, transcribed voice-over.
[36] Nichols 2001, p. 33.
[37] Nichols 2001, p. 107.
[38] Renov 2004, xxi.

are silent, something that also happens often, leaving the audience with non-vocal sounds, such as music, and imagery.

**The Male And Female Disembodied Voice In *Erysichthon***

The two voices in *Erysichthon* are a male and a female one. As mentioned earlier, the voice has been a topic of discussion in the study of documentary film and the nature of this knowledgeable, privileged and unquestioned interrogator has been for the most part that of the male.[39] I have not encountered any publications on male and female voices in documentary film. However, in the study of the acted feature film key publications have been published on sexual differences between the female and the male voice, of which some will be applied here. Art historian Kaja Silverman states that ''at its most crudely dichotomous, Hollywood pits the disembodied male voice against the synchronized female voice.'' According to her, disembodied male voices in narrative film usually occupy an anonymous and invisible position close to the recording apparatus, a position of power, whilst the female voice often functions as a fetish, becoming identified with ''spectacle and the body,'' which is visible and powerless.[40] What does this imply for my study, one that is solely preoccupied with disembodied voices?

It seems that the disembodied female voice *an sich* is already the exception to the rule, a position where the woman escapes the visual scrutiny of the male gaze. Film scholar Sarah Kozloff has written extensively on voice-over narration in feature films and states that she only has encountered two never-seen female narrators in American fiction film, leading her to argue that ''the barriers against women serving as third-person narrators in feature films have been so many and so high that their thorough exclusion ultimately seems overdetermined.'' Kozloff employs the term 'third-person narrators' to describe the disembodied voices in narrative feature film. She continues that ''if a woman were to serve as a third-person narrator, not only would she be allowed dominion over the public sphere as opposed to private, not only would she potentially wield great power and authority, not only would she speak as the film's image-maker,

---

[39] Doane 1980, p. 42.
[40] Silverman 1988, p. 39. Kaja Silverman was one of the first scholars to notice that theoretical attention to film sound on the representation of women was lacking, whereas the same subject had been studied with regard to the image track. Topics like the male gaze did not remain unseen, while the operations of the female voice in film remained unheard for the greater part of the twentieth century.

but she would escape being objectified or eroticized"[41] Similarly film scholar Amy Lawrence posits that "the authorial voice is rarely heard as a *woman's* voice in classical cinema."[42] However it is important to note that Kozloff's and Lawrence's study date from 1989 and 1991 respectively, making it impossible to cover more recent cases of the female disembodied voice. Moreover, the object of study in these publications primarily concerns classical Hollywood cinema, not contemporary artist film.

In *Erysichthon* the audio-spectator is listening to two disembodied voices, a male one and a female one. In what has been described earlier we have seen that the voices in the film have a suggesting nature as well as a poetic one. In the appendix I outline how much is being spoken in Rafman's eight minute film and define the characteristics of the two disembodied voices in order to give a better understanding of their presence and dominance (the transcriptions of the upcoming case studies can also be found in the appendix). The female voice is responsible for not only the major part of the text spoken but also for uttering the most sentences that have a voice of god-like nature. For example, the female voice has nineteen lines whereas the male has only seven, furthermore the female voice addresses the audience three times in an authorial manner whilst the male does this twice, lastly the woman utters four poetic sequences and the man two. Concludingly, there is a certain dominance that is emitted from the female voice in this film, since she speaks more and especially instructs more. As a result the authorial voice in *Erysichthon* is not male but female. However, as we shall see later, this authority is in turn subverted by its sonic manipulation.


**Interpreting The Non-verbal in *Delivery to the Following Recipient Failed Permanently***

As the words that are transmitted by a disembodied voice can be experienced as something divine and directing, what do the vocal sounds that do not employ words to communicate tell us? Aspects like the quality of the voice, its demeanor, also have an impact on the transmission of meaning. Furthermore, non-verbal sounds not originating from the disembodied voice but juxtaposed to it can negotiate with it, questioning the voice's *verbocentrism* (not its *vococentrism*). Lastly, anyone who has ever transcribed colloquial speech into words knows that humans generally do not speak in fluent sentences, rather they stutter, pause, reformulate,

---

[41] Kozloff 1989, p. 101.
[42] Lawrence 1991, p. 169.

cough, stumble over their words, etc., and there is meaning to be found in these non-verbal phenomena, beyond the vibrations of articulated speech. Ed Atkins's 2011 *Delivery to the Following Recipient Failed Permanently* contains cases of the non-verbal that negotiate with the disembodied voice, problematizing its common powers. To what extent do they suggest, instruct or control and do they account for an authoritative quality?

Atkins's one channel digital film *Delivery to the Following Recipient Failed Permanently* features a digitally altered voice-over expanding on the word 'smoke' over images of the silhouette of the back of a person's head in the dark, black and white backgrounds and clouds of green smoke. The seventeen minute film is full of sudden cuts. Visually it can abruptly change from a black image, to a white image and subsequently to a short insert of a digitally rendered human nose and mouth (used in the cover page of this thesis). The images most recurrent are the ones of white or green smoke moving either swiftly of slowly over a dark background. Added to these images is a disembodied and synthesized voice uttering poetically and essay-like on the word 'smoke', or, as the voice tells the audience of its intent: ''I want to make you aware of my mouth, to map my mouth, comprehensively using the word 'smoke'.''[43] This delivery of the disembodied voice sounds as if it takes the center stage since the film has a clear beginning and ending, marked by the voice starting its monologue and ending with it. Also, it seems that the voice is almost always present, constantly delivering a lengthy text during the film's full duration (see appendix). But this voice is accompanied by myriad of human non-verbal communicative sounds, sound effects and music that are regularly inserted like jump cuts (appearing and leaving without clear notice).

According to film scholar Robert Stam the theory of Russian semiotician Mikhail Bakhtin offers a conceptual tool to deal with questions concerning all kinds of exchanges and interactions within a film and between a film and its audio-spectator named *taktichnost*, literally meaning 'speech tact'. Stam states that ''the notion of 'tact' is extremely suggestive for film theory and analysis, applying literally to the verbal exchanges within (...) as well as to the 'dialogue' between film and spectator.''[44] He goes on by explaining that tact evokes power relations between audience and film, being able to suggest phenomena like intimacy and distance or camaraderie and domination. However Stam does not elaborate much on how this 'speech tact' is to be analyzed in film. Furthermore, Bakhtin did not specifically theorize the notion for studying film. Luckily in the writing of film scholar Germain Lacasse I have found a

---

[43] Atkins, *Delivery to the Following Recipient Failed Permanently*, 2011, transcribed voice-over.
[44] Stam et al., 1992, p. 224.

useful explanation of Bakhtin's concept applied to film theory. As stated by Lacasse *taktichnost* ''refers to every non-verbal but important aspect of communication: tone, attitude, gestures, codes and physical mannerisms.''[45]

In the following I will discuss the *tone* and *attitude* of Atkins's disembodied voice and its position towards other sounds in a selection of sequences that occur early in the film, while gradually clearing the ground for the analysis of the voice's synthesized aspect.[46] Considering the *tone* of voice in *Delivery*, Atkins himself is the original source of the voice that is audible in the film and his way of speaking is reflected in the disembodied voice. When listening to interviews with Atkins aspects of his vocal tone are reflected in his characters (Atkins voices almost all of his films himself). The artist is well articulated in British English, the standard dialect of English that conjures stereotypes of intellect, and employs a wide vocabulary of difficult and hard pronounceable words. Atkins's speech is hardly disfluent, especially in his films, giving the voice an even more well-articulated air. Being disembodied all the more focus lies upon how Atkins's voice sounds since there aren't many contextualizing factors present for the audience. Stam argues that ''in the sound film, we not only hear the words, with their accent and intonation, but we also witness the facial or corporeal expression that accompanies the words'' that are evoked by 'tact'.[47] In the case of *Delivery* however these facial and corporeal expressions do not partake in this game of tact. The consequence is that the importance of verbal tact in the question of power is augmented, as the tone of Atkins's voice is a dominating force and not contesting with the tact of a visible body.

Considering the voice's *attitude*, initially it seems Atkins's eloquent and manly voice dominates over all other sounds on the auditive track. However when listening closely the disembodied voice also has to give way to other sonic phenomena, and is sometimes even interrupted by them, giving it a flexible attitude. Already early in the film the voice (at 02:41, fig.1.4, 1.5 and 1.6.) is cut off by a loud and sudden burst of sound, after which a ringing noise (reminiscent of the sound one hears when suffering from noise-induced hearing loss), together with other tones, gradually gains in volume and depth only to be rewinded and followed up by the sound of a reversed cymbal. Only after this sonic sequence the disembodied voice continues its survey of the word 'smoke', as if the voice was waiting for the other sounds to end.

---

[45] Lacasse 2012, p. 491.
[46] I omit *gestures* and *physical mannerisms* out of this analysis since they are not of importance in the context of disembodied voices. What Lacasse means with *codes* remains unclear to me. Does he refer to codes in a semiotic sense, etc.?
[47] Stam 1989, p. 46.

This is also the case when the disembodied voice negotiates with the film's music, sometimes it refrains from speaking when music is heard, but the voice can also be silenced by it. At other times voice and non-vocal sounds work together, as audible when the word 'smoke' is uttered and the sound of a lighter is heard (04:47) and at another time a smoke machine (06:33) (even though smoke itself is a mute phenomenon, its form is triggered by the sound of its cause). Cooperation is also present when for example music is being used to set a mood for the narrative uttered by the voice. Lastly, it also occurs that voice and the other sounds seem to work independently from each other: while the disembodied voices speaks, music, foley work and noise are sometimes injected like aural jump cuts, however, it is as if the voice and these ruptures (jump cuts generally account for abrupt transitions in a film's visuals) do not notice each other since the voice continues speaking and the sonic cuts keep on appearing, normalizing a disrupted rhythm. Altogether the attitude of the disembodied voice is not as dominating as it might sound when primarily considering its tone. Atkins's voice gives way to, collaborates with, disregards and hegemonizes the other sounds, adopting various attitudes.

Also, the general tone of the disembodied voice is contrasted by its use in the beginning of the film, wherein it is heard breathing, hesitating and wetting its mouth. The voice is heard uttering 'uhhh' several times before starting its eloquent sixteen-minute speech. This hesitation, reminiscent of other non-lexical fillers like 'erm' 'um' and 'huh', indicates either a certain disfluency in speech, or a preparation of the mouth before speaking. Either way, the film has a disfluent start and resembles articulation problems such as stuttering. According to linguist Steven Connor "stuttering has sometimes been thought of as a kind of alienation from the human…"[48] It is exactly this *alienation* that I address in the following part of this chapter on synthesized speech.


**Power After Synthesization: The Exaggerated Voice And The Voice That Has Seen Too Much**


The voices in the two previous case studies are not only disembodied but also subject to a certain synthesization. The voice in Atkins's work is not a clear human voice, by digital means the pitch of the voice has been lowered. On the other hand the disembodied voices in Rafman's film sounds as if it has been created through speech synthesis, the artificial production of

---

[48] Connor 2014, p. 29.

human speech (however I can't firmly state whether the voice is artificially created or manipulated digitally in order to sound synthesized), creating a mechanical voice resembling the vocal sounds of robots as made audible in popular culture, such as HAL-9000 in *2001*. Film scholar Pamela Robertson Wojcik reminds us that a cinematic performance, also one that omits a visible body, is not only constructed by actor labor but also by sound design, because it is also shaped by microphones, recording methods and, especially in my case studies, the following manipulation in the sonic mix.[49] Firstly I discuss how the vocal transformation of the voice in *Delivery* influences and falters its position of power, and subsequently examine the effect and consequences of the robotic quality of *Erysichthon*'s speakers.

Atkins's disembodied voice has been digitally altered to have a decreased pitch, in other words, it sounds lower, making it sound deeper. A voice with a relatively low frequency is sex-typical for men, however Atkins's voice has been lowered in such a way that it would be improbable to connect it to a living person. On the one hand it is recognizably human, on the other it will feel unnatural to perceive it synchronous with a speaking person. It has been shown by cognitive scientist Katarzyna Pisanski "that humans can spontaneously manipulate vocal frequencies to deemphasize or accentuate various biosocially relevant traits with meaningful variation across social contexts, within the limits imposed by various anatomical or mechanistic constraints on vocal production."[50] In *Delivery* there is no mechanistic restraint, the artist has attached to the disembodied voice an exaggerated quality, sonically overemphasizing its manliness. Atkins himself has stated that his character's words "seem affectionate, caring– but which are in fact exploitative tropes conveyed with a cool impunity to anyone who'll listen."[51] This delivery characterized by a cool impunity infers that the voice somehow goes unpunished, that it has a responsibility it is not taking. How is this excessively male voice evading responsibility?

The voice in *Delivery* actually resembles the transformation of the voice by a technique used in hip hop music named *chopped and screwed*. This style is characterized by slowing down the tempo of a beat, creating a sedated motion (it was designed to complement drowsy drug highs) which can communicate a number of things: calm, leisure, grogginess, chromaticism, and an almost numbing sense of ease.[52] This comfortable sonic environment of slowness in conjunction with a disembodied and deep male voice creates a hypnotized state in

---

[49] Wojcik 2006, p. 73.
[50] Pisanski 2016, p. 307.
[51] ''Ed Atkins: 'Delivery to the Following Recipient Failed Permanently' (2011)'', Frieze, 2011.
[52] Pearce 2017.

the audio-spectator, and just as in *Erysichthon*, the voice in *Delivery* speaks *at* the audience but not *with*, leaving no space for dialogue. As with hypnosis, the voice suggests and instructs in a comforting aural environment (visually this state is also induced, especially through its continuous focus on a particular object: the opaque back of the head of an unknown person).

The result is a detachment and alienation from the space in which the audio-spectator perceives the film work. The synthesized voice created by Atkins makes the audience less aware of their surroundings and creeps, as the voice argues, ''inside your brain,'' in order to control.[53] Atkins's voice beguiles, evading a responsibility to be criticized, monologues instead of dialogues, and exaggerates the common disembodied male voice, amplifying its effects but also, by making him at times obedient to other filmic components and disfluent in speech, falters its position of power.

As opposed to Atkins's film, *Erysichthon* features a disembodied voice that sounds more mechanical and more akin to the popular sonic portrayal of robot voices. Characteristic to the sound of these voices stands the vocoder, an instrument designed in 1928 to alter a voice's pitch and frequency in order to encrypt speech for US war efforts. Next to quick transmission (vocoded voices basically exist of less data making them easier to transmit) one of the characteristics of the vocoder's military use was the exclusion of voice recognition.[54] This quality of the voice to sound impersonal turned to the attention of artists and musicians like Laurie Anderson claiming it to be authoritarian, robotic and corporate, or the German band Kraftwerk, employing the sound to produce an alienating effect.[55]

Two years after HAL-9000 the vocoder gave voice to the supercomputer Colossus in Joseph Sargent's film *Colossus: The Forbin Project* (1970)[56]. The Colossus's voice became one of power and paranoia, since the characters in the film let the computer control the U.S. Defense Department's complete nuclear arsenal. According to musician Wendy Carlos ''Colossus is a much more frightening talking computer than HAL, which may be partially due to

---

[53] Atkins, *Delivery to the Following Recipient Failed Permanently*, 2011, transcribed voice-over.
[54] Thompkins 2010.
[55] ''The Secret History of the Vocoder'', The New Yorker, August 20 2014, <https://www.youtube.com/watch?v=OvR4qK0B--w>.
[56] The vocoder as an aid to produce the voices of machines only started in the sixties and seventies. The voices of robots and computers like Robby the Robot in Fred Wilcox's *The Forbidden Planet* (1956) or Alpha 60 in Jean-Luc Godard's *Alphaville: A Strange Adventure of Lemmy Caution* (1965) were not created with the vocoder technology.

the vocoder."[57] However, these films are examples of *embodied* synthesized voices, the case studies in this study contain *disembodied* ones  unconnected to an alien form.

It is interesting to note that upon hearing a disembodied synthesized voice an audience will think that it will have a mechanical source, though the source might very well be Anderson, Kraftwerk or the voice actors in *2001* or *Colossus*, automatically attaching to the voice the qualities of computers even when there is no immediate proof. It seemed that the voice in Rafman's film had all the qualities of a powerful, authoritarian and fear instilling voice. However as we have seen in the preceding, the voice only partly functions as a voice of god, similarly, the common feeling of fear towards a voice that is normally embodied by an all-controlling supercomputer is not applicable to the voice in *Erysichthon*. Rather, our fear is not towards source of the voice, although it sounds inhuman and mechanical, and while its invisibility bestows it in a position of power. Our fear is for what it warns for and the mental state it embodies.

The male and the female voice in *Erysichthon* sound as if they originate from a machine. Their qualities are both human and inhuman. In writing on the mechanical woman Hadaly in Auguste Villier's speculative fiction novel *The Future Eve* (1886), film scholar Alain Boillat argues that even though the movement of machines animates *things*, "only the voice truly 'animates' the *beings* represented."[58] The disembodied voices in Rafman's film are therefore subject to a paradox, sonically they are both animate (the presence of a voice) as well as inanimate (the voice is not that of a living being). The synthesized character of the voice hints at the presence of something that is ordinarily suppressed while listening to a non-mediated human voice - it is the sound of the mechanism of the body, which logically must be mechanical, especially since there is no visual proof that the voice is from a *being*.[59] However, in the case of *Erysichthon* I will go one step further and define the disembodied and synthesized voices not just as inhuman and mechanical, but as *alien* and *disturbed*, positioning the audio-spectator ambiguously within the power relations of the film. Concerning the 'nature' of synthesized vibrations sound designer David Sonnenschein states that:

 "Synthesized sounds cover a large variety of sound waves that are not originally generated by physical vibration of air molecules, but rather begin as sounds with electronic analog or digital

---

[57] Thompkins 2010.
[58] Boillat 2010, p. 239.
[59] Chion 2009, p. 332.

sources. 1950s sci-fi films exploited the earlier analogue synthesizer's other-worldly effects, which worked so well because they obviously were not being produced by anything natural on Earth.''[60]

Synthesized sounds have the special quality of being perceived as alien or unnatural. There is one moment in the fifth minute of *Erysichthon* where the female voice starts to stall, loop and warp. The word that is repeated and emphasized during the voice's short deformation is 'fail', singling it out as the one word that is treated with more distortion, making Rafman's film (next to its obsession with self-devouring and falling into a void) even more grim. Sonnenschein argues that when one listens to someone in a recognizable language, the focus will be upon its signification. Contrarily, when listening to something alien, the focus tends to shift to the pure acoustical characteristics of a voice.[61]

In the case of *Erysichthon* the focus is on both signification and acoustics, the language is at the same time understandable and alien. It speaks in hypnotizing phrases, occasionally acts like a voice of god but also sounds tired and disturbed. One critic remarks that ''its piercing soundtrack and disturbing voiceover conjure a mind pushed to its limits, reiterating that it has seen too much.''[62] This tired and endured voice is in stark contrast to what we generally perceive as pleasant and ideal voices, since, according to linguist Steven Connor, ''our celebrations of the voice are too monotonously pitched in the register of fullness, richness, clarity and penetrativeness, the bay is too regularly accorded to the energetic out-loud and 'haute voix'.''[63] The disembodied and synthesized voices in the film are not full, rich and clear, neither are they sharp and loud.

The power of the voices are the result of their hypnotizing phrasings and disembodiedness, however, unlike a typical voice of god, *Erysichthon*'s ghosts do not see or know everything, they do not have complete power (unlike HAL or Colossus), and have the tendency to keep silent. The voices sound tired and fragile, subverting their position of power, leaving the audience to overthink the nature of the speaker and even creating a strange feeling

---

[60] Sonnenschein 2001, p. 42.
[61] Sonnenschein 2001, p. 137.
[62] Twerdy 2015.
[63] Connor 2014, p. 31.

of intimacy.[64] It might be dominant and suggesting at times, 'direct addressing' the audience (as Doane would put it), but it is its disturbed nature that is most powerful here.

The synthesization of the disembodied voices in both case studies make for a different value than the effects attributed to the disembodied voices in the theories of Chion and Doane. Yes, being disembodied, the voices in the films still make a voice of power heard. However, its sonic transformation has serious consequences for the value of the authority such voices commonly have. Their alien demeanor weakens their power, either by exaggeration and its subservient position within other audio components (*Delivery*) or by sonically revealing that the voice is used up and almost insane (*Erysichthon*). The voices do not dictate everything, and deviate from the powers stated in theoretical framework. In the following chapter another common conception of the function of sound will be scrutinized: the notion of sound as a harmonizing agent.

---

[64] Kozloff 1988, p. 50. Kozloff's *Invisible Storytellers* posits that voice-over narration in fiction film is well suited to create intimacy through direct address. According to her, when a voice addresses the audience ''it implicates us in their world and value system.'' *Erysichthon*'s fragile sounding voice achieves a similar effect.

**Chapter 2**: **The Fractured Voice: Finding Value In Disharmony**

This chapter addresses the question of unity and disunity. The voice becomes less isolated and will be analyzed in consideration to the other components of the audiovisual whole in the films *Dust and Piranhas* by Helen Marten, *Woods of Arcady* by Jon Rafman and *How Not To Be Seen: A Fucking Didactic Educational .MOV File* by Hito Steyerl. The question of this chapter emanates from the observation by film scholar Kathryn Kalinak that multiple film sound theorists state that the function of sound (and especially music) is to create audiovisual unity within a film. [65] The purpose of this chapter is to show its disunifying qualities. To what extent does the disembodied and synthesized voice create rupture in a film's unity and what are the consequences for the notion of sound as harmonizing agent in the context of the case studies?

This chapter will start with a discussion of sound film's fundamental lie, a lie film scholar Rick Altman perceives as pivotal in understanding sound film and reflects on the consequences of this lie for disembodied and synthesized voices. Subsequently three topics in the context of the three case studies of the chapter are explored. First, I discuss Michel Chion's notion of the 'real and the rendered' with regard to the 2011 film *Dust and Piranhas* by Turner Prize winning British artist Helen Marten to investigate what the disembodied voices in the film render. Second, I explore how the synthesized voice can function as a counterpoint device, a term made famous by Soviet director Sergei Eisenstein, in Jon Rafman's 2010 film *Woods of Arcady*. Lastly, I stress how the synthesized voice can be considered as, following economic and social

---

[65] Kalinak 2010, p. 24. Kalinak provides an overview of a selection of film scholars who deal with questions of sound and music. For example, Claudia Gorbman states that the sound track has 'reconstructed time into a ''relentless linearity'' with music being ''the one sound element capable of freeing up that temporal representation.'' This explains why music is called upon to attend film's most fractious moments in terms of time - flashbacks, montages, and slow-motion sequences - that threaten the unity of the film.' Furthermore, Soviet director Sergei Eisenstein states in his publication *The Film Sense* that the possibility of hearing opens up the possibility of creating a single, unifying sound-picture image (Eisenstein 1957, p. 73.). Similarly, composer Aaron Copland argued that one of the functions of film music is ''building a sense of continuity,'' glueing the images together (Aaron Copland, 'Tip to the Moviegoers: Take off Those Ear-Muffs', in *The New York Times*, 6 November 1949). Lastly, Chion also writes that the most widespread function of film sound is to create an unifying sound bath, since: First, in temporal terms, it unifies by bridging the visual breaks through sound overlaps. Second, it brings unity by establishing atmosphere as a framework that seems to contain the image, a ''heard space'' in which the ''seen'' bathes. And third, sound can provide unity through nondiegetic music: because this music is independent of the notion of real time and space it can cast the images into a homogenizing bath or current'' (Chion 1994, p. 47).

theorist Jacques Attali's concept, 'noise' in the film *How Not To Be Seen: A Fucking Didactic Educational .MOV File* by German artist, filmmaker and lecturer Hito Steyerl.

**The Illusion of Unity**

According to Rick Altman the sound film is based in a fundamental lie: "the implication that the sound is produced by the image when in fact it remains independent from it."[66] The sound we hear when experiencing a film event does not originate from the image. It is a persuasive illusion that upon watching on-screen people talking we perceive them as speaking real words. With the exposition of this lie Altman famously challenges the idea that sound is subservient to the image. He compares the sound film to ventriloquism, with the sound being the ventriloquist who, "by moving his dummy (the image) in time with the words he secretly speaks, creates the illusion that the words are produced by the dummy/image whereas in fact the dummy/image is actually created in order to disguise the source of the sound."[67] By some filmmakers sounds are synchronized to the image in order to hide their actual separation, others play with this appearance of reality to "reveal the inherent disjunction between sound and image."[68] The examples accompanying this argument usually exist of discussing how images and sounds occurring simultaneously fuse, for example making a voice embodied. However, do disembodied voices not achieve a similar kind of fusion? Or is this fundamental rupture of the sound film emphasized when hearing a disembodied and synthesized voice?

      In the case of disembodied voices, it is clear that a mouth and a voice will not occur simultaneously (of course, there will always be exceptions). This fusion or *synchresis*, as Michel Chion terms it, is not applicable in our case studies. This means that the concept of *added value* is also functioning differently as "added value is especially at work in the case of sound/image synchronism, via the principle of *synchresis*."[69] Of course, *added value* is still at work when no easy sound/image combinations are there to be made, such as attaching a voice to a mouth, an splash to a something dropping in the water or a chirp to a bird, it is also at work when thinking

---

[66] Altman 1980, p. 6. It needs to be stressed that, as mentioned in the introduction, there is an important distinction between analogue and digital film and the medium of video. In video sound and image are not fundamentally separate as in film.
[67] Altman 1980, p. 67.
[68] Beck 2016, p. 31.
[69] Chion 1994, p. 5.

less straight, for example when attaching the sound of frying bacon to rain, the sound of a lion's roar to a revving car engine or the sound of traffic to a cow.[70] In the case of the disembodied voice there is a split between the voice and the body since there is no physical and visual body to attach the voice to.

The sound film's fundamental illusion is at risk of becoming more exposed due to the presence of disembodied voices. However, even though there is a split between the acousmatic voice and the body, the unity of a film is not necessarily lost. According to Doane the loss of a body to attach the voice to can be reconstituted by the technology and practices of the cinema in the form of a *fantasmatic body,* ''which offers a support as well as a point of identification for the subject addressed by the film.''[71] Although it is hard to derive a clear definition of the term *fantasmatic*, Doane does not deliver her full definition, it is widely used in psychoanalysis. Most relevant here is to state that the fantasmatic is a ''particular network of fantasy and desire that subtends and motivates (...) manifestations of the unconscious.''[72] The fantasmatic body in film is incorporeal and desirable. Doane posits that one attribute of the fantasmatic body is to bring unity to a film. Even when an union seems displaced (as would be the case with disembodied voices), the ideology of organic unity will be at work, as ''combinations of sound and image are described in terms of 'totality' and the 'organic' (...) and sound is 'married' to the image.''

Also, as one sound engineer puts it, ''one of the basic goals of the motion picture industry is to make the screen look alive in the eyes of the audience…''[73] With the development of digital surround sound (DSS) the possibilities for fuller and life-like representations are amplified since here sounds can be ''carefully placed in the multi-channel soundscape to ensure that the aural environment created matches the visual one suggested by the onscreen images.''[74] Furthermore, DSS is aimed at diminishing the noise of the system, concealing any unwanted sounds that are not part of the film's representation, diminishing the sonics of the exhibitionary situation. Although it is difficult to trace, it is very likely that the films by Marten, Rafman and Steyerl employ DSS since they are defined as digital animation, digital video and HD video respectively (digitally produced films are often created with products like Adobe Premiere Pro

---

[70] Kisner 2015, *Guardian*. In fact all these sound/image combinations work and have been used, even if they sound as they are impossible to connect. A well-known Hollywood sound editor Skip Lievsay has arranged these sounds to sound natural to the image. According to him people hear what they are conditioned to hear, not what they are actually hearing.
[71] Doane 1980, p. 33-34.
[72] Brett Farmer 2000, p. 64.
[73] Doane 1980, p. 35.
[74] Kerins 2011, p. 265.

CC or Apple Final Cut Pro X, software that offers the possibility of professional surround sound mixing). However, I posit that even in our age of more than life-like simulacra unity and harmony are challenged by the strange sounding disembodied voices in our case studies.

**The Real and the Rendered Sensation in *Dust and Piranhas***

Helen Marten's five channel digital film *Dust and Piranhas* (2011) is hard to interpret (the film is projected on five-screens simultaneously). The twenty five minute film features CGI animated columns moving in single colour environments with CGI objects such as plants, furniture or ancient Greek pottery. Over these images multiple voices speak, some clearly meant to be embodied and others more unclear, on a mostly silent soundtrack, and sometimes rhythmically rhyming along with an inserted hip hop beat. A quote from the film contains an important cue to its interpretation: ''how to protect beauty from dust and piranhas, how to promote beauty to dust and piranhas?''[75] The film seems to be concerned with design, style and aesthetics and basically questions what beauty should be forgotten and left to dust or even devoured and left to piranhas and what should not. Even though *Dust and Piranhas* does not translate to a clear and identifiable meaning, Marten takes an interest in the sculptural aspects of digital animation and also creates digital renderings of popular or fashionable imagery and products. As mentioned earlier, the film features CGI animated columns, one of which argues that high-end technology (such as guns or luxurious cars) is only appreciated for being high-end and not for its function. Subsequently the argument spreads and starts being applied to refrigerators, televisions, chairs, cupboards, coffee tables, lamps, forks, the lemon squeezer, etc. The column declares that the function of the appliance is not there to be used, rather the appliance will resonate on the owner and support personal importance, for  he or she owns a technologically perfect product. But as the film is quite lengthy the complete content uttered by the voices will not be the center of focus. However the focus of this section will lie on selected sequences.

Marten's film features a scripted text delivered by several voices. Firstly, a female and male one, both disembodied but not synthesized. Secondly, two voices, of which one is synthesized, that are synchronized to a CGI animated doric column with eyes and eyebrows (it has no mouth) and an old stone sculpture with a face (fig. 2.1 and 2.2.). The column claims the disembodied voice to itself by moving itself when the voice is speaking and the sculpture

---

[75] Marten, *Dust and Piranhas*, 2011

asserts ownership by having its mouth animated, moving along with the speech of the disembodied voice. Both objects function as the puppets of a ventriloquist, claiming an otherwise hidden or sourceless voice to be their own. The sculpture's voice is multiple: when it speaks both a female and a pitched down male voice are heard simultaneously uttering the same phrases. Lastly, there is a sequence that contains two synthesized and disembodied voices, again male and female, monologuing, either by taking turns speaking or by speaking simultaneously. It is this last sequence that will be the focus of my discussion.

The sequence I analyze concerns a pondering by multiple synthesized and disembodied voices on the concept of glamour. The sequence opens by a female disembodied voice and a male disembodied and pitched-down voice simultaneously proclaiming their "treaties on gorgeousness", after which they start uttering reflexive phrases on glamour, either simultaneously or consecutively, frequently alternating between non-synthesized and disembodied and all sorts of manipulated voices .[76] These reflections are played over a digital animation of the doric column laying (seemingly) dead on a tombstone while a faceless wooden ionic column with Mickey Mouse gloves glides by, entering the frame from the left (fig. 2.3.**)**. The ionic column bends over the tomb, then starts circling to the right side of the frame before leaving it completely (fig. 2.4 and 2.5.). After that, a blue light shines on the doric column and erects it, raising it from its stasis (fig. 2.6.).

With regard to this sequence I will discuss the notion of the real and the rendered, coined by Chion. According to him, film sound is not used to reproduce a sound and deliver a better simulacrum by using sounds that have a high degree of fidelity to an original sound. In contrast, film sound, as Chion posits, *renders* and instead translates "the physical, psychological, even metaphysical impact of the action on the sender or the receiver," and it can "give us a vast array of luminous, spatial, thermal, and tactile sensations that extend far beyond realist reproduction."[77] The idea that a recorded sound is a reproduction is false, Rick Altman has called this the 'reproductive fallacy', as each recording is an interpretation of the original sound.[78] According to Chion at the basis of rendering stands *synchresis*, as only when a sound and an image occur in sync will the sound be able to render something. But what are the

---

[76] Marten, *Dust and Piranhas*, 2011, transcribed voice-over.
[77] Chion 2009, p. 239 - 240.
[78] Altman 1992, p. 40. Circumstances like the "recording location, microphone type, recording system, post production manipulation, storage medium, playback arrangement and playback locations" will necessarily distinguish the recorded sound from the original.

consequences for the disembodied voices? Since these voices are not synchronous to anything visual, what do they render?

In order to examine what the disembodied and synthesized voice in *Dust and Piranhas* renders I turn to a term Chion also uses in his discussion of the real and the rendered: materializing sound indices (MSI's in short). Chion explains that ''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.''[79] A human voice that speaks with many mouth or breathing noises or throat clearing will have a high MSI because it reminds the listener that it issues from a body, its physical source. The voices in the sequence of Marten's film do not make any of these noises audible and complicate a connection to a human body. Usually, ''voice-over commentaries are most often kept dematerialized, and at every stage of their reading, recording, and editing, sound technicians are careful to suppress MSI's that might leak in, with the stated aim of deflecting attention from the physical person behind the voice.''[80]

However, in *Dust and Piranhas* the attention *is* somehow directed towards some of the bodies of the voices heard in the sequence and this is due to their synthesized qualities. In the second half of the spoken treaty a female voice is manipulated to sound with a mechanical hoarseness and a higher pitch. Even though this voice has a very low MSI, it makes the voice sound more human, especially since it is juxtaposed with a male disembodied voice that is reminiscent of robot voices. The female voice sounds at unease and at times a little hysteric, qualities that usually make it sound emotional and thus human, even though its synthesized form eliminates reference to a non-mechanical body. Furthermore, the rhythmic structure of the lines spoken, or the flow of speech, also hints at a human source as it is not characterized by the more monotonous flow of synthesized speech such as HAL-9000's or the Colossus's, rather it has a large tone reach making its speech flow sound more close to home.

The variety of voices audible in the sequence generally have low MSI's but still manage to render sensations. This is achieved, not through the presence of human noises but of synthesized ones, giving the audio-spectator various sensations by rendering in different ways. Marten has stated on the process of visual rendering: ''in creating the final gloss of CGI, what you're doing is layering skins and light in an algorithmic jumble that is actually incredibly fleshy.''[81] I argue that the same is the case for her voices, its synthesized aspect is a certain gloss,

---

[79] Chion 1994, p. 114.
[80] Chion 2009, p. 245.
[81] Marten quoted in: Cattelan 2012, p. 54.

layered in artificiality, but somehow renders something 'fleshy', bringing about a tactile sensation or impact on the audio-spectator. In *Dust and Piranhas* many sensations are rendered.

For example, the pitched-down and exaggeratingly deep voice that speaks ''should the dance be arresting, or is it more glamorous when seen and not heard?'' renders a hypnotizing and calming sensation (reminiscent of Atkins's use of the voice in *Delivery*). The simultaneously speaking voices uttering ''should we glam into that content or vice versa?'' render contrasting feelings, with the male one sounding at ease and the female one fidgety (it is pitched-up and made to sound restless). Subsequently, a single female voice starts asking ''is it glamorous to be glamorous?'' making the music stop for a moment (during the complete sequence music is playing), framing it as the sole audible sound for a short while and therefore attracting attention to its hysteric demeanor (this voice sounding even more restless than the other female voice).[82] Finally, a robot like voice claims the stage and pronounces ''there are no glamorous people, no glamorous events, we know glamour is artificial'', with its mechanical quality also referring to something artificial.[83]

The variety of synthesized effects laid upon the disembodied voices in the sequence render a composite of sensations. There is no reproduction of something real, the original recorded voices have been reinterpreted and artificially transformed to render something completely different than the original voices would have. Although the content uttered by the disembodied voices comes across as cohesive speech, the sensations that are rendered do not give the sequence an 'organic' quality, on the contrary, it is as if the film suffers from multiple personality disorder. In *Dust and Piranhas* there is no single *fantasmatic body* or unity to be found.[84] The disembodied voices complicate the definition of cinematic unity, the idea that sounds and images start out in 'natural harmony', still a common perspective in film.[85]

---

[82] Unlike Rafman's use of gendered voices in *Erysichthon*, Marten does not give the female voice subversive power. On the contrary, even though the female voice is exaggerated it does not raise specific questions on gender.

[83] Marten, *Dust and Piranhas*, 2011, transcribed voice-over.

[84] This is also apparent upon analyzing the other sounds in the sequence in relation to the onscreen visuals. There is only one sound that is synchronous to something occupying the visual track of the film (all the noises and music are non diegetic). It is emphasized that the two tracks are separate, for no important difference would arise if many sounds would play over the images with a few seconds delay. The exception comes from the animated mouth of the statue. In the style of a ventriloquist's puppet, Its mouth sluggishly moves along with the spoken text but this synchronization is far from smooth. Sound and image only match because of *synchresis*, but the actual relation is arbitrary.

[85] Chion 1994, p. 95.

**The Synthetic Voice as Counterpoint in *Woods of Arcady***

One of Rafman's early works *Woods of Arcady* (2010) mixes recordings of digitally created classical statuary, architecture and awe inspiring landscapes from the online virtual world Second Life with a synthesized rendition of the poem 'The Song of the Happy Shepherd' by Irish poet William Butler Yeats. The four minute long film plays an almost equal length musical composition, at its core only consisting of two chords, reminiscent of the synthesizer pieces of the Greek composer Vangelis. The film is presented as a harmonious arrangement, with its disembodied voices steadily delivering the poem's text, never quickening or slowing down its pace of speech, the expansive and reverberating tones of the music overlapping several images, and a seemingly ever-present sound of wind binding the flow of images.

We have seen that it is not uncommon to perceive film sound as a component that 'glues' everything together, as evident in theory on film music and film sound giving a voice to the fantasmatic body of film, thereby animating it. It is all the more remarkable then, that the introduction of sound technology in the late 1920s caused a reaction among the Soviet filmmakers Sergei Eisenstein, Vsevolod Pudovkin and Grigori Aleksandrov to embrace the new technology not for its unifying possibilities but for its use as *counterpoint*.[86] Originally, counterpoint is a musical term used to describe that one or more independent melodies are added above or below a given melody. According to the Soviet filmmakers, a contrapuntal use in film is characterized by a "distinct nonsynchronization with the visual images."[87] This definition is of course far from specific, as any disembodied voice in any film would be considered counterpoint in this view.

Based on Eisenstein's early montage theory film scholar Kristin Thompson has laid out five types of counterpoint devices and its functions, a schema that will support the analysis of the case study's contrapuntal aspects. One main type of this sound-image function is "perceptual roughening", a term that "implies that the film confronts the spectator with an unusual device which is difficult to perceive smoothly; the purpose (...) is to stimulate the

---

[86] The three filmmakers issued the text 'A Statement' in 1928 in which they were cautious to embrace the new technology of film sound. Perceiving it as a threat to their montage aesthetic, they would only embrace it as a component that has to serve the culture of montage and disliked the idea that sound would bring forth a naturalistic style of filmmaking, prioritizing the illusion over intellectual montage where filmic components are used in a conflicting and dialectical way to stimulate thought.
[87] Eisenstein 1928, in: Weis 1985, p. 84.

spectator to a more intense, active perception.''[88] However, I will argue that counterpoint can also be at work when a film work is presented smoothly, as in the case of *Woods of Arcady*. Counterpoint does not simply have to be based on the idea of a contradiction or audiovisual dissonance (we hear A, but should hear B), and theorists like Chion advocate the use of a *free counterpoint*. Even though it is still undertheorized, *free counterpoint* does not just indicate the use of wrong or inappropriate sounds, rather it can suggest and render something completely out of the box.[89] Following Thompson's five-type schema, I will conduct an in-depth discussion of the third type and function of counterpoint as laid out by Thompson: Conceptual Juxtaposition of Sound and Image.[90]

In *Woods of Arcady* there are three main filmic components that are always present, with the exception of the first and last sixteen seconds, when the film consists of an image of a digitally rendered ocean and either the introductory sound of the musical theme (in the film's opening) or the sound of the wind (in both the film's opening and ending). The first component is the disembodied and synthesized voices that speak for three and a half minutes of the film's four minute duration. The second is the film's music, as discussed in the above as a simple synthesizer piece, and the third are the images of ancient statues and buildings the artist has found in the online virtual world game Second Life, its cuts solely consisting of fade-in and outs (fig.2.7, 2.8 and 2.9.).

   Within the auditive track a subtle use of counterpoint exists. The voices and the music, almost always active and never discontinuous, seem to work together, with the music trapping the voices in a rhythm, creating a sonic flow with an internal logic. According to Chion, sounds and images with an internal logic, as opposed to external logic, appear to follow a flexible, organic process of development, variation and growth.[91] However, when listening closely to the

---

[88] Thompson 1980, in: Altman 1980, p. 120.
[89] Chion 1994, p. 39.
[90] Thompson 1980, in: Altman 1980, p. 121-138. The third type and function of counterpoint devices as schematized by Thompson is the most interesting for my case study. The first, 'manipulations of only the sound track' and the second 'manipulations of only the image track' are problematic because these types mostly encompass the simple sound and image contradictions we are not looking for, moreover in Rafman's film it is never the case that either only the auditive track or the visual track is being manipulated (both tracks are non-naturalistic and digitally created and are inherently manipulated). The fourth 'temporal disjunction', when the sound occurs at a different part of the plot time from the image, is too obvious for analysis since almost every sound in *Woods of Arcady* is non synchronous with the film's plot (and the film does not consist of a clear series of events). The disembodied voice speaks from another place and time and the music is non-diegetic.
[91] Chion 1994, p. 46.

sonic demeanor and content of the disembodied voices in comparison to the atmosphere generated by the music there is a case of free counterpoint.

The voices are characterized by a text-to-speech robotic sound, giving it a stiff rhythm, a flat intonation and no audible accents or emphasis on words or segments. This makes the voices delivering Yeats's poem sound indifferent to its content, unlike the voices in Marten's film. The content of the original poem 'The Song of the Happy Shepherd' appeals to an ideal lost paradise, Arcadia (actually an old region in the center of the Greek peninsula Peloponnesus), and, according to Anglicist Enrico Reggiani "attempts to come to terms by substituting for Arcadia the supremacy of art in the form of song or poetry, and the song of the shepherd can therefore be called 'happy' because it can offer an alternative to Arcadia."[92] Rafman, on the other hand, does not find this alternative in poetry as is audible in the indifferent and unemphatic recital of the poem. Rather this alternative can be found in the images he presents the viewer: it is in digital technology and, more specifically, online games like Second Life where contemporary access to Arcadia can be found. The promise of this lost paradise is augmented in the triumphant and reverb clad synthesizers of the film's music. However, the attitude of the disembodied voices also influences our reading of the digital Arcadian images. When the voice speaks that "The woods of Arcady are dead / and over is their antique joy," this transitoriness can also be attached to the paradise of Second Life, for its immersive social space offers very real promises for an alternative life.[93] In *Woods of Arcady* an idea (the recited poem) is juxtaposed with another idea (the images of the game) and a sensation (the music), creating not just a contradiction but a counterpoint that creates ambiguity. The manipulated voices are not appropriate, to the words, to the images and to the music, and act in *free counterpoint*.

---

[92] Reggiani 2012, p. 83.
[93] Meadows 2007, p. American author Mark Stephen Meadows has argued in the publication *I, Avatar: The Culture and Consequences of Having a Second Life* that an alternative personality in Avatar-based games, such as Second Life, can become a powerful force in a user's life, as "the alternative personality can become predominant and begin to take over the primary, daily one."

**The Sound of Text-to-Speech: the Voice as Noise in *How Not To Be Seen: A Fucking Didactic Educational .MOV File***

The title of the 2015 film *How Not To Be Seen: A Fucking Didactic Educational .MOV File* by Hito Steyerl is a reference to a sketch by British comedy series Monty Python's Flying Circus called *How Not To Be Seen* (1970). In the sketch a disembodied voice (later revealed to be the voice of John Cleese) claiming to speak for the British government on the importance of not being seen. The film is initially presented as a public service information film, but soon turns into violent joke wherein every person that becomes visible or whose location is traceable gets blown up. Here, the disembodied voice of Cleese has control over every scene as every time he discovers someone behind a bush, tent or house an explosion follows. In Steyerl's film the disembodied voice does not possess similar power, rather it has an explanatory function. *How Not To Be Seen: A Fucking Didactic Educational .MOV File* exists of five chapters that instruct the audio-spectator how not to be seen (fig. 2.10). As the artist observes, our contemporary situation is marked by the idea that ''we are visible all the time to certain modes of capture (...) and people might want to escape from this inordinate amount of surveillance.''[94]

Again, in this film the disembodied voice is typified by its text-to-speechlike (TTS) sound. Shortly mentioned in the previous paragraph, TTS refers to the ability of computers to read text aloud, translating digital text into sonic output. Upon comparing the text as delivered by the voice in *How Not To Be Seen* with a common TTS system as featured in Google Translate, it can be stated that the artist has used a similar system to recite the text she has written for the film.[95] To the human ear a TTS voice sounds uncanny, on the one hand the presence of a voice refers to life, to an animate being sourcing it, and on the other hand such a voice will sound

---

[94] ''Hito Steyerl - 'Being Invisible Can Be Deadly' | TateShots'', Tate Modern, May 17 2016, <https://www.youtube.com/watch?v=kKAKgrZZ_ww>.

[95] For the analysis I have inserted the complete transcription that is delivered by the disembodied voice in Google Translate's 'Listen' function. After inserting a text for translation, users of this function have the option to listen to the text through speech synthesis, the artificial production of human speech. When comparing the two voices there are important similarities to be found, such as intonation, rhythm and use of accent. There are three differences, firstly Steyerl's manual is explained by multiple voices, a male one and a female one. Secondly, the speed of the recital is slower than common TTS speech, especially the male one. Finally, the male voice has been slightly pitched down, resulting in, combined with its low delivery speed, the sonic unveiling of the mechanical system producing the speech. Its talking speed makes audible the mechanical quality of the voice, releasing cracks and noises not present in organic and human voices.

alien or disruptive. It is clear something is not right. In this light I compare the disembodied TTS voice in Steyerl's film to the concept of *noise* as studied by theorists like the economic and social theorist Jacques Attali, Caleb Kelly and Andy Birtwistle.

Attali defines noise as "a resonance that interferes with the audition of a message in the process of emission."[96] Noise always disturbs or interferes, and in extreme cases is capable of killing a transmission of information. Similarly, the Oxford dictionary also translates noise as "a sound, especially one that is loud or unpleasant or that causes disturbance."[97] As noise is also translatable to just 'sound', I stress that I take Attali's approach and the disruptive definition of 'noise' in my discussion. Considering this, it is remarkable that voice has been imagined by linguist Steven Connor as the antonym of noise. According to him "noise is anonymous, mechanical and meaningless; voice is personal, animate and expressive. Noise is accident, voice is intent. Noise has no importance, voice is full of portent."[98] But what happens when a voice contains noise, when it interferes, or more precisely, how exactly does it interfere in the audiovisual whole of *How Not To Be Seen*, and what are its consequences?

Steyerl's film is a playful but serious approach to the problem of contemporary hypervisibility. On the one hand it explains how resolution capture technology has evolved from the installation of resolution targets and grey scales in the California desert in the nineteen-fifties (used to calibrate aerial photographs and videos) to the contemporary pixel-base resolution charts (fig. 2.11 & 2.12). As the voices inform us that "resolution determines visibility" and that pixel-based resolution has a higher image quality than ever before, so in order not to "become invisible one has to become smaller or equal to one pixel."[99] Of course, such an advice does not translate easily to practical use. To resist pervasiveness and being visible at all times, instructions vary from temporary solutions such as hiding or camouflaging oneself to the outrageous such as owning an anti-paparazzi handbag or being a superhero. Appropriately, the source of the voices that offers this inconsistent set of instructions is also not visible. Its disembodied nature supports

---

[96] Attali 1977, p. 26.

[97] "Noise | Definition of noise in English", Oxford Dictionaries, accessed August 2 2018, <https://en.oxforddictionaries.com/definition/noise> It is remarkable that the publication *Sound: A Very Short Introduction,* also originating from the Oxford University Press, is very outspokenly negative on noise, stating that it must be ensured "that noise is universally regarded as the pollutant that it is" (Goldsmith, 2015, p. 123.). The author does not take into account that many cultural theorists see noise as something disruptive but therefore also as something that is capable of creating something new.

[98] Connor 2014, p. 7.

[99] Steyerl, *How Not To Be Seen: A Fucking Didactic Educational .MOV File*, 2015, transcribed voice-over.

the case it is trying to make. However, the fact that Steyerl has chosen a TTS voice to instruct the audio-spectator as opposed to one sourced from a human has the following consequences.

The TTS voice is disruptive, it contains noise. A lecturer on the theory of art, Caleb Kelly, argues that various manifestations of noise (such as the crack) are "a continuing and productive approach in the domain of contemporary digital production. We are, that is, in what Kim Cascone describes as the postdigital environment, in which deliberate errors are used to produce media that turns its back on the clean, methodical and calculated digital studio."[100] By opting for a TTS voice Steyerl attacks the identity of the voice. According to media scholar Andy Birtwistle "the synthetic nature of these sounds [electronic sounds, such as synthetic voices] gives them the potential of almost infinite duration, presenting a post-human/inhuman musical sound…"[101] The TTS voice, being completely artificial, is not just disembodied, it makes no reference to a body at all. The irony created here is the combination of non-human agents uttering in disfluent speech and deliberately left to sound 'noisy' that are engaged in the task of instructing human agents, the audience, to resist visibility and start struggling for their anonymity. In other words, we have to become just like the voices in *How Not To Be Seen*, not only invisible but also inhuman by transforming into digital matter, into something smaller than one pixel, something noisy.

The deliberate use of noise in the voices does not result in a more fuller representation, rather its interference as noise emphasizes the importance of what is considered as bad quality. The noise of the TTS voice is the sonic counterpart of low resolution in images. Both might not have naturalistic or harmonious ends, rather Steyerl has pointed quite specifically at the value of a low quality image in her essay 'In Defense of the Poor Image', as "it builds alliances as it travels, provokes translation or mistranslation, and creates new publics and debates."[102] These poor images are often without owner and unidentified, since they are hurled in the speed of digital circulation and this is exactly where its value lies, as something unidentified and unseen can not be controlled. Also applicable to the voice, disembodiedness and synthesization exemplifies how to resist hypervisibility, and this resistance is acted out through sound, either eliminating a voice's location or transforming it.

---

[100] Kelly 2009, p. 315. Steyerl does not only employ the voice as an instrument to disrupt and deliver an unpolished film product. This also manifests itself in the presence of an on-screen non-keyed green screen, a camera crew and the artist herself, but also in the inserts of deliberate low-quality imagery, generic fonts and cheap special effects.
[101] Birtwistle 2010, p. 177.
[102] Steyerl 2012, p. 42.

It is clear that the voices in Marten's, Rafman's and Steyerl's films do not work in an harmonious arrangement with the other sounds and images. The multiple artificial voices in *Dust and Piranhas* rendering a variety of sensations keep the audio-spectator on his or her toes as they constantly change, confusing the audience in what it is listening to. The disruption in *Woods of Arcady* results in a *free counterpoint* where loosely connected tracks (voice, music and image) contradict in tone, resulting in ambiguity oscillating the audience between the interpretations of each track. When approaching the disembodied and synthesized voice as noise in *How Not To Be Seen*, it is clear that a noise containing voice does not reduce in value, rather it reverses its common traits, and makes its new value heard to the audience.

It is interesting to note that all three artists are occupied with the disruptive effects of sound and that it does not correspond to the often mentioned theorized effects of the voice in the existing literature on film sound. I carefully put forth the idea that filmmakers who are not defined as or do not identify themselves as 'artists' could be more interested in the harmonizing qualities of film sound. Of course, there are many exceptions, as the theories that have been discussed bring forward filmmakers concerned with sound's disruptive qualities. But this chapter has shown that the artist films are very much involved in disturbance through sound and voice and less with harmony. To finish this thesis I will take a broader understanding of film's spatiality. Does the voice as catalyst of disharmony, disruption and resistance in a filmic text function differently in a different spatial understanding of film? The works by the artists discussed all have been exhibited in very specific exhibitionary situations, whereas most of the theory applied to the voice in cinema has been studied in a textual bubble. In the following chapter the exhibitionary conditions of the films will be taken into account since it is something all artists that have been discussed are concerned with.

# Chapter 3: The Expanding and Roaming Voice

Artist films tend to be more reflexive to their exhibitionary situation and therefore experiment more with space. Unlike in cinema, where the 'black box' display format has been a defining standard for many decades, artists working with film have not been confined to the movie theatre. Instead  museums and galleries allow for many kinds of audiovisual presentation, with the approach of film and video as installation as a major influence on the release of its black box spatial confinement. As art and media scholar Andrew Uroskie has noted ''as the moving image was brought both literally and figuratively inside the gallery space, it would disrupt the established traditions of exhibition and spectatorship therein.''[103] In fact, all of the films discussed in the previous chapters have not been presented as a simple black box screening, but as moving images installed in carefully created environments.

For example, Rafman's *Erysichthon* has two main display formats, one in which the audience can watch the film while moving back and forth on a swing, and one wherein the film is shown in the ruins of a Greek temple that has been reconstructed with undecorated, geometric pillars and blocks (fig. 3.1 and 3.2.). Furthermore, Ed Atkins is also known for presenting his films in a style that differs from the black box, exhibiting in split-screen or wall-sized projections that reach the floor of a space.[104] Marten's *Dust and Piranhas* has been presented in a five-screen open air setting, Rafman's *Woods of Arcady* has been released on his YouTube channel and Steyerl's *How Not To Be Seen* is accompanied by props from the film and a floor drawing of a resolution target (fig. 3.3 and 3.4).

Due to the differences in exhibitionary form and the logical consequence that every presentation will offer a distinct audiovisual experience the final chapter recedes from the analysis of the video works as texts in the Barthesian sense of the word and instead treats them as events. According to Barthes a text is an assembly of codes for the reader to interact, interpret and question, a ''moving play of signifiers, without any possible reference to one or some fixed

---

[103] Uroskie 2014, p. 85. Moreover, established modes of experiencing film are now also being disrupted, either with or without an artist's intent, as films are also being watched online on smartphones, laptops and other types of screens that do not reside in the cinema, museum or gallery.
[104] Unfortunately I can't provide an example of the exhibitionary situation of *Delivery to the Following Recipient Failed Permanently* since none of its installations seem to be documented.

signifieds.''[105] Even though here nothing *inside* of the text is fixed, everything *outside* of it is. The theoretical remnants of this text-oriented era still linger in the study of film sound (but also in the study of imagery), and as Rick Altman states ''sound itself is most often treated as if it were an ideal conveyer of linguistic or musical information, received by an ahistorical audience in a generic viewing situation, with no particular moviegoing purpose.''[106] It is curious that the seminal works by Chion on the sound film also do not take into account the dimensions outside of the film text. However, I do not wish to generalize as many contemporary scholars do pay attention on the sound film's extra-textual dimensions.

Following Rick Altman in this chapter I will not consider the case studies as autonomous aesthetic entities. Instead, the cinema as *event* ''is constituted by a continuing interchange, neither beginning nor ending at any specific point.''[107] Altman employs the notion of 'cinema as event' instead of 'film as event', however, to our specific case studies classifications such as 'cinematic' are less applicable. Because of this, I will reformulate Altman's notion to 'film as event' in order to also encompass artist film. Here, studying a film event will also take into account the spatial conditions of the screenings. As a result this chapter has little concern for the inner workings of the film text (e.g. how a synthesized voice relates to the mise-en scène) and shifts its focus to the spatial dimension of sound in the space of the gallery. The events for the films *Factory of the Sun* (2015) by Hito Steyerl and, returning to the first case study, *Erysichthon* by Jon Rafman will be moments of focus. What are the consequences to the interpretation when the voice leaves the worlds of *Factory of the Sun* and *Erysichthon* and what does this event-based approach deliver in contrast to a text-based one?

I will discuss this through an investigation of the exhibitionary situations of Steyerl's and Rafman's works. This chapter enlarges its theoretical framework to art history and consults, next to previously discussed authors like Chion, Altman and Kerins, a text by media scholar Paul Hegarty, and various online sources, such as non-scholarly articles, a video and a website to substantiate the argument and logically also to provide more documentation on exhibitionary situations. The chapter is structured in two analyses of the selected works and within these discussions Altman's introduction to his 1992 publication *Sound Theory Sound Practice*, a text that has been referenced various times in the preceding chapters, is guiding. In his text Altman has summarized thirteen key attributes to the film event, of which I will discuss four and apply as

---

[105] Heath 1977, p. 10. In: Barthes 1977.
[106] Altman 1992, p. 1 - 2.
[107] Altman 1992, p. 4.

the structure of the chapter: *three-dimensionality*, *materiality*, and *heterogeneity* in relation to *Factory of the Sun* and the notion of *multi-discursivity* with *Erysichthon* in mind.

**The Voice from the Factory in *Factory of the Sun***

Hito Steyerl's 2015 film *Factory of the Sun* was first exhibited during the Venice Biennale of 2015. Like many installation artists, Steyerl also involves the space around the screen in the film event. The upper part of the screen has been tilted towards the audience, as if it can fall upon them, almost threatening. The screen is surrounded by a steel frame structure that is commonly used to attach lighting for a musical performance. The screen and its supporting frame have been placed in a dark rectangular room scattered with beach chairs. The surface of the floor and walls consists of a luminous blue grid that makes reference to the fictional device called the 'Holodeck' that is present in the television and film series *Star Trek* (1966 and onwards), which is a staging environment wherein participants can enter a virtual reality environment (fig. 3.5.).[108] The twenty three minute film was just one of the four artworks that were presented in the German Pavilion of the biennale and was presented in the basement of the Pavilion (fig. 3.6 and 3.7.). For my analysis of the disembodied and synthesized voice in the film event of *Factory of the Sun* I investigate a short sequence of the film since not all the film's voices are disembodied. In my investigation I take into account three of Altman's attributes of film sound. First, within the space of the installation three-dimensionality will be discussed. Secondly, exploring the remaining space of the pavilion, materiality is addressed. Thirdly, while leaving the confines of the Biennale, heterogeneity is at stake.

Sound can only exist in a three-dimensional context. The black box display format of the movie theatre serves to convince us that ''film-viewing is limited to the experience of the two-dimensional rectangle before us,'' as Altman notes.[109] However, as the disembodied voice in *Factory of the Sun* is not clearly connected to what is presented in that two-dimensional rectangle, its location has to be somewhere else. Especially in an era of Digital Surround Sound

---

[108] There are many online articles stating that the grid-based space is an obvious nod to the motion picture *Tron* (1982), such as Ferdinand Krag in the Nordic journal of contemporary art *Kunstkritikk* and the online platform Google Arts & Culture. However Steyerl herself has stated that it is actually based on the Holodeck, as Hyunjee Kim in the contemporary art magazine *this is tomorrow* has noted.
[109] Altman 1992, p. 5.

(as previously discussed in chapter 2), the disembodied voice can be located anywhere on the soundtrack. But essentially sounds are never tied to the screen, making the disembodied voice thus hardly unique. Mark Kerins argues that in the fully spatialized environment of the DSS style "*most* sounds are carefully placed in the multi-channel soundscape to ensure that the aural environment created matches the visual one suggested by the onscreen images." But Kerins reserves a special location for the disembodied voice, according to him:

"With no physical body to tie it down, the acousmetre is free to roam from channel to channel throughout the theater. It can choose to be in the plane of the screen - where it becomes the *only* such sound with no corresponding image - or out in the space of the auditorium. It can even occupy multiple channels simultaneously, a power generally withheld from sounds whose sources *are* or *have been* visible in the image track."[110]

This statement addresses an important difference between the visual and the audio track in analogue film works. Chion emphasizes that the information originating from the screen is a container of time and space with clear borders but the information vibrating from the loudspeakers is the opposite, since it has no clear borders or a frame.[111] Similarly media scholar Paul Hegarty notes that "the image is tethered to some solid support while the sound roams."[112] The screen has a specific location and is two-dimensional while sound is everywhere and only at work in a three-dimensional context (the same goes for receiving information through our senses 'sight' and 'hearing', as one only sees what is in sight while one can hear anything that is not visible, but in 'earshot').

In the event of *Factory of the Sun* this three-dimensionality and the power to be ubiquitous manifests itself in the audience being approached by the voice in order to involve them in the work. In the sequence to be analyzed a disembodied and synthesized voice exclaims: "I am coding a game called Factory of the Sun / but you will not be able to play this game / it will play you." Subsequently a loading screen of a videogame is shown and right after its completion a close-up of a man in a gold jumpsuit is shown. As the camera slowly recedes from the person the disembodied voice appears to tell the man: "This is your mission / you start out as forced labourer in a motion capture studio / every movement you make will be captured

---

[110] Kerins 2010, p. 265.
[111] Chion 2009. p. 226.
[112] Hegarty 2014, p. 13.

and converted into sunshine.''[113] In an interview Steyerl explains that ''Factory of the sun is, among other things, about a fictional computer game, in which every human impulse, every type of work, every movement, every emotion, is translated into light. And this light is captured and collected. It's like a currency, like money, something that's universally valid.''[114]

With Altman's insistence on three-dimensionality and Kerins's statement on the disembodied voice, it can be stated that this voice occupies the space of the auditorium and also addresses its information to the audio-spectator (not just the on screen character). Add to this the fact that the audience is physically situated *in* the artwork upon entering the room and, according to the rules of the game, is occupied with work as he or she is seated on a beach chair, a product that helps someone 'collect' light (*Factory of the Sun'*s currency), and the voice becomes more powerful: it is not instructing from the perspective of the screen, it operates outside of the text, engulfing the addressee. The voice talks the audience into the position of gamer or worker and speaks for ''the thematic and ideological environment'' that encloses them. [115] This strengthens the film's message that we live in a world that is more and more adopting the mechanics of video games. As the artist notes: ''It's basically like a computer game where you're constantly encouraged to do something, but you press all the buttons and nothing happens - other than that you yourself are read. Your own participation is used as work, but you don't get anything out of it.''[116]

The crossing of the disembodied voice, originating from the film's virtual world, to the spatial environment the audience inhabits, or the 'real world', helps blurring the boundaries between two realities, as both the film characters, consisting of zero's and one's, and the audio-spectators, consisting of flesh and blood, are addressed by the voice. It expresses that digital actors have a very real impact on the real world, influencing and determining our lives in such a way that we can't speak of two separate realities anymore.

Another attribute of the sound film is its materiality and seen ''as a material product, cinema quickly reveals the location and nature of its sound track(s)'' and ''encourages us to move past the imaginary space of the screen to the spaces and sounds with which cinema must compete.''

---

[113] Steyerl, *Factory of the Sun*, 2015, transcribed voice-over.
[114] ''German Pavilion – La Biennale di Venezia 2015 (Deutsche Welle)'', ifa Institut für Auslandsbeziehungen, 30 Juli 2015, 23:36 - 23:58 <https://www.youtube.com/watch?v=T6tH-mA0Eto>.
[115] Linnert 2017.
[116] ''German Pavilion – La Biennale di Venezia 2015 (Deutsche Welle)'', ifa Institut für Auslandsbeziehungen, 30 Juli 2015, 26:32 - 26:50 <https://www.youtube.com/watch?v=T6tH-mA0Eto>.

[117] In this section I gradually move towards a *social materiality*, an aspect that in this case concerns me more, however the accent on its *physical materiality* also has consequences. Film's materiality is emphasized in Steyerl's installation due to the visibility of the ceiling mounted film projector, the loudspeakers that have been attached to the metal structure, the tilt of the screen and the diagonal placing of the screen in the Steyerl's Holodeck-inspired installation. Recordings by visitors of the Biennale make audible that the film's sound, and especially the voices from the film (both embodied and disembodied,) are subject to a long reverberation time. The acoustics of the Holodeck allow for the voices to persist for a long time after its production and this reverberation directs attention towards the space, as a person will be more aware of a certain space when upon speaking the space transforms him or her voice, and less aware of it when nothing in particular occurs (compare this to the situation when a space produces an echo, here the space demands more attention).

The material qualities of the installation allow for the disembodied voice to fill the room, persisting long after its output. Its persistence competes with what Altman calls components of cinema's *social materiality*, such as ''the kids in the front rows, the air conditioner hum, the lobby cash register, the competing sound track in the adjacent multiplex theater…'', and many other things.[118] Of course, these more cinema oriented extra-textual sounds are less applicable in the Biennale event and, judging from the recordings of the event, the reverberation of the voice more or less covers up the sounds that are not part of the film. However in the spaces further away from the installation the disembodied voice starts competing with other sounds, as the voice does not only fill the room, it also roams outside of the basement of the German Pavilion.

As Paul Hegarty has argued sound is an important cue and marker for the mobility of the gallery visitor.[119] It announces the installation when it is not visible, spatially expanding the work out of the basement, silencing and luring visitors to the Holodeck and subsequently defining them as game participants (fig. 3.8 and 3.9.). Moreover when the screen is not visible, all the voices in *Factory of the Sun* can be considered as disembodied. For the audio-spectator who is ascending the staircase to experience Steyerl's work every audible voice is sourceless. In this exhibitionary situation the roaming disembodied voice competes with and overwhelms the other sounds. Unconnected to the screen the voice holds a different but still dominating function.

---

[117] Altman 1992, p. 6.
[118] Altman 1992, p. 6.
[119] Hegarty 2014, p. 12.

Film works are exhibited in a variety of spaces, to different kinds of audiences and in many social and cultural contexts. Why is it then that critics typically mold a film's reception to fit a single mode, homogenizing it? Altman, arguing for film's heterogeneous nature, warns that only ''as we move away from film as a single, homogenous phenomenon, we become aware of the heterogenous chain of objects and spaces which serve as a vehicle for sound.''[120] In the following paragraph I make a comparison between two exhibitionary situations the films has been presented in, Factory of the Sun at the Venice Biennale in 2015 and at the Kunsthal Charlottenburg in Copenhagen in 2016. What does the reverberation of the voice imply for the two presentations?

At the Venice Biennale Steyerl's film is situated in the basement of the German Pavilion which has been dubbed 'a factory' by curator Florian Ebner as all works in the pavillion are reflections on the notion of 'work'. In the building's verticality, Factory of the Sun's placement at the lowest point is contradicted with the artwork by conceptual artist Olaf Nicolai on the roof which, according to Ebner, ''appears as heterotope as 'another place', in which freedom is evoked.''[121] In the basement the work is being done, it is an unlit and invisible space from where the visitor can hear the reverberation of its workers voices, and it is the opposite of the non-hegemonic condition presented by Nicolai. The pavilion is a vehicle for the disembodied and synthesized gaming instructions such as ''but you will not be able to play this game / it will play you'' and ''this is your mission / you start out as forced labourer in a motion capture studio''.[122] These paralyzing and demanding instructions resonate through the pavilion as the voices of a gamified world wherein its players have no control. In a constellation of works that is preoccupied with the restrictions on human mobility that are imposed on us by 'work', the voices from the basement speak against the freedom that is pursued by Nicolai on the pavilion's rooftop, negotiating with and complementing the notion of Ebner's Pavillion.

The environment in which the disembodied voices of *Factory of the Sun* roam, the sounds that escape the spatiality of the holodeck, gives the content more meaning. The work is juxtaposed to other similar-themed works and interrupts them. According to the Slovene philosopher Mladen Dolar in his publication *A Voice and Nothing More* there exist two widespread uses of the voice. On the one hand the voice can be a ''vehicle of meaning'', serving as a symbolic interest, but on the other hand the voice can be ''the source of aesthetic

---

[120] Altman 1992, p. 7.
[121] Ebner 2015.
[122] Steyerl, *Factory of the Sun*, 2015, transcribed voice-over.

admiration'', turning it into a fetish object.[123] At the Venice Biennale the signification of the voice of *Factory of the Sun* is accentuated (however this does not mean that the uncanny presence of its synthesized demeanor is lost), but this emphasis is a result of being a specific link in the heterogenous chain of spaces *Factory of the Sun* is situated among, in the vicinity of other artworks reflecting on the notion of 'work'.

For example, at Kunsthal Charlottenburg Steyerl's film has not been exhibited in a juxtaposition to other work. This does not imply that a dialogue does not exists between the work and other works that are presented on location, but the significance the work obtained during the Biennale is altered, which is a result of the diversity and specificity of its spatial events. In Copenhagen *Factory of the Sun* is presented in a smaller holodeck (fig. 3.10.) as a singular work in the museum's white cube (fig. 3.11.). It has a less interesting dialogue with other exhibitions, as the Kunsthal's main exhibition 'An Age Of Our Making' examines how shaping or claiming the public space is a form of citizenship, less concerned with notions of work. At the Kunsthal the gallery visitor will not be as immersed in this theme as he or she will be at the Biennale. Also, the meaningful roaming voice of Venice is transformed into a voice with more expressive potential. As the speech of the disembodied voice will resonate less symbolically with the visitor, its synthesized characteristics will be channeled to the fore and emphasis will be laid upon its sounding presence, invisible and ambiguously human and mechanical. But more importantly, the voice, after acknowledging its spatial nature, has actually more in common with the space or installation than with the projected images. Unlike the images, it roams through the installation. In the event of *Factory of the Sun* we are not just dealing with the familiar notion that the audio-spectator 'enters the work', rather he or she is added to the work by the disembodied voice uninvitedly. The audience becomes not only part of a work in a physical understanding but also, and maybe even more so, in its sonic dimension. The artist film's heterogeneous nature shows that its spatial and sonic dimensions strengthen each other, diminishing the part of the two-dimensional screen.

.

**The Points of Listening in *Erysichthon***

Already discussed in relation to the power of the voice within the text in chapter 1, Jon Rafman's *Erysichthon*, seen as an event, offers more interesting reflections on the function of the vocal

---

[123] Dolar 2006, p. 4.

instrument. The sound film, and especially in relation to the voice, is also characterized by its multi-discursivity. As Altman notes "a film does not carry a single message, unified, unilinear, and univocal. Instead it is more like a *scarred palimpsest*, at various points revealing diverse discursive layers, each one recorded at a different point in time."[124] Sound can carry its own independent discourses. Who addresses the audience through the disembodied and synthesized voice and how do the film's specific events influence its message?

As we have seen in chapter 1 *Erysichthon* contains two disembodied voices, one male and one female, that suggestively monologue towards the audio-spectator, are also silent allowing for the audience's contemplation, but in the end (due to its synthesized form) sound disturbed and tired. To repeat the film's theme, the voices murmur on how contemporary circulation of images has put us in a position of insatiable hunger for them (the reference to the Greek tale), wanting to consume more and more to the point of losing ourselves in the images. Literally, we are not listening to Rafman's transformed voice, instead he has used other voices, transformed them in order to sound tired and alienated. The voices speak as if there is no turning back, a dead end, as seen in sentences like "holding hands with death / a moment without hope and without an opening" and "in those moments when consumption accelerates, beyond the desire to endure."[125] Rafman lets voices speak that have consumed too much and already experienced the effects of over consuming images.

This pessimistic rendition of the voices is influenced by the spatial configuration of the film's exhibition. The first installation at the Stedelijk Museum in Amsterdam (fig. 3.1.) features a swing set for the visitor to experience the film from. The swing is clear reference to one of the film's images, a man on a swing in a continuous loop, tied to a situation in which he is chasing his own trail, muck like the king of Thessaly or the snake devouring itself. However, it is not the visuals that are of interest to me, rather the combination of the swing set and the disembodied and synthesized voice that speaks, as discussed in chapter 1, in a hypnotizing yet unnerving manner. The movement of the swing is akin the stereotypical use of a swinging clock of hypnotic induction. By swinging back and forth the audio-spectator occupies a position of inducement with the voice acting as hypnotist. The symbolic placement of the visitor strengthens the voice's power to suggest and instruct, as it becomes more intrusive. The case is different in the presentation of *Erysichthon* in the Zabludowicz Collection in London (fig. 3.2.).

---

[124] Altman 1992, p. 10.
[125] Rafman, *Erysichthon*, 2015, transcribed voice-over

By letting the disembodied and synthesized voices roam an environment of ruins the references to classical themes are emphasized like the self destruction of the Greek king Erysichthon.

Another classical reference is augmented. In the film there are two sequences wherein a data centre is shown, but only the first sequence is of real interest to this study since the second sequence is not influenced by a disembodied and synthesized voice. The first sequence superimposes an image of the sculpture *A Marble Figure of a Young Satyr Wearing a Theater Mask of Silenus* by an unknown artist from the first century a. d. (fig. 3.12 and 3.13.). The sculpture shows a faun with a mask of the satyr Silenus attempting to scare another figure, by putting his hand through the mouth of the mask.[126] In *Erysichthon* the mask is directed towards the audio-spectator, making it seem that a head is devouring an arm, and in turn attempts to scare the audience. The disembodied and synthesized voice is present during this sequence and utters "when you leave consciousness, you will do so contented with one thought / you know you did not intentionally create another human being / you did not bring into being another consciousness."[127] The reference to death (leaving consciousness) by devouring or overconsumption is clear. The voice unnervingly contemplates on the audience's death and by making the visitor experience the film while standing among ruins Rafman's theme of decay is accentuated. Later in the film a similar sequence involving the data center and another mask is shown (fig. 3.14.), this time without the presence of a voice. The second mask's details and identity are difficult to trace, it is either another Silenus mask or a mask representing the Greek god Oceanus, and for this reason I will not extensively elaborate on this mask.[128]

This section has shown how Rafman uses a voice that has seen too much to talk a pessimistic view on the intake of images into the audio-spectator. The varying environments wherein the voice addresses the audience direct the communication. In the swing installation of *Erysichthon* the voice as hypnotizing force is brought to the fore, foisting a certain state of consciousness on the dangling visitor. Contrastingly, when experiencing the film between the ruins the voice does not achieve this quality. Here, the juxtaposition of the voice to the images and installation results in a more complex reflection on the film's theme of decay. Just as in *Factory of the Sun* Dolar's two uses of the voice can be applied here, with the hypnotizing

---

[126] Kranz 1984. Kranz identifies the same figure on a sarcophagus where it clearly tries to scare another faun. This other figure is not present in the single sculpture *A Marble Figure of a Young Satyr Wearing a Theater Mask of Silenus*.

[127] Rafman, *Erysichthon*, 2015, transcribed voice-over

[128] If I could confirm the mask as representing Oceanus, an interesting case could be made on one of the classical deity's forms, an enormous river encircling the world, in relation to Rafman's view of the intake of images that permeate our surroundings. The circle or loop is a recurrent motif in the film.

quality of the voice that is working together with the swing set functioning more aesthetically and the voice of the ruins being more concerned with the transmission of meaning. As evident in two cases of the same film, the function and interpretation of the voice can vary greatly in different environments.

This last chapter has expanded from the usual text-based approach of studying film to the treatment of voices in film approached as event that are three-dimensional, material, heterogenous and multi-discursive in nature. By structuring the analysis to four of Altman's attributes of the film event I have tried to point out that sound clearly occupies dimensions that are 'outside' of the film text and that the travelling voice can seriously influence the film's interpretation. Since the disembodied voice has no visible source it is free to roam anywhere in the gallery space, amplifying its relation with the space or installation and reducing its cooperation with the projected images, creating consequences for the interpretation of the films.

 In focusing on *three-dimensionality* in *Factory of the Sun* I have argued that the voice blurs what is and what is not considered as the film world since it engulfs the audience, turning the audio-spectator in a participant. Furthermore, investigating the attribute of *materiality* it is clear that the voice has to compete with other sounds and not only acts as a conveyor of information but also as bait, signifying only that the audio-spectator has to approach and focus. Also, seen as an event and taking into account its *heterogeneity* the voice in the film can become more or less significant in diverse spatial contexts, lean towards a negotiation with the space and the information to be communicated thus dependent on a specific space. Lastly, in discussing *multi-discursivity* in *Erysichthon*'s exhibitionary situations, the emphasized content of the voices can also be subordinated to specific versions of the same work, as differences in the installation guide the accent of the text recited by the voice. This more intrusive audio-spectatorship and sensitivity to specific situations is not taken into account while conducting a text-based study. However, the results from these two case studies do not mean that the consequences in my findings are applicable to all artist films and it has to be stated that the works *could* be representative. Concludingly, when seen as an event I have shown that sound, the voice, can strongly determine the film's interpretation by the audio-spectator.

# Conclusion

To conclude in the following I sum up the results of my inquiry into the disembodied and synthesized voice as it is roaming in the artist films that have been under investigation. First, I repeat my main question as it is stated in the introduction: what is the added value of the disembodied and synthesized voice in the selected works of Rafman, Atkins, Marten and Steyerl? Again, ''added value'' is ''the expressive and informative value with which a sound enriches a given image so as to create the definite impression, in the immediate or remembered experience one has of it, that this information 'naturally' comes from what is seen, and is already contained in the image itself.''[129] In the first chapter we have observed that the voices in the case studies exert a different power over the audience than commonly used in feature and documentary film. Whereas the 'usual' disembodied voice has many ties to traditional voices of power, the authority of the voices in the case studies is undermined by sounding alien, delivering a situation wherein the spectator is to a smaller extent pushed to a single interpretation.

Considering the second chapter, if we describe something as 'organic' we refer to something living, components working together to create a functioning organism, however the case studies have shown that their disruptive sound do not just result in destabilization of the image and therefore the filmic whole. Rather they challenge the audience by not working in sync to create new interpretations of unusual juxtapositions and inhuman voices. In this chapter it is disharmony and a fragmented medium that confuses the audio-spectator and, similar to the consequences stated in the first chapter, opens possibilities for him/her own interpretation.

Lastly, when approaching film as an event the interpretation of the case studies is dependent on the scope of the spatiality employed in the research. Its voices do not just serve to create a disruptive exchange between the film and the audio-spectator, they also interfere in the places of exhibition and it is the disembodied voice that is negotiating most strongly with the spatial conditions of the case study. Also, specific events of the same work can result in a different function for the voice to fulfill.

In these pages a number of criticisms on established theories on film sound have been presented. I have sought to make audible that a number of conceptions on the voice, in regard

---

[129] Chion 1994, p. 5.

to power and harmony, are more complicated when applied to film works that owe more to the history of video art than to the cinema. I do not dismiss the value of the research conducted in the field of film studies as the texts of many scholars, such as Chion and Altman, have proven to be very insightful to my argument. In my discussion of the synthesized voice it was challenging to work with existing research and felt the need to conduct a more unorthodox and less theory-based study on the sonic qualities of the transformed voice, however I found the research in this uncharted territory to be most valuable to add to the existing body of knowledge on the cinematic voice.

Briefly mentioned in chapter two, recommendations for further research may inquire in this supposed split between artists and filmmakers, with the former being more involved in disruption and the latter in harmony. Of course, this assumption needs to be put in perspective and disputed in order to learn something from it, but in this light it is curious, as film scholar Cormac Deane mentions, that "when we remember that the sounds of computers are, phenomenologically speaking, unnecessary and unmotivated, the insight that they are, psychoanalytically speaking, necessary and motivated, reveals the extent to which desire is at work.[130]" The addition of alien sound synthesis has to be a motivated choice wherein a certain value is being pursued by the artist and this accentuates the value of listening to the sounds and process of mediation.

However in no way am I arguing for bringing forward the idea that all artist film will be necessarily more interested in sound's disruptive qualities than its harmonizing ones. Rather, I have tried to point at differences between specific cases of the disembodied and synthesized voice and theory that is based on results of the study of films from another discipline. An important difference is the focus on either signification (in film studies) or a more medium-based research (in the arts), wherein I identified that studying the medium is very valuable to the interpretation of a work, an interpretation that is also tied to a specific situation, being event-dependent. Concludingly, I have sought to reopen the study of sound while taking into account its specific spatial dimensions with regard to the voice. Also, in focusing on the disruptive quality, seemingly powerful and noisy medium a disembodied and synthesized voice can be I have tried to make it sound that it is impossible to ignore its tendency to interfere and discard it as racket, not ruling out that there is value to be found in the dissonance and cacophony of many invisible free-roaming and confusingly alien voices.

---

[130] Deane 2016, in: Greene 2016, p. 463.

# Bibliography

Adorno, Theodor W., and Eisler, Hanns, *Composing for the Films,* New York, Oxford University Press, 1947.

Anderson, Mitchell, 'Jon Rafman/Erysichthon review', *Flash Art,* 30 September 2015, <https://www.flashartonline.com/2015/09/jon-rafman-erysichthon/>.

Boillat, Alain, 'On the Singular Status of the Human Voice': *Tomorrow's Eve* and the Cultural Series of Talking Machines', in: Albera, Francois, and Tortajada, Maria, *Cinema Beyond Film: Media Epistemology in the Modern Era*, Amsterdam, Amsterdam University Press, 2010.

Alter, Nora M., *Sound Matters: Essays on the Acoustics of German Culture*, Berghahn Books, New York, 2004.

Altman, Rick, *Cinema/Sound*, New Haven, Yale University Press, 1980.

Altman, Rick, *Sound Theory Sound Practice*, New York, Routledge, 1992.

Attali, Jacques, *Noise: The Political Economy of Music,* Minneapolis, University of Minnesota Press, 1985.

Barthes, Roland, *Image, Music, Text,* London, Fontana Press, 1977.

Beck, Jay, *Designing Sound: Audiovisual Aesthetics in 1970s American Cinema*, New Brunswick, Rutgers University Press, 2016.

Birtwistle, Andy, *Cinesonica: Sounding film and video*, Manchester, Manchester University Press, 2010.

Bordwell, David, and Thompson, Kristin, *Film Art: An Introduction*, New York, McGraw-Hill, 2013.

Cattelan, Maurizio, 'Helen Marten: A little bit naked', *Flash Art*, feature 283 March-April 2012.

Chion, Michel, *Film, A Sound Art,* New York, Columbia University Press, 2009.

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

Chion, Michel, *The Voice in Cinema*, New York, Columbia University Press, 1999.

Cooke, Mervyn, *A History of Film Music*, Cambridge, Cambridge University Press, 2008.

Connor, Steven, *Beyond Words: Sobs, Hums, Stutters and Other Vocalizations*, London, Reaktion Books, 2014.

Copland, Aaron, 'Tip to the Moviegoers: Take off Those Ear-Muffs', *The New York Times*, 6 November 1949

Doane, Mary Ann, 'The Voice in the Cinema: The Articulation of Body and Space', in *Yale French Studies* no. 60, Cinema/Sound (1980), pp. 33-50.

Dolar, Mladen, *A Voice and Nothing More*, Cambridge and London, MIT Press, 2006.

Ebner, Florian, 'German Pavilion La Biennale Di Venezia 2015', Website of the German Pavilion 2015, accessed 2 August 2018, <http://archiv.deutscher-pavillon.org/2015/en/>.

''Ed Atkins: 'Delivery to the Following Recipient Failed Permanently' (2011)'', *Frieze*, 2011.

Eisenstein, S., Pudovkin, V. and Alexandrov G., 'A Statement', in Weis, Elisabeth and Belton, John, *Film Sound: Theory and Practice*, New York, Columbia University Press, 1985.

Eisenstein, Sergei, *The Film Sense*, Meridian Books, New York, 1957.

Farmer, Brett, *Spectacular Passions: Cinema, Fantasy, Gay Male Spectatorships*, Duke University Press, Durham, p. 63.

Flinn, Caryl, *Strains of Utopia: Gender, Nostalgia and Hollywood Film Music*, Princeton University Press, Princeton, 1992.

Goldsmith, Mike, *Sound: A Very Short Introduction*, Oxford University Press, Oxford, 2015.

Gorbman, Claudia, 'The Master's Voice', in *Film Quarterly*, vol. 68. no. 2. (Winter 2014), <https://filmquarterly.org/2015/01/12/the-masters-voice/>.

Gorbman, Claudia, *Unheard Melodies: Narrative Film Music*, Bloomington and Indianapolis, Indiana University Press, 1987.

Greene, Liz et al., *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks*, London, Palgrave Macmillan, 2016.

Hegarty, Paul, *Rumour and Radiation: Sound in Video Art*, Bloomsbury, London, 2015.

Horton, Justin, 'The Unheard Voice in the Sound Film', in *Cinema Journal,* vol. 52, no. 4 (Summer 2013), pp. 3-24.

ifa Institut für Auslandsbeziehungen, ''German Pavilion – La Biennale di Venezia 2015 (Deutsche Welle)'', 30 Juli 2015, <https://www.youtube.com/watch?v=T6tH-mA0Eto>.

Jay, Robert, *Downcast Eyes: The Denigration of Vision in Twentieth-Century French Thought*, University of California Press, Berkeley, 1993.

Jutte, Robert, *A History of the Senses: From Antiquity to Cyberspace*, Cambridge, Polity Press, 2005.

Kahn, Douglas and Whitehead, Gregory, *Wireless Imagination: Sound Radio, And the Avant-Garde*, Cambridge and London, MIT Press, 1992.

Kalinak, Kathryn, *Film Music: A Very Short Introduction*, New York, Oxford University Press, 2010.

Kelly, Caleb, *Cracked Media: The Sound of Malfunction*, MIT Press, Cambridge, 2009.

Kerins, Mark, *Beyond Dolby (Stereo): Cinema in the Digital Sound Age*, Bloomington and Indianapolis, Indiana University Press, 2011.

Kim, Hyunjee Nicole, 'Hito Steyerl: Factory of the Sun', *this is tomorrow*, 14 August 2017.

Kinney, Katherine, 'The Resonance of Brando's Voice', in *Postmodern Culture*, vol. 24, no. 3 (May 2014).

Kisner, Jordan, 'Rain is sizzling bacon, cars are lions roaring: the art of sound in movies', *The Guardian*, Wednesday 22 Juli 2015.

Knight, Arthur, and Wojcik, Pamela Robertson, *Soundtrack Available: Essays on Film and Popular Music*, Duke University Press, Durham, 2001.

Kozloff, Sarah, *Invisible Storytellers: Voice-Over Narration in American Fiction Film*, University of California Press, Oakland, 1988.

Krag, Ferdinand Ahm, 'Sun, Sky, Screen and Cloud', *Kunstkritikk*, 13 Januari 2017.

Kranz, Peter, *Jahreszeiten-Sarkophage. Entwicklung und Ikonographie des Motivs der vier Jahreszeiten auf kaiserzeitlichen Sarkophagen und Sarkophagdeckeln*, Gebr. Mann, Berlin, 1984.

Lacasse, Germain, 'The Film Lecturer', in: Gaudreault, Andrė, et al., *A Companion to Early Cinema*, Wiley, Hoboken, 2012, pp. 487-497.

Lawrence, Amy, *Echo and Narcissus: Women's Voices in Classical Hollywood Cinema*, Berkeley, University of California Press, 1991.

Licht, Alan, 'Sound Art: Origins, Development and Ambiguities', *Organised Sound* vol. 14. no. 1 (April 2009), pp. 3-10.

Linnert, Nicolas, 'All Access Politics: Reality and Spectatorship in Two Film Installations by Jean-Luc Godard and Hito Steyerl', *Xtra*, vol. 19. no. 3. (Spring 2017), <http://x-traonline.org/article/all-access-politics-reality-and-spectatorship-in-two-film-installations -by-jean-luc-godard-and-hito-steyerl/>.

Meadows, Mark Stephen, *I Avatar: The Culture and Consequences of Having a Second Life*, New Riders, San Francisco, 2007.

Meigh-Andrews, Chris, *A History of Video Art,* Bloomsbury, London, 2014.

Mitchell, W. J. T., 'There Are No Visual Media', in Kelly, Caleb, *Sound: Documents of Contemporary Art,* MIT Press, Cambridge, MA, 2011, pp. 76-79.

''The Secret History of the Vocoder'', *The New Yorker*, August 20 2014, <https://www.youtube.com/watch?v=OvR4qK0B--w>

Nichols, Bill, *Introduction to Documentary*, Indiana University Press, Indianapolis, 2001.

Oxford Dictionaries, ''Noise | Definition of noise in English'', accessed August 2 2018, <https://en.oxforddictionaries.com/definition/noise>

Pearce, Sheldon, 'From DJ Screw to Moonlight: the unlikely comeback of chopped and screwed', in *The Guardian*, Tuesday 24 Januari 2017.

Pisanski, Katarzyna et al., 'Voice Modulation: A Window into the Origins of Human Vocal Control?', in *Trends in Cognitive Sciences*, volume 20, issue 4 (April 2016), pp. 304-318.

Reggiani, Enrico, 'Rewording in melodious guile» W.B. Yeats's The Song of the Happy Shepherd and its Evolution Towards a Musico-Literary Manifesto', *A Journal of Irish Studies*, no. 2. (2012), pp. 73-92.

Renov, Michael, *The Subject of Documentary*, University of Minnesota Press, Minneapolis, 2004.

Ruoff, Jeffrey, 'Conventions of Sound in Documentary', in Altman, Rick, *Sound Theory Sound Practice*, New York, Routledge, 1992.

Silverman, Kaja, *The Acoustic Mirror: The Female Voice in Psychoanalysis and Cinema,* Indiana University Press, Indianapolis, 1988.

Sonnenschein, David, *Sound Design: the Expressive Power of Music, Voice, and Sound Effects in Cinema*, Michael Wiese Productions, San Francisco, 2001.

Spielmann, Yvonne, *Video: The Reflexive Medium*, Cambridge, MIT Press, 2008.

Stam Robert et al., *New Vocabularies in Film Semiotics*, Routledge, London and New York, 1992.

Stam, Robert, *Subversive Pleasures: Bakhthin, Cultural Criticism, and Film*, Johns Hopkins University Press, Baltimore, 1989.

Steyerl, Hito, 'In Defense of the Poor Image', in: Steyerl, Hito, *The Wretched of the Screen*, Sternberg Press, Berlin, 2012 (originally published in 2009), pp. 31-45.

Tate Modern, ''Hito Steyerl - 'Being Invisible Can Be Deadly' | TateShots'',  May 17 2016, <https://www.youtube.com/watch?v=kKAKgrZZ_ww>.

Tompkins, Dave, *How to Wreck a Nice Beach: The Vocoder from World War II to Hip-Hop, The Machine Speaks*, Melville House Publishing, New York, 2010.

Twerdy, Saelan, 'This is Where it Ends: The Denouement of Post-Internet Art in Jon Rafman's Deep Web', in *Momus*, (9 july 2015).

Uroskie, Andrew V., *Between the Black Box and the White Cube: Expanded Cinema and Postwart Art*, University of Chicago Press, Chicago and London, 2014.

Viola, Bill, *Reasons for Knocking at an Empty House: Writings 1973 - 1994*, MIT Press, Cambridge, 1995.

Wojcik, Pamela Robertson, 'The Sound of Film Acting', in: *Journal of Film and Video*, volume 58, no. ½ (Spring/Summer 2006), pp. 71-83.

# List of Illustrations

**Chapter 1:**



Fig. 1.1. Jon Rafman, *Erysichthon*, 2015, single channel digital film, colour, 8 minutes. A snake devouring his own tail, referring to the old Greek myth of Erysichthon of Thessaly about a king who got cursed and finally ate himself in hunger.
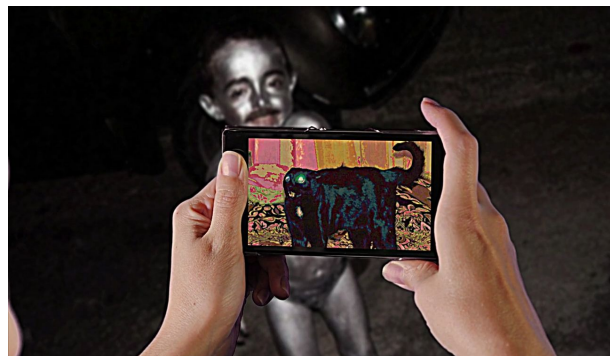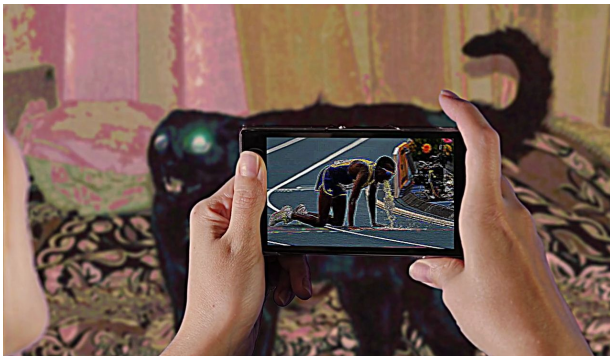


Fig. 1.2 and 1.3. Jon Rafman, *Erysichthon*, 2015, single channel digital film, colour, 8 minutes. Images are successively being captured by a mobile phone. Each image engulfs the film frame before being captured and consequently forgotten.

Fig. 1.4, 1.5 and 1.6. Ed Atkins, *Delivery to the Following Recipient Failed Permanently*, 2011, single channel digital film, colour, 17 minutes. The disembodied and synthesized voice in Atkins's film is interrupted by other sounds (but also visuals) giving way to different filmic components.

**Chapter 2:**



Fig. 2.1 and 2.2. Helen Marten, *Dust and Piranhas*, 2011, five channel digital film, colour, 25 minutes. The sole two synchronized disembodied and synthesized voices are connected a jumpy doric CGI column and a stop-motion animated statue.
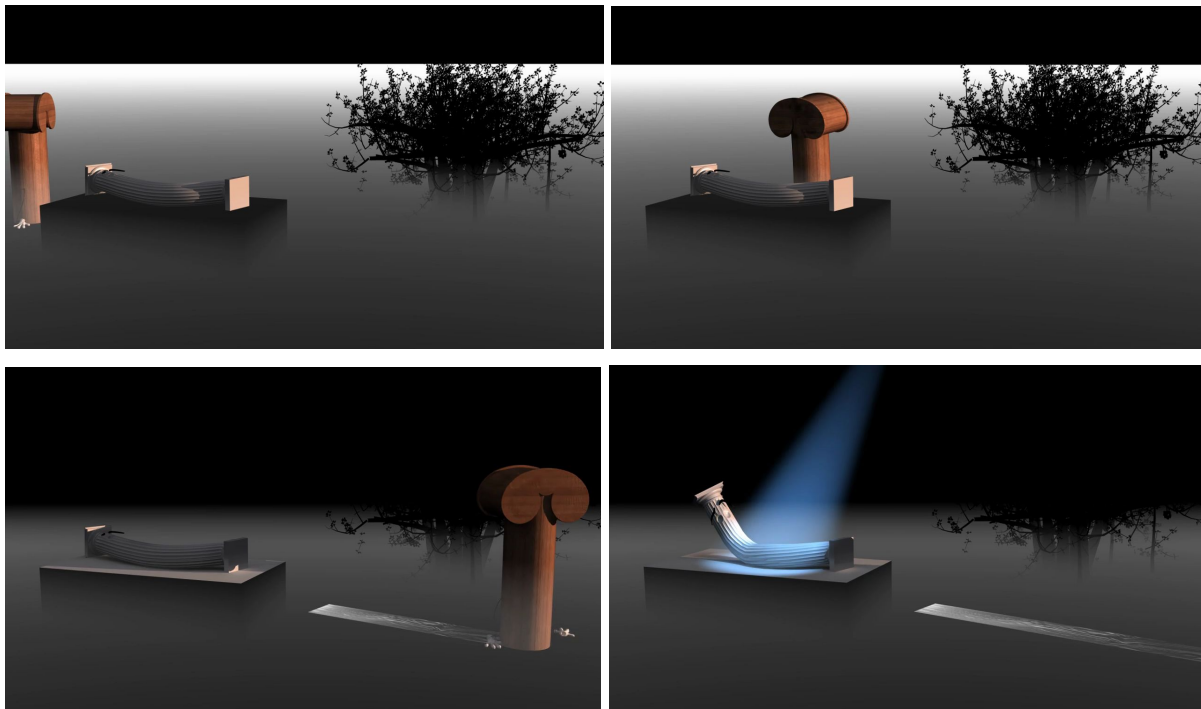


Fig. 2.3, 2.4, 2.5 and 2.6. Helen Marten, *Dust and Piranhas*, 2011, five channel digital film, colour, 25 minutes. The animation that is playing while the multiple voices are proclaiming their 'treaties on gorgeousness'.
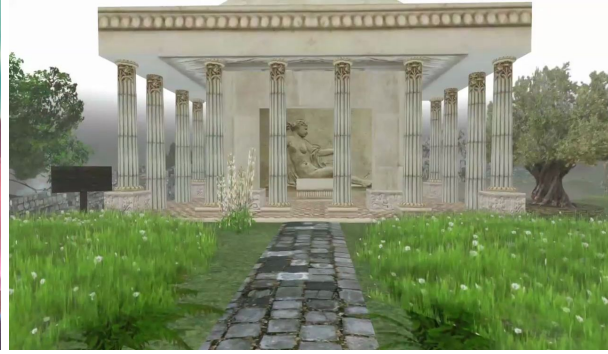
Fig. 2.7, 2.8 and 2.9. Jon Rafman, *Woods of Arcady*, 2010, single channel digital film, colour, 4 minutes. The first three Second Life recordings that are shown, representing a digital arcadia.



Fig. 2.10. Hito Steyerl, *How Not To Be Seen: A Fucking Didactic Educational .MOV File*, 2013, single channel digital film, colour, 16 minutes. The introduction of the first lesson 'How to make something invisible for a camera'.
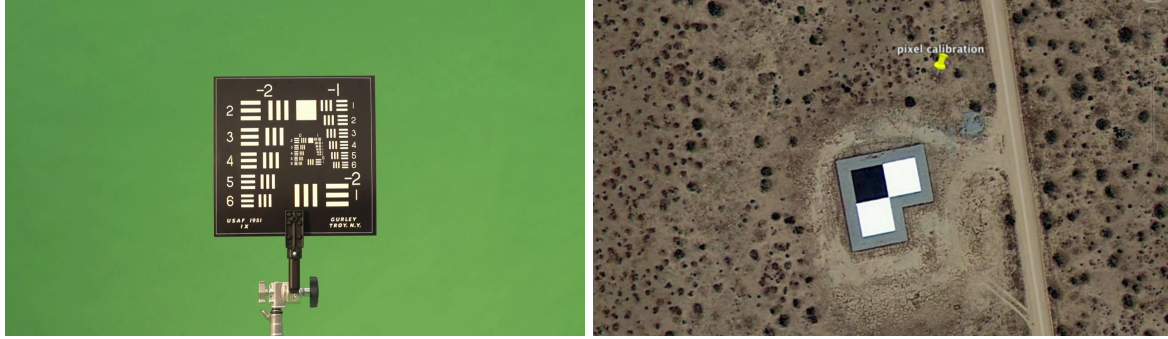
Fig. 2.11 and 2.12. Hito Steyerl, *How Not To Be Seen: A Fucking Didactic Educational .MOV File*, 2013, single channel digital film, colour, 16 minutes. Left: A resolution target as used in the nineteen fifties by the US Air Force, now decommissioned. Right: A pixel-based resolution chart, contemporary resolution is a quarter of a meter per pixel.

**Chapter 3:**



Fig. 3.1 and 3.2. Two display formats of Rafman's *Ersysichthon*. Left: the installation as it was presented during the exhibition *I have ten thousand compound eyes and each is named suffering* at the Stedelijk Museum in Amsterdam in 2016. Right: the installation as presented in the exhibition *Jon Rafman* at the Zabludowicz Collection in London in 2015.



Fig. 3.3 and 3.4. Left: The installation view of Marten's *Dust and Piranhas* at the Serpentine Gallery in London in 2011 (Park Night Project). Right: Hito Steyerl's *How Not To Be Seen* in the exhibition *Hito Steyerl: Too Much World* at the Institute of Modern Art in Brisbane in 2014. The film was shown in a darkened adjacent space.

Fig. 3.5. Hito Steyerl, *Factory of the Sun*, 2015, single channel digital film, colour, 23 minutes. Installation view of Steyerl's *Factory of the Sun* during the Venice Biennale in 2015.



Fig. 3.6 and 3.7. Left: The German Pavilion at the Venice Biennale, a neoclassical style building that has been altered in 1938 to National-Socialist architectural ideals. It has remained largely unchanged up until present day. Right: The architectural plan of the pavilion in 2015. After crossing through the facade of the building the central gate led to the 'motion capture studio', referring to *Factory of the Sun*.

Fig. 3.8 and 3.9. The stairs to the entrance of *Factory of the Sun*. The sound of the film is free to roam in the spaces outside of the 'motion capture studio'. The audio-spectator of the pavilion is subject to a continuous flow of voices even when he or she is not physically present in the installation.
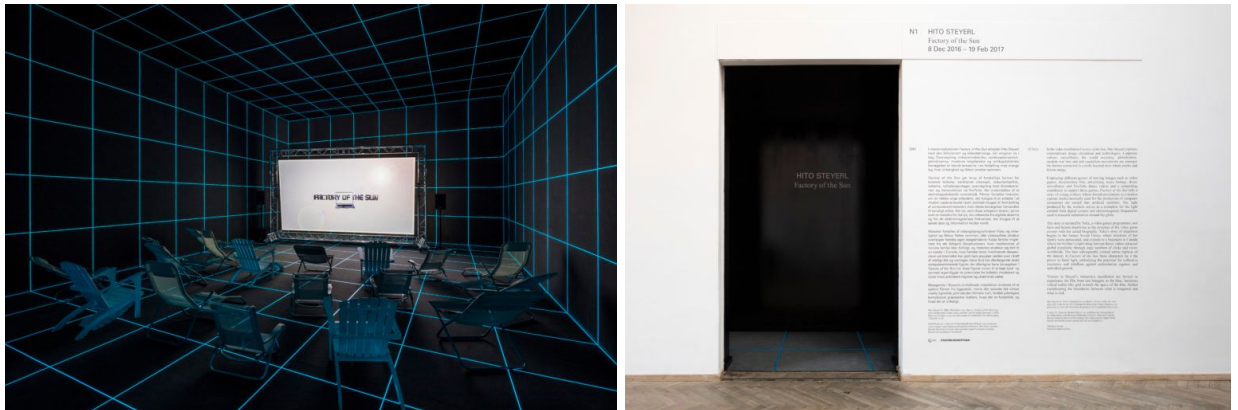




Fig. 3.10 and 3.11. Left: *Factory of the Sun*'s installation view in Kunsthal Charlottenburg in 2016, the screen has been placed in the left corner of the holodeck (as opposed to its right placement in Venice) and the space of the holodeck is smaller than previous presentations (this is easily confirmed by counting the squares of the deck's grid). Right: *Factory of the Sun* as a singular work in the exhibition 'Hito Steyerl: Factory of the Sun' from 08-12-16 to 19-02-2017.

Fig. 3.12 and 3.13. Jon Rafman, *Erysichthon*, 2015, single channel digital film, colour, 8 minutes. Left: The sculpture *A Marble Figure of a Young Satyr Wearing a Theater Mask of Silenus*, unknown artist, first century a.d. Right: The sculpture as it is imposed in Rafman's film, directed towards the audience.



Fig. 3.14. Jon Rafman, *Erysichthon*, 2015, single channel digital film, colour, 8 minutes. Unidentified mask.

# Appendix

To support my argument I have transcribed the text that is uttered by every disembodied and synthesized voice in my case studies. In some cases, such as Marten's *Dust and Piranhas* and Steyerl's *Factory of the Sun,* I have not completely transcribed the spoken text of the films but only the text in the sequences to be analyzed. In some cases the transcription of the words has turned out to be problematic due to pronunciation. For example, the voice in Ed Atkins's *Delivery to the Following Recipient Permanently* employs a large number of uncommon words that even a native English speaker will find difficult to receive. To denote inaudible pronunciations I have inserted '(inaudible') in some of the transcriptions. In order to clarify the nature and source of the spoken text I have created some key characteristics to create a better understanding of the transcription:

---

W: Female voice

M: Male voice

MW: Multiple voices (male and female) speaking simultaneously

<span style="color:red">Red</span>: Voice of God phrases that address the audience directly (only relevant to *Erysichthon*)

<span style="color:blue">Blue</span>: Poetic phrases that do not address the audience directly (only relevant to *Erysichthon*)

(1): a number between brackets indicates the amounts of sentences spoken (only relevant to *Erysichthon*)

---

**Transcribed voice-over from Jon Rafman, *Erysichthon***

W: You fell asleep from exhaustion at 7am this morning.
You were awake all night and the only thought in your mind was that to exist is to devour oneself. (2)

M: If you look at these images enough, you begin feeling like you composed them. That you took the photo. You remember the dream in which you had glimpsed the germ of the idea. (3)

M: These images posted at random are forgotten, have ways of permeating their surroundings. and thus effecting you even though you are barely conscious of perceiving them. (1)

W: When you leave consciousness, you will do so contented with one thought. You know you did not intentionally create another human being. You did not bring into being another consciousness. You hope this gives you courage while your mind departs. You hope you remember to meditate on this while you depart. You will hope you remember, to remember. (6)

W: A moment of complicity and intimacy. Holding hands with death. A moment without hope and without an opening. (3)

W: I finally said it: my start-up failed, failed, failed (repetitions and variations) (1)

M: I have then thousand compound eyes and each is named suffering. (1)

W; In those moments when consumption accelerates, beyond the desire to endure. (1)

M: It is a far, far better thing that I do, than I have ever done. It is a far, far better rest that I go to, than I have ever known. (2)

W: I never thought I'd die alone. I laughed the loudest, who'd have known. I trace the cord back to the wall, why did building seven fall? (3)

W: The void also attracts you, otherwise you wouldn't have any vertigo. But you will die if you fall. And what can you do with the void, except, fall into it? (3)

**Transcribed voice-over from Ed Atkins, *Delivery to the Following Recipient Failed Permanently***

W: I will survive, without you.

M: Uhh…
Uhh…
Uhh…
Uhh…
Delivery to the following recipient failed permanently.
When I speak the word 'smoke' for example, there certainly seem to be ways in which I might invoke a more material sense of the word, inside your brain, i mean, as a hemorrhage of sorts, is how you might like to think of it.
If I nurse the word in my mouth and on my lips and with my throat. If I shape it, turn it in the right fashion with my yellowing elegenic tongue and my weak teeth. I might manage to send the word spinning off more smoke than if I really say it, hidden among so many other words in some banal sentence.
There's no smoke without fire, etcetera.
That suctioned backwards stream of smoke from mouth to nostrils.
I want to make you aware of my mouth, to map my mouth, comprehensively using the word 'smoke'.
To make the word lap and plot the position of every surface in there, and so turned, carefully release the word. And the word fanning out into the cool evening air. In the still gulf between my mouth and your ear.
Coagulating as it goes, thickening, so that when it arrives at your ear, it's only just petrifying into matter. It barely makes it on such hardening bakerlight wings, landing ominous as a fat black beetle on your earlobe. Just outside your vision, its surprising weight understood as an earring. From there it works its way in, dousing lazily with its antenna.

M: Are you recording?
M: Yeah.

M: Then purposeful the word has been fashioned by me to fit perfectly inside your convoluted ear, snugly, it's a tailored word. Every surface of its ever stouted body correlating with every surface of your diminishing inner ear. Prodding, caressing purposefully, in the way that one might communicate with the blessed death blind. The shape of my mouth mentioned instructions for the reformation of the word 'smoke', which then convulses up to your brain, then swerves left and down into your gorgeous mouth. Do you even have a mouth? Without matter? If you didn't have a mouth.
Perhaps it is like smoking, like drugs, like certain sexual practices, like liver, so long as you tried it once, so long as you once harboured a mouth I reckon it will work. So long as you tried a mouth out before, so to speak. So long as you've let stuff in, expelled stuff via a mouth, shoved a salty finger in there to retrieve a nugget of wet wadded crisp pulp.

A craterous mola(inaudible 4:19)

Felt the relief of removing a two thick-strand in you, herbal hair of a…

From between two of the more plate-like teeth. Suffered an alse, or a cut, or bit in a lump of cheek, clean off, or temporarily disabling the tongue in a bite meant for other dead meat. Burnt your tongue to leave it craving abrasive toast like textures, detained something in there, smoke, an egg, a momentary orb of springwater, expanded the mouth and proceeded to aerate and cool some scolding morsel of…

Performed that particular sucking move, matched by, maybe, a look of utter contempt in the eyes. Concentrating the saliva in your mouth, into that depression in the tongue, solely conjured for the purpose of spitting. Similarly, hawked up some though slimed chunk from a sineous passage to complement or empower or justify the spitting. As long as you've confused the scale of things in there, most importantly, is for you to have some appreciation for the complexity of the tongue. To have licked an ice-cream, a plate, softened wood, a clitoris, a stamp, a wound, a penis, etcetera. So long as you can appreciate something of the mouth and its tongue hegemony. Then, when that word, smoke, reaches into you and reveals its shape and its weight and the ways in which these correspond to the movements of my stinking mouth, you should be fine materialising it. Making it (inaudible)

I have tried to swallow words. To force them down with a gulp of saliva bulge recovered from around my teeth. To cosset them, swaddle them in saliva to give them a fighting chance of being digestible.

I found that the most instinctive thing to do is just breathe the word like smoke, though that should be resisted. In this perverse account, smoke is too much like air, is too much like nothing.

What little body there is to smoke is predominantly visual. Little scent, sure, little sting in the eyes, but no weight no splashing turd. You should swallow it properly, down the wrong pipe. You may choke a little at first, to be expected. You may gag. These are good signs, it means there's something there. Something is taking shape, something is clinging up, becoming itself, solidifying, fleshing out, thickening. Where, in an instant, and as your tongue spasms imperceptibly, you stuff wads of stuff around the constituent letters of words, draping steaks of stuff over the crossbar of the A. Packing sausages of stuff into the snaked scaffold of the S. What on earth are you spelling?

Every metaphor here should relate to the tongue, to your own tongue. Every metaphor should ideally be your cuttlefish tongue.

Certain licks will tattoo, so careful.

I'm sure you can picture as well as I, that children seemingly licked their lips perpetually, to the point when their lips where outlined clumsily with sore red skin. Is that from the acid content of your saliva? Why did you persist? Surely the first sign of your face eroding, you'd stop licking? A terrible narcissism, especially as evidenced in a child. Though it's surely of interest to the child. You, would be addicted to tasting their own lips or the skin immediately surrounding their lips.

71

Perhaps the flavour improved the sorer the skin became? I imagine the taste would become more metallic, but I could be wrong. Perhaps some flavour would be revealed? Somewhere between the fifth and the sixth layer of downie child skin, the defied analogy, an original flavour, like coke, or metal, metal-lic.

Can you remember if those children who licked their lips raw, or the same children who took a while to speak, or that they mispronounced words, or that they swore shockingly, and that no one, particularly their parents, could make out where they have heard such appallingly course biological language. Sam papering off their mouths with their rasping kitten tongues. Tongues for blowing raspberries or for eakin into the dint (inaudible) on raspberries, or for rolling as a demonstrable birthright. Tongues for being pressed down with the flat side of a wooden lolly stick, either for some unknown examination of after the lash bit of red slush.

Your tongue lapse this way and that, gesturing, and acting some convulsive spell to summon the body of the word while simultaneously expunging its symbolic order. Your tongue calling upon the word to shrug off its fears, its aspirations, its fucking being. Your tongue, the murderer, sitting there in your ignorant mouth, failing in ability, when really it's the most mobile thing I can think of. Picture the uncanny swaying head of a cobra, before a strike, only looking just like a mole rat. Only speckled, flecked with those marks of abuse. Chili, smoking, coffee, cripplingly sour gummy sweets etcetera. Wrong with pride, as testament to its impressive grim (inaudible).

And so the swaddled word, imaginatively rendered in a pertinent typeface…(inaudible on purpose)  is swallowed, hold, don't involve with these, to be dispersed biologically by the various acids, ammonia, bleaches, (inaudible), etcetera. I have no idea. Absorbed into the bloodstream and carried, illicitly, about the body, swept along that cardiac thigh to affect its changes. To transform, ultimately, every single cell of your oblivious body into something always ready appended. Syntactically, but also systically with the effective word.

It's hard to describe really. Moulting. There are some correlatives. An ovoid of mercury placed on the back of your hand, slowly impossibly passing through said hand. Over the course of a year maybe, emerging birth from an invisible stigmata on the palm and dropping to the linoleum floor like a fatted grub.

Only having shrunk, having shed some of its volume inside your hand. Your blood then shuttling these fugitive glimmering globules, only they wouldn't glimmer in that darkness. Glimmering globules sliding about the body. Touching the sides, inducing a thick ache in your veins and a dull thrum in your brain. If we don't operate now you might….

Or drugs. The flavour of some drugs. Summoning the thought that perhaps you're in fact tasting yourself. Stripped, marinated by the drug. Swallow hard. And that the effect of the drug is in fact a sub functionly of yourself. A subcutaneous seam of affect that was inside prepared by the drug, sure, it was in fact yourself, a regurgitating process. (inaudible) whereby you ingest yourself in order to affect yourself. Parts of yourself you had no idea existed. For instance, there

is a purple layer of skin somewhere in the middle of the dermic sandwich that if licked, summons frightening or erotic visions of… that's right.

And for instance, behind your right eye is something like an autolith, in a jet black and mercurial, an obsidian slug murmuring something about the woods. Or more likely you quaffing (inaudible) with impunity. The doctors said that was fine, which was a little worrying really. Because the warnings on the bottle explicitly contradicted that. It's morphine after all. So you might extrapolate that you were in some way beyond or exempt from those universal cautions. That you would enter a different phase, an advanced phase, an inhuman phase, where you may as well… drink as much of the stuff as you like. As much as it takes to dull the pain, because it's all irreversible now.

From herein, my friend, it can only be palliative. That's how she said it, cow. And with every ever easier swig the effects appeared duller. With every ever greedy a swig it's harder to achieve the same alleviation from the pain. You find it very difficult to remember what it's like to be relieved by that first prescribed swig. Let alone what it was like to live without the pain, the illness.

I find it equally hard to remember sex with any accuracy, or the weather. Both exist in an incoherent fog that harbours solid formations, lurking at out reach. Or a morph approximates a representation of that, but only a representation. You are not well, my friend, that's how she said it, with an ignorant condescending crumpling with the brow. I'm not your friend, cow. Is what you thought. This life you lead now, slushing about, full of oramorph, is as if on the brink of death beneath your own smothering pillow, and above that, the ferocious arm of a loved one.

Imagining all of the air in a given space, a bedroom for example, replaced instantaneously with concrete. All the air, all the apparent space in the world exchanged in a blink, for concrete. Again, everything full, close, cold, dead, dark. Instant death. That's how i'd like to go. All the way from sensation. grounded in the comprehension of greater or lesser distances between myself and everything else to abject insinuation. Nothing but infinite motionless density. All those gaps in the cosmos, those unaccounted for spaces between everything and everything else. The infinitesimal rifts between quats, the vast drifting nothings between galaxies. All of that, suddenly filled in with concrete.

At night, I dream of Pluto, the dejected ex-planet. A sphere of ice, four billion miles away from the bedroom.

**Transcribed voice-over from Helen Marten, *Dust and Piranhas***

Transcription starts at 03:38 and ends at 04:20.

MV (inanimate statue, normal female voice, pitched down male voice): But let's grab some air already, here in the beginning. For maybe these considerations, these indulgences, full heart and encaring, unless we can agree to be lighthearted.

There will be time later spent rubbing the dust of our backs. Thumbs and spit to wipe our cheeks. Let's scrabble instead through the perfect facets of gemstone that give ourselves a personality. Let us make a chocolate bar that fits into the pocket of every sports jacket. Square, practical, good. No shame in the precision max, in our new and improved, wanting to taste the difference, to be fitter, happier, more productive.

Transcription starts at 16:00 and ends at 17:24.

MW (inanimate statue): But, let's keep going please. Among all these heights, let's raise companion to fill our libraries, stock our fridges and drag our neighbours out to town.

MW: Treaties on gorgeousness

M: With art that (inaudible), art directors surely can't help asking. How much drama, how much mood, how much information? Should the dance be arresting, or is it more glamorous when seen and not heard?

MW: In so many cases (inaudible) is too glamorous to remain as content. Should we glam into that content or vice versa?

W: (hysteric) Is it glamorous to be glamorous? Are we stuck with metaglamour? Is an illustrated magazine such a hot format for glamour? Are we marketing (inaudible) male order out, internet shopping or vice versa?

W: As long as it remains possible to tackle vice versa to the end of each paragraph. We must be doing something right or vice versa.

MW: we know that glamour is not an object, not an action, not an idea. We know glamour never emerges from the nature of doing things.

M: There are no glamorous people, no glamourous events. We know glamour is artificial. We know that in order to be glamorous we have to become playdress, intellectual parasites. We move in on history and occupy images.

M: Emptying them of meaning, reducing them to shells, we fill these shells with glamour... (rest of the sentence is inaudible, as it becomes distorted and starts echoing)

**Transcribed voice-over from Jon Rafman, *Woods of Arcady***

W: The woods of Arcady are dead,
And over is their antique joy;
Of old the world on dreaming fed;
Grey Truth is now her painted toy;
Yet still she turns her restless head:
But O, sick children of the world,
Of all the many changing things
In dreary dancing past us whirled,

M: To the cracked tune that Chronos sings,
Words alone are certain good.
Where are now the warring kings,
Word be-mockers? — By the Rood
Where are now the warring kings?
An idle word is now their glory,
By the stammering schoolboy said,
Reading some entangled story:
The kings of the old time are dead;

W: The wandering earth herself may be
Only a sudden flaming word,
In clanging space a moment heard,
Troubling the endless reverie.
Then nowise worship dusty deeds,
Nor seek, for this is also sooth,
To hunger fiercely after truth,
Lest all thy toiling only breeds
New dreams, new dreams; there is no truth
Saving in thine own heart.

M: Seek, then,
No learning from the starry men,
Who follow with the optic glass
The whirling ways of stars that pass —
Seek, then, for this is also sooth,
No word of theirs — the cold star-bane
Has cloven and rent their hearts in twain,
And dead is all their human truth.
Go gather by the humming sea
Some twisted, echo-harbouring shell,
And to its lips thy story tell,

And they thy comforters will be,

M (different voice): Rewarding in melodious guile
Thy fretful words a little while,
Till they shall singing fade in ruth
And die a pearly brotherhood;
For words alone are certain good:
Sing, then, for this is also sooth.

M (different voice); I must be gone: there is a grave
Where daffodil and lily wave,
And I would please the hapless faun,
Buried under the sleepy ground,
With mirthful songs before the dawn.
His shouting days with mirth were crowned;

W: And still I dream he treads the lawn,
Walking ghostly in the dew,
Pierced by my glad singing through,
My songs of old earth's dreamy youth:
But ah! she dreams not now;

M: dream thou!
For fair are poppies on the brow:
Dream, dream, for this is also sooth.

**Transcribed voice-over from Hito Steyerl, *How Not To Be Seen: A Fucking Didactic Educational .MOV File***

M: How Not To Be Seen: A Fucking Didactic Educational .MOV File

M: Lesson 1

There are four ways to make something invisible for a camera.
To hide.
To remove.
To go off-screen.
To disappear.
This is a resolution target.
It measures the visibility of a picture.
This is a resolution target.
It measures the resolution of the world as a picture.
Resolution determines visibility.
Whatever is not captured by resolution is invisible.

How not to be seen: Lesson 2
There are seven ways of making something invisible in plain sight.
Pretend you are not there.
Hide in plain sight.
To scroll.
To wipe.
To erase.
To shrink.
To take a picture.

W: Today most important things want to remain invisible
Love is invisible.
War is invisible.
Capital is invisible.

M: In the nineteen fifties and sixties the US Airforce installed grey scales and resolution targets in the California desert to calibrate aerial photographs and videos.
Resolution determines visibility.
It calibrates the world as a picture.

How not to be seen: Lesson 3.
There are seven ways of becoming a picture.
To camouflage
To conceal.

To cloak.
To mask.
To be painted.
To disguise.
To mimicry.
To key.

Around 2000 a new standard for resolution targets is introduced.
This is a pixel-based resolution chart.
It serves to shoot pixels.
In 1996 photographic resolution in the area was about 12 metres per pixel.
Today it is one fourth.
To become invisible one has to become smaller or equal to one pixel.

How not to be seen: lesson 4.
There are thirteen ways of becoming invisible by disappearing.
Living in a gated community.
Living in a military zone.
Being in an airport, factory or museum.
Owning an anti-papparazzi handbag.
Being fitted with an invisibility cloak.
Being a superhero.
Being female and over fifty.
Surfing the dark web.
Being a dead pixel.
Being a wifi signal moving through human bodies.
Being undocumented or poor.
Being spam caught by filter.
Being a disappeared person as an enemy of the state.
Eliminated
Liquidated
And then dissimilated

W: In the decades of the digital revolution 170000 people disappear.
Disappeared people are annihilated.
Eliminated.
Eradicated.
Deleted.
Dispensed with.
Filtered.
Processed.
Selected.
Separated.

Wiped out.

Invisible people retreat into 3d animations.
They hold the vectors of the mesh to keep the picture together.
They re-emerge as pixels.
They merge into a world made of images.

M: How not to be seen: lesson 5.
There are 54 ways to merge into a world made of images.
To hide.
To remove.
To go off screen.
To disappear.

This is a resolution target.
This pattern has been decommissioned in 2006 as analog photography lost its importance.
Rogue pixels hide in the cracks of old standards of resolution.
They throw off the cloak of representation.

**Transcribed voice-over from Hito Steyerl, *Factory of the Sun***

Transcription starts at 07:00 and ends at 08:03.

W: I am coding a game called Factory of the Sun.
But you will not be able to play this game.
It will play you.

W: This is your mission.
You start out as forced labourer in a motion capture studio.
Every movement you make will be captured and converted into sunshine.
For part of this game, from this scenario.
They were shot by corporate drones to increase global uprisings.