Department of Humanities

Master's Degree in Philosophy and languages of modern times

Final Dissertation

Recordings: the problem of recurrence in the philosophy of sound

Supervisor

Prof. Federico Laudisa

Prof. Claudio Calosi

Student

Raffaello Deturres

No.231470

Academic year 2023/2024

...To Anna and Rossana

Contents

Chapter 1

Introduction

In August 1999, a musical review in "Guitar One" featured an interview with a young John Frusciante, who shared his thoughts on recordings: "Every time you put on a record or a CD, it's a dimension in time; it's a piece of time that's suddenly moved from one dimension to another. A moment in 1927, or a series of moments that accumulated into being one song in 1967, suddenly appears again in 1999 because I press a button. That's like dying and waking up in another world". Twenty-five years later, this idea still holds a certain fascination, not just for the trance-like state music can induce, but for the idea that recordings can *resurrect sounds from the past*, performing a sort of magic trick. While Frusciante expresses this concept poetically, his intuition uncovers philosophical concerns in at least three areas: the nature of sound in general, the possibility of sound recurrence, and the nature of recordings as a source of information.

In the theoretical realm, the argument about recordings gained prominence at the end of the last century, sparked by a renewed interest in sound as a matter of perception. This field, often referred to as *sound ontology*, encompasses various perspectives on the object of auditory perception and its ontological status. In this framework, recordings, akin to echoes and other acoustic phenomena, challenge traditional notions of what sound is.

Recordings particularly support the intriguing idea of sounds returning from the past, blurring the distinction between events and objects. Our lives are filled with things that can be encountered repeatedly, such as people, cars, or thoughts. For example, we might not see a friend for a whole year but still recognize him upon reunion. In contrast, singular events like New Year's Eve 2020 or the Beatles' last concert on a London rooftop are said to occur *only once in history*. However, we still preserve memories of such events through audio recordings, photographs, and documentaries, even for those who were not present. In a nutshell, recordings may prompt us to reconsider our intuition about sound, aligning them closer to long-lasting entities rather than momentary, unrepeatable

events.

Before delving into the world of sounds, it is crucial to clarify the relationship between sound and music. It is important to note that this work *is not primarily focused on music*. When we think of recordings, it's natural to reminisce about favorite vinyl records or tapes—perhaps recalling the iconic 'It's My Life' by Talk Talk, as a person born in the 70s might. Indeed, recordings are deeply intertwined with the world of music, and thoughts of recordings often lead us to their musical content. However, it's essential to recognize that vinyl records, tapes, and CDs can reproduce music *because* they can reproduce sounds. While music falls within the realm of arts and aesthetics, sound, as the raw material from which music emerges, is fundamentally a perceptual phenomenon that conveys information about the external world. In the words of Kania:

> The distinction between sound and music is difficult to limn precisely, but there is a general agreement that human beings (unlike dogs, say) hear music in certain sounds, just as we see three-dimensional arrangements of pigments [2](p.121).

Throughout this work, as we illustrate scenarios where recordings play a role, we will reference numerous examples of musical recordings. However, any mention of tunes, albums, and music should be understood in the context of the more fundamental dimension of sound, defined as what enters auditory perception.

The present work articulates around a fundamental question: do audio recordings really entail *recurrence* for sounds? What follow serve as my attempt to answer this fascinating question. On the other hand, to provide a concrete answer we must before introduce the more general framework of studies which is the ontology of sounds.

Beginning with 2, this work embarks on a comprehensive exploration of the intricate world of acoustic phenomena and its profound philosophical implications. We initiate our journey by delving into the philosophy of sound, recognizing its pervasive presence in our lives, and acknowledging its historical underrepresentation in philosophical discourse, especially when compared to the extensive discussions surrounding vision (2.1). Our investigation delves into the complexities of sound perception, contrasting theories of *direct* and *indirect* perception, and their implications for our understanding of sound as either a private, subjective experience or an integral part of our shared reality. We scrutinize the pivotal role of auditory perception in shaping our overall experience of the world and delve into the nature of sounds themselves as distinct entities. Within this framework, we explore various ontological positions about sounds categorizing them, as proposed by Casati,

Dokic, and di Bona, into *proximal* (2.2), *medial* (2.3), *distal* (2.4), and *aspatial* theories. Particular emphasis is placed on describing sound as individual events, an interpretation that stands as one of the most compelling.

Central to this work is the examination of the status of recorded sounds, which takes center stage in 3. In this section, we compile the significant observations from philosophical literature, shedding light on how recordings contribute to shaping our understanding of sounds. We introduce the intriguing problem of *recurrence*, intricately linked to the experiences of sound playbacks (3.1). To elucidate this debate, we commence with an exploration of the prominent stance on events, known as *particularism*, and delve into Davidson's defense of this perspective (3.2). Subsequently, we delve into philosophical interpretations inspired by recordings that aim to challenge this conventional position regarding sounds. However, as we approach the conclusion of this chapter, we present a defense of the standard position by drawing a thoughtful comparison between recordings and other forms of representation, such as photography and sculpture (3.3).

The final chapter, 4, explores the possibility of considering sound playbacks akin to other forms of representational systems, providing a potential solution to the puzzle of recurrence. This section also encompasses personal contributions to the topic, addressing three fundamental questions: How many audio images exist (4.1)? What do recordings represent (4.2)? And what kinds of changes can recordings undergo (4.3)?

Ultimately, this work extends an invitation to reconsider our relationship with sound, encouraging us to perceive its essence beyond mere auditory sensation. It prompts us to appreciate the intricate nature of sound representation in recorded formats. The overarching goal of this work is to offer a nuanced perspective on the metaphysics of sound, challenging readers to contemplate the profound philosophical inquiries that arise from our auditory experiences.

Chapter 2

Metaphysics of sound

2.1 The philosophy of sound

Imagine this cosy scene: we are in the comfortable embrace of our living room. The soft glow of ambient lighting sets the perfect mood for an evening of musical enjoyment. With a shared enthusiasm for jazz, we decide to enjoy one of the most iconic albums of all time: "Undercurrent" by Bill Evans and Jim Hall. Soon the loudspeakers began to play the sweet and haunting notes of 'My Funny Valentine'.

Sounds are ubiquitous entities that require no formal introduction. For anyone with the gift of hearing, sounds are a fundamental part of everyday life, filling our world to the extent that we occasionally find ourselves compelled to plug our ears. Sounds do not need to be searched for; they enter perception as a primary source of data, as do tactile impressions, smells, tastes and, of course, colours and shapes.

Historically, the realm of audition has not generally played a prominent position in philosophy. Sporadically it has been included in disquisitions on the reliability of senses, as witnessed in ancient skepticism, or in doctrines of natural physics, as in Aristotle's. In essence, we can say that audition has rarely been studied in its own right. In most cases it has played second fiddle to its more celebrated counterpart: vision. This neglect of audition has been so conspicuous that certain scholars have coined the term 'visuoventrism' to describe the seemingly "tyrannical" dominance of vision over other senses [50](1.1).

Nevertheless, philosophical questions about sound have often been raised in the landscape of *Philosophy of Perception*. Broadly speaking, the philosophy of perception is a theoretical discipline that asks whether one can ever perceive the physical world directly. It's within this framework that inquiries into sensory data often fall. As Di Bona and Santarcangelo explain in [17], this investigation applied to the world of hearing has produced two main currents: those who

assert a direct contact with sound sources through perception, and those who think that sounds are rather mental entities formed from basic sensory data.

With regard to this second strand, the theory of *indirect perception*, we can sum it up as a combination of two claims:

1. Sounds are private, subjective and detached from their apparent physical sources;

2. in an auditory experience, nothing more than sound and audible properties are heard.

The first claim is sometimes supported by the existence of some unusual hearing experiences. As recorded in the Old Testament, the Lord purportedly spoke to King Solomon in a dream. While the authenticity of this event may be debatable, it is undeniable that only Solomon could have heard this conversation. Regarding the second assertion, sounds always appear in perception with some audible attributes: they can be high or low, loud or soft, short or long. This cannot be said of their sources.

This concept of indirect perception was, at one point, the stance adopted by some representatives of British empiricism concerning hearing. For instance, as Berkeley argued in "Three dialogues between Hylas and Philonous" (1713), hearing the sound of a carriage on the street does not permit us to infer the existence of the object that produced it. According to his theory of perception, which posits that all empirical data, whether primary or secondary, exist solely within the realm of perception, sounds are not reliable carriers of knowledge. Indirect perception aids in our comprehension of how sound operates, as auditory experiences are often influenced by external factors such as taste or internal mental states.

However, upon further examination, Berkeley's idealism, when taken to its extreme, seems to pave the way for what could be termed 'sonic isolation'. Assuming that sounds are entirely private and disconnected from the external world, then how can we ensure that that two individuals attending a performance of "The Barber of Seville" in a theater are essentially witnessing the same opera? And do we not perhaps be puzzled when the loud noise of a carriage driving along the streets occurs but in a desert road?

On the other hand, those who subscribe to a theory of *direct perception* believe that in addition to sounds, sources also enter into auditory perception. According to this idea, perceiving a whistle is not the same as perceiving a mental entity that is detached from the external world and mediates our experience of reality. Rather, hearing, the ability to perceive sound, makes us immediately aware of something happening outside in the world—a whistle in this case.

Going back in time, one of the most important proponents of this view was Galileo Galilei in [22]. He was the first to suppose that the audible properties of sound were directly related to the properties of the movement of air. Shortly afterwards, the concept of *mechanical waves* was developed and was destined to play a major role in physics. The same idea of direct perception is the one adopted by modern acoustics, that is the discipline devolved to study sound-waves, their generation and their propagation in the space.

In the last five decades, philosophical interest in sounds and audition has entered a new phase. Questions that previously piqued the curiosity of empiricists and scientists, have now converged in the spectrum of metaphysics[1]. Under the umbrella of this renovated philosophical doctrine, sounds have unleashed a series of articulated debates: What are sounds as entities? Are they products of the mind? Is our perception capable of faithfully representing the qualities of sound? When can two sounds be considered identical?

In the present chapter, we will delve into the metaphysics of sound, examining its primary currents and debates. First, we will introduce the most widely accepted criterion for categorizing theories based on where they locate sounds. Once the concept of sound's location is established, we will explore the three main positions one by one: Proximalism, Medialism and Distalism. By the end of this chapter, we will have a solid foundation to move into the core subject of this work: the status of recorded sounds.

2.1.1 Matters of space

In everyday contexts, we often describe sounds based on their temporal characteristics, such as their duration, the shape of their envelope, or their rhythmic patterns. These features are readily apparent and influence our perception. For example, think about statements like "the alarm ringing ten times this morning" or "the notes on the left side of a piano are more sustained compared to those on the right". However, we less frequently consider the spatial properties of sounds. Do sounds have spatial properties? It appears they do. Upon close examination of our auditory experiences, we recognize that environmental noises don't merely materialize randomly in our hearing; instead, we perceive them *as originating from specific positions in the surrounding space*.

[1]In the literature it is possible to find references to both a metaphysics and an ontology of sounds: generally speaking, while ontology is the discipline that answers the question *an sit* (what there is), metaphysics prefers the form *quid sit* (what is its nature). As far as the present topic is concerned, metaphysics and ontology are closely related. Indeed, questioning the existence of sounds has consequences for what sounds are supposed to be; at the same time, metaphysical choices often determine ontological positions. However, given the nature of the arguments in this chapter, focused on the essence rather than the existence, I have preferred to use the more general label 'metaphysics'.

The relevance of spatial information in the realm of sound extends to several domains, including *music composition*, *auditory scene analysis*[2], and *philosophy* [16]. In this context, philosophy serves as the focus of our discussion. The significance of spatial attributes in the metaphysics of sound is largely driven by a meta-theoretical consideration: many philosophical theories about the nature of sounds propose distinct spatial statuses for them. For instance, some theories contend that sounds occupy the same region as their sources, while others suggest that sounds reside within the mind of the perceiver. From this perspective, inquiries like "What is sound?" are reformulated as "Where are sounds? Do they occur in any specific location?"

Casati and Dokic, as discussed in their work [11], offer a taxonomy that classifies these diverse theories based on their positions regarding the localization of sounds. This taxonomy identifies four main groups of theories:

1. *Proximal Theories*: these theories assert that sounds are localized either in the listener's head or in the immediate vicinity of the perceiver.

2. *Medial Theories*: mainstream acoustics aligns with medial theories, suggesting that sounds are located in the medium in which both the listener and the sources are immersed.

3. *Distal Theories*: this category posits that sound is located at the resonating sources itself.

4. *Aspatial Theories*: contrasting with the others, aspatial theories contend that spatial properties do not directly inhere in sounds, but rather in their sources.

It's important to note that this partition is not a rigid framework necessitating theoretical monism. Recently, some scholars have proposed hybrid descriptions of sounds that draw from multiple spatial localization concepts, *e.g.*, [33]. Furthermore, even aspatial theories, which deny sounds a specific location, implicitly assume a metaphysical stance on auditory perception and its objects. This underscores the importance of spatial cues not only in the philosophical investigation of sound but also in structuring the entire field of study.

Before presenting proximalism, I would like to inform the reader that aspatial theories will not be explored further in the present work. This stems from several considerations combined. Firstly, aspatial theories of sound suggest more than just the idea that sounds lack a *precise location*, a

[2]Auditory Scene Analysis is the discipline that studies the process by which our brain organises and interprets the complex mixture of sounds that occur in the environment. Albert Bregman's research [7] showed that the human auditory system adopts a natural problem-solving strategy for disentangling the chaotic acoustic field of disorganised auditory stimuli. Sounds are split into basic sensory elements, grouped and then attributed to the correct sound sources; during this mental process, spatial continuity is an essential cue for organising sounds in space.

notion which in itself might be tenable. These theories propose that sounds are entirely devoid of any relationship to any spatial region. In stark contrast to this, I firmly believe that spatial information is an intrinsic aspect of our auditory experience, forming a fundamental property of sounds themselves. Of course, this justification alone is not sufficient to remove aspatialism from the framework. Secondly, my choice to bypass in-depth analysis of aspatial theories is influenced by the fact that discussions surrounding these theories are frequently referred to the work of Strawson [63]. This work, which initially introduced the concept of aspatialism in the context of sounds, was the focal point of an extensive and spirited debate during the late 1970s. In essence, this theory represents a distinct chapter within the broader domain of the metaphysics of sound. By the way, the ontological commitments of aspatialism will be further elucidated when we will explore Martin's abstractism concerning sounds (3.2.2)—a concept notably influenced by Strawson's philosophical contributions.

2.2 Proximal theories

A metaphysical theory of sound is defined as proximal if it *locates sound at or below the physical surface of the listener*. In part, this kind of idea has already been introduced when speaking of *indirect perception*. Proximal theories typically align with the notion that sounds are equivalent to sensations, essentially viewing them as a private mental experiences[3]. For a proximalist, exotic scenarios like the one of King Solomon or disturbance caused by malfunction of the ear (like tinnitus) count as genuine instance of sounds.

Proximalism aims to make sense of propositions that delve into the phenomenological and psychological dimensions of auditory experiences. It give importance to the *private* dimension of listening and perception in general. However, someone who tolerates a reduction in terms of sensation will also emphasise that sounds do not *exist* outside the interaction with the listener, but at most they are somehow *related* to external events or objects.

From a wide perspective, this attitude represents an instance of what is called *indirect realism*. An clear definition is that given by Brown:

> The most common forms of indirect realism hold that we directly perceive mental entities—variously referred to as 'idea', 'sensa', or 'sense data'—that are caused by a purely physical interaction between the perceiver and the and the physical world. But once such entities are introduced, the existence of illusion, hallucinations and

[3] An example of proximal theory of this kind is the one of Maclachlan in [43].

> dreams forces us to conclude that we cannot deduce properties of physical objects—or even that physical objects exist—from the properties of the objects we perceive [8](p.341).

When applied to the field of acoustics, this perspective implies that sounds are nothing but a *collection of auditory impressions*. Any attempt to draw inferences about the external world from these impressions involves a complex process akin to passing through a "black box" in which perceptual stimuli are processed. The truth of these inferences therefore depends on the extent to which this process is intelligible. Many scientists, such as Hermann von Helmholtz, have argued for the possibility of gaining knowledge about the inner mechanisms that govern the perception of sound[4]. This latter, more moderate form of indirect realism forms the basis of psychoacoustics, the scientific study of sound perception.

As mentioned earlier, proximal theories sometimes encompass hallucinations as pure sounds. But what exactly are hallucinations? Take three different kinds of sensible cognition, in order: *veridical perception*, *illusion* and *hallucination*. All these three forms have in common that they arise as sensations. Illusion differs from veridical perception in that the former reveals properties that do not belong to the perceived objects. For example, a sequence of tones that creates the auditory illusion of a Shepard scale appears to be continually ascending or descending in pitch, even though it's not[5]. This is a systematic illusion of the ear, which can be uncovered through frequency analysis. Importantly, illusions can often be recognized upon closer examination of the object generating the sensation.

Conversely, hallucinations differ from both veridical (accurate) and illusory perception in that they have no physical source. A person suffering from tinnitus may hear a constant ringing or whistling in the ear. These disturbances are not caused by a physical event hitting the unfortunate person's eardrum, but by a malfunction of the nervous system. It is no coincidence that the disturbance caused by tinnitus is called 'phantom noise'. Hallucinations can't be measured or perceived by other subjects, but since the object of these experiences is naturally categorised as sound, some proximal theories are designed to include them in the list.

[4]The renowned scientist Hermann von Helmholtz was concerned, among many other things, with harmony and the perception of sound. His 1863 book *Über die Sinneswahrnehmungen von Tönen als physiologische Grundlage der Musiklehre* is considered a milestone in psychoacoustics.

[5]A Shepard tone, named after Roger Shepard, is a sound consisting of a superposition of sine waves separated by octaves. When played with the base pitch of the tone moving up or down, it is called a Shepard scale. This creates the auditory illusion of a tone that is continually ascending or descending in pitch, but which ultimately does not appear to be getting higher or lower.

This position, which considers private experiences as sounds, can lead to a *solipsistic* understanding of the acoustic world, and we might wonder how to avoid falling into this drift. A simple step towards mitigate this attitude is to rid of hallucinations from the list of sounds by assuming a claim:

(C) Any reliable metaphysical theory must classify as 'sounds' all and only those sonic entities that stand in a *relation* with physical objects or events[6].

This constraint, although consistent with indirect realism, is strong enough to exclude hallucinations from the list of sounds. As regards the *publicity* of sound, it is not automatically inferred from the presence of (C). Consider a proximal scenario of two subjects, *a* and *b*, in a room at the same distance from a vibrating tuning fork. Although their experiences are caused by the same object, *a*'s sensation will be different—or simply not comparable—to that of *b*. Lastly, it's important to note that acoustic illusions should not be considered as hallucinatory; they are indeed sounds. Many non-proximal positions have engaged in extensive discussions of acoustic effects like the Doppler effect and echoes, which are associated with illusory perceptions of position and pitch [11](3.2.2).

2.2.1 Sound at the eardrum

One case of proximal theory in line with (C) is that supported by O'Shaughnessy in his work "Consciousness and the World" [52]. He contends that the sound heard is located in the same position as the listener. In contrast to reductionist approaches, O'Shaughnessy's theory leaves room for *unheard sounds to exist at a distance from the listener*. However, when it comes to auditory experiences, sounds must be situated at the surface of the eardrum:

> there is a tendency to fail to grasp that the auditory perception of space is largely directional in character, and to model it upon the visual perception of physical objects, which are visibly set both in directions out from us and in addition at *depths* away from us in three dimensional space. In imputing to the auditory mode more than it can achieve, we misunderstand its true nature and tend as a result to overlook its actual accomplishments [52](p.445).

For O'Shaughnessy, a heard sound is a *proximal stimulus*. It represents only a fragment of a more extensive acoustic event that consists of spatio-temporal components. The proximate stimulus

[6]The most commonly discussed relations in this context are *causation* and *identity*. However, other less stringent candidates are also admissible. For instance, O'Callaghan proposes a *mereological* relation. Other potential relations to consider include *supervenience*, *grounding*, *dependence*, among others.

represents only a moment in the "life" of the whole acoustic event. This perspective is motivated by the understanding that when we listen, the information available to us does not encompass the entirety of the event. While there may be a natural inclination to associate a sound with its source, the initial temporal part of the event, which includes the source as participants, remains inaccessible to the listener. However, the listener does have access to a subsequent temporal part of the event, which coincides with the intersection between the region of space occupied by the listener and the region occupied by the propagating sound waves in the medium.

The previous example of the tuning fork could play a role in supporting this theory. Assuming that the vibrating tuning fork is a *directional radiator*[7], depending on the position of *a* and *b* in relation to the beam of the source, their impression could be different regarding volumes. Now imagine that the source is a loudspeaker, *a* is in front of its directional axis and *b* is on the opposite side: whatever they hear, the auditory sensation of *b* will typically be muffled in respect to that of *a*. This is due to the fact that the sound heard at the rear of the speaker has a different spectrum to that heard at the front.

Perplexities on this interpretation have been raised by Casati, Dokic and di Bona in [11]. As they pointed out, even though proximal waves are the only thing that excite our tympanum, they also transport information about the source events. Furthermore, qualitative inferences on the intensity of the sound at its source are often true: along with the proximal perception we possess a notion of *distal volume* of sounds[8]. Secondly, there is a sense in which this proximal theory demands either *multiplication* or *multilocation* of sonic events:

> The locatedness of the sound a the hearer's position entails that there are as many sounds as there are actual (or potential) hearers around. An alternate account would consider that one and the same sound is present which is, however, multiply located [11](p.5).

O'Shaughnessy's position does not reduce sounds to sensations and is compatible with (C). However his view is not enough to preserve the fact that both participants in the proximal scenario hear exactly the same set of sounds. Everyday circumstances and common language seem to necessitate such a feature of *univocality*. Indeed, If we consider that each perceiver experiences a unique sound, it becomes challenging to address the concerns raised in the context of Berkeley's

[7]That is, it does not radiate vibrations equally in all directions.

[8]*Distal loudness perception* is the ability of the human ear to postulate the intensity of a disturbance in the proximity of its source by listening to it from a distance.

idealism. Proximal theories illuminate the importance of subjective representation of sounds in the mind; nevertheless, they might be inadequate for providing a comprehensive description of the entire sonic front. After all, even a proximalist who is not an idealist must still address what external phenomena are the subjects of auditory perception.

2.3 Medial theories

Let us return to the initial scenario. We decides to turn up the amplifier's volume when the record is on its third track, 'Dream Gypsy'. At this point, we might feel that the piano's nostalgic sound begins to physically permeate the room. This idea may be supported by a tactile sensation. Our hand, lying on the floor, begins to feel the vibration emananting from the speakers. All these impressions might lead us to conclude that the sound we hear travels through the air and causes other objects to shake.

A theory that assumes a medial position *locates sound in the medium in which the sounding object is immersed*. For example, in the scenario just mentioned, the block of air in the shape of a room surrounds both us and the sound reproduction system play the role of a medium. Medial theories often identifies sounds with waves, that is physical disturbances that occur in the medium. Auditory impressions are directly caused by, or at least supervene, on waves. This position, which dates back to modern science is now the widely accepted. Medial theories are in most cases compatible with a realism about sounds, i.e. *sonic realism*. According to this view, sounds exist even in the absence of a listener, and perhaps the inhabited Saturn's moon Titan is a very noisy place. Furthermore medialism tends to preserve the publicity of sounds.

In what follows, we will briefly introduce some terms from the field of acoustics in order to become familiar with the concept of the 'sound-wave'. Subsequently, the strength of this theory will be described from a metaphysical point of view, and finally, an essential critique will be presented.

2.3.1 Physical description of sound

Acoustics is a discipline that is situated in the broad landscape of the physical sciences. It's a multidisciplinary field that explores the production, transmission, and effects of sound-waves. From understanding how musical instruments produce harmonious melodies to designing noise-canceling mechanism, acoustics plays a crucial role in shaping our world.

In acoustics we speak of sound when a perturbation propagating through an elastic material

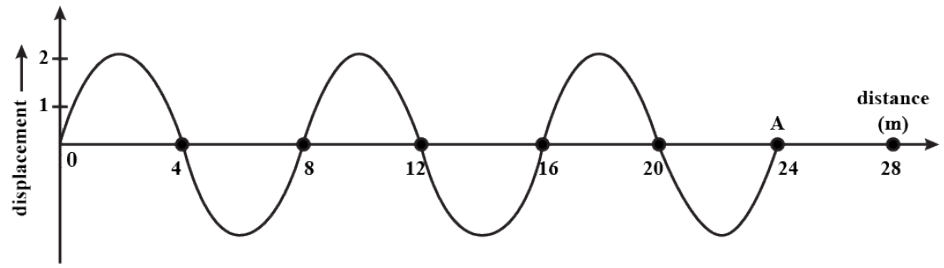

Figure 2.1: a sine wave graph displaying periodic oscillations.

causes a change in pressure detectable by a person or by a device[9]. A disturbance in a physical medium can be described as the *displacement* of particles with respect to a stationary state in which the displacement value is zero (fig. 2.1). Different values of displacement correspond to changes in the *pressure* transmitted through the medium and such alteration is typically measured in units of Newtons per square centimeter (N/cm^2).

The disturbances that produce sound waves are by nature *mechanical* waves. They are different from electromagnetic waves (like light) in that they pass through matter but cannot propagate in vacuum. In gaseous materials, such as air, sound waves behave as *longitudinal* waves, meaning that the vibratory motion of particles is aligned with the direction of wave propagation—as opposed to *transverse* waves, where the vibratory motion is perpendicular to the direction of the wave.

In the study of acoustics, the *medium* assumes a central role in being the body that receives and transmits the disturbance. It represents a necessary, albeit not sufficient, condition for sound-wave to exist. For a medium to transmit sound, it must be elastic. *Elasticity* designates the inherent capacity of certain materials to return to their original displacement state after the removal of an applied stress.

Several characteristics of perceived sound depend on the properties of the medium through which it travels. For example, the acoustic *impedance* of the medium fundamentally determines the degree of energy conservation during sound transmission. Gaseous media such as air, which have a low acoustic impedance, are remarkably efficient at conserving energy over long distances. On the other hand, *temperature* is more responsible for the speed of sound propagation. Assuming an average air temperature of 20 degrees Celsius, the speed of propagation is about 345 metres per second, so a sound produced by a source 1 km away from us will be perceived with a delay of 2.9 seconds.

The robustness and reliability of acoustics as a scientific discipline rest on the sturdy pillars of mathematics. Indeed a waveform disturbance can be represented by a function in algebraic structure.

[9]In writing this section, I made use of the first chapters of a well-known acoustics textbook by Benareck [3].

Environmental sounds are often described by complex sinusoidal function. What is represented by a function are the quantifiable attributes of the disturbance, i.e., the intrinsic properties of the wave[10]:

1. *Frequency*: it represents the number of wave cycles occurring within a single second, typically measured in Hertz (cycles per second).

2. *Amplitude*: this is the magnitude of the particle displacement at a given point in the wave.

3. *Wavelength*: it measures the spatial separation between adjacent points of the wave in the same phase, such as two consecutive crests.

Environmental signals, despite their intricate vibrational nature, can be deconstructed into elementary vibrations. The Fourier analysis method reveals that complex sound waves result from the summation of a finite number of simple sinusoidal signals. The lowest-frequency component is referred to as the *fundamental*, while the other components are known as *harmonics*. This method forms the basis of spectrum analysis. This analysis has a wide range of applications, including audio signal processing for tasks such as filtering, compression and synthesis.

As we stated at the above, acoustic experiments make use of technical devices that transcend mere perception. Think of a signal: with the help of an acoustic detector, such as a microphone, one could discover all the relevant features of the signal—for example, that it is a five-second sawtooth signal at 480 Hz with an intensity of 20 db. Using these devices, critical features of sounds can be grasped without necessarily invoking the realm of perceptual experience.

However, a disturbance that behaves like a sound wave that doesn't produce auditory perception is considered to be either infrasound or ultrasound, depending on whether it is above 20,000 or below 20 hertz. This particular threshold defines the range of frequencies typically perceived by the human ear, making it a benchmark for distinguishing what may or may not be considered audible; therefore, even in the absence of specific perceptual encounters, the human auditory system formally determines which physical disturbance is eligible to be classified as sound.

2.3.2 Sounds as waves

To consider sound waves as the main candidate for sound is to embrace the *ontological commitment* of medialism. So far we have seen what acoustics considers a sound-wave to be: a mechanical

[10]For intrinsic property I assume the definition of [8]: "*F* is an *intrinsic property* of a physical object *O* if and only if *O* is *F* independently of any relations in which *O* enters"(p.345).

disturbance occurring in an elastic medium. On the auditory side, the theory of sound waves offers the advantage of adeptly bridging the perceptual realm with physical descriptions. It aligns some of the intrinsic properties of sound waves with the *perceptual attributes* of the sounds we encounter. This second class of properties, the class of *relational properties*, includes all the audible qualities that we often attribute to heard sounds.

As reported in [11], medial perspective is effective in multiple ways: First, medialism provides a reliable link (nomological coextensiveness) between the intrinsic and relational properties of sound, which is vividly demonstrated by the compression of perceptual qualities such as *pitch* and *volume* into the frequency spectrum and amplitudes of sound waves, respectively[11]. Secondly, the perceptual aspect of *directionality*, which is often experienced in listening, is closely related to the way in which mechanical waves propagate in media. Sound seems to come from a direction because we occupy a point with respect to the source's beam. Thirdly, certain *acoustic effects*, which may seem confusing at first, are elegantly explained by adopting a medial perspective.

This last point is crucial, as the exploration of acoustic and auditory effects—particularly those considered as exotic or unusual scenarios of perception—plays a pivotal role in determining the acceptance and viability of different metaphysical theories of sound. Consider, for example, the apparent change in the observed frequencies of a wave due to relative motion between the source and the observer, namely the Doppler effect. This seemingly strange situation occurs more often than one might realize. If we are stuck in traffic and a fast ambulance whizzes past us, the sound of the siren is heard to drop in pitch as it passes us. This is exactly what the *Doppler effect* describes. In fact, if the source is taken as the coordinate system, any variation in frequency (f) can be detected; but if the perspective of a relatively stationary observer is taken, the resulting frequencies (f') can be calculated by means of a simple formula:

$$f' = f \cdot \frac{c - v_o}{c - v_s}$$

[11]The way in which pitch and loudness are related to the measurable hues of sound waves would deserve more attention. However, studies conducted by Stevens have shown that pitch is not the same as frequency, but rather it can be considered as a *function* of frequency; in [60] he demonstrated the possibility of constructing a *extensive scale* of tones based on subjective perception. An extensive scale is one in which the property of distance between its members is preserved: "the numeral 8 is as far from 4 as it is from 12, and it is half as far along the numeral series as is 16. Incorporating these properties in a scale we assign numeral in a such way that the relation among them reflect relation among objects" [60](p.331). As far as loudness is concerned, studies have suggested a less strong functional relationship. In fact, the value of perceived loudness is better projected on an *intensive scale*, i.e. a scale that only allows us to say whether one sensation is louder than another [61].

Where c is the speed of sound in the medium, v_o the speed of the observer and v_s the speed of the source, i.e. the sirens. The Doppler effect is intimately related to the medium's properties and holds a pivotal place in the defense of the medial perspective. Reflections and echoes, among other acoustic phenomena, add to the credibility of the idea that sounds instantiated by sound-waves travel through the medium. These effects, initially regarded as exotic aspects of perception, validate the medial perspective, while distal or proximal positions might be hesitant to accept them as genuine auditory experiences.

In summary, we have presented an average of the most salient points in which mainstream acoustic and medial theories are committed, such as the *identification* of sound with sound-waves and a a *close relation* between disturbances features and perceptual features. We will now survey some arguments constructed to challenge the medial position.

2.3.3 Challenging medial position

Certainly, medial theories provide a clear answer to the question "where are sounds", which is "sounds are in the medium of propagation". Unlike proximalism, this answer does not involve a controversial entity such as the mind.As a consequence, the acceptance of the medial answer is so widespread that many people are sometimes reluctant to consider alternative views. Nevertheless, its weaknesses are significant enough to stifle discussion.

One primary concern regarding medial views is whether the reduction to waves is effective in individuating *all and only* the class of sounds. As we notice above, ultrasound are not considered to produce perceptual stimuli for human ear because they overflow the audible limit of 20.000 Hz of frequency. An analogous are the ultraviolet rays which are electromagnetic waves not visible in the domain of colours. Nevertheless, both in the case of ultraviolet and ultrasound no radical changing in the nature of waves occurs. The only difference between sound and ultrasound resides in value of wavelength, which in the latter case is too short to be captured by human's perceptual apparatus. The same principle applies to infrasound. This leads to the question: should we consider these as sounds?

The complexity deepens when we consider other animal species. For instances, a cat's audible range goes from 55 to 77.000 Hz of frequency, while bats can hear up to 100.000 Hz. For these animals ultrasound produce auditory impression[12].

[12]While variations in the audible range across different species in the animal kingdom do not pose a significant challenge to medial theories, they do necessitate a *relativization* of the concept of 'sound'. This involves considering what is audible in relation to specific species.

In connection with the domain of light, the existence of *metameric* sounds poses a challenge to medial theories. Two sound are said to be *metameric* if they produce *identical* auditory perception, even if they have a *different* spectral distribution. Now, a medialist who considers sounds to be sound-waves could reasonably adopt a criterion of identity based on the frequency spectrum (along with length, volume, etc.). As a result, where the theory individuates two or more different objects, we can individuate at least one. The same could be said of any sum made up of a sound and an ultrasound (or infrasound), leading to a proliferation of overlapping entities. Metameric sound, on its own, does not necessarily lead to the refutation of medialism. In many fields, we find ontological differences that remain imperceptible without the aid of devices. On the other hand, metameric sounds highlight distinct *differences* between perception-based and acoustics-based criteria for identity. However, it's important to note that not all medial theories rely on a frequency spectrum for identity, as demonstrated in [48].

Another area of controversy within medial theories concern the nature of waves. Remember, medialists think that sounds are in the medium and we directly perceive sound-waves, but what is the ontological status of these physical entities? Obviously there is no straightforward answer. Because of the importance of wave theory in physics, the metaphysics of waves has undergone great development and refinement. However, an old-fashioned approach is to consider waves as either (1) *processes* (long-term events) or (2) non-eventive *individuals*.

1. As for the first branch of the dichotomy, medialism's challengers propose this argument: processes are said not to move [19]. Even though we are used to expressions like "the meeting moved from the classroom to the pub", we can hardly say that this is a pure instance of spatial movement. Rather, events, inasmuch as they are four-dimensional entities, are made up of one or more temporal slices, each of which takes place in a region. Therefore, instead of saying that the event E has moved, it is more appropriate to say that it has happened in a region L sufficiently inclusive to encompass all those sub-regions L_1,L_2, ... L_n in which slices of E have occurred. Now, mechanical waves are said to travel in space in a way that resembles genuine motion as much as the flow of a baseball from the pitcher to the receiver. Moreover, its participants (the mediums particles), who remain to some extent stationary during the transmission, are not affected by the direction the waves travel. So if waves are really moving, they cannot be considered as processes[13].

[13]This argument resembles the 'no-change objection' commonly raised against four-dimensionalism. To fully engage with this discussion, it's important to acknowledge that several responses have been proposed to counter this objection.

2. On the other side, sound waves are said to be non-eventive individuals as much as, for example, physical objects. Even if this path leads to a less intuitive description, it could at least preserve the motion of the wave. However this seems to not preserve the intuition that disturbances and perturbation happening in time as pure events.

In particular, both (1) and (2) state that sounds, as well as waves, travel through the medium. From a phenomenological standpoint, this could be rejected because it does not align with the typical content of auditory experience. In fact, sounds are commonly perceived at their source, not in the eardrum, not in the medium, and certainly not in motion[14]. This is precisely the objection advanced in [11], and O'Callaghan also presents a very similar argument. It is supported by the following example:

> Imagine a scenario in which engineers have rigged a surround-sound speaker system to produce a sound that seems to be generated by a bell across the room. this sound subsequently seems to speed trough the air toward you and to enter your head like an auditory missile [50](p. 29).

O'Callaghan concludes that such an auditory experience would be different from what we are used to. Setting aside this straw-man argument, what really emerges is perhaps that medial theories often overlook to the way in which sounds are initially presented to experience: that is, as stationary (with respect to the source) signals. At this point we can formulate this claim once and for all by calling it the *requirement of fidelity to auditory content*:

RFA A theory of sound should be true to the phenomenological content of auditory perception [11](p.14).

It's important to note that RFA does not strictly derive from scientific evidence; in fact, some might argue against the existence of such thing. Nonetheless, RFA represents a crucial condition for any comprehensive theory that aims to align with our primary phenomenological intuitions regarding sound. For this reason, distal theories are specifically formulated to maintain this claim.

For example, one might argue that change, including movement, is analogous to possessing different spatial properties at different times. For a comprehensive overview see [57](6.2).

[14]This feature of hearing is known in psychoacoustics as *extracranial location* [23], i.e. the tendency to locate sounds in a field that has the listener as its centre. If one imagines this described in the figure of a circle, the position of a source would be the conjunction of *direction* and *distance*, the former being a point on the circumference and the latter the radius.

Whereas O'Callaghan takes RFA as primitive in his theory, Casati and Dokic in [10] infer it from the representational power of hearing[15].

Obviously, the RFA is a very stringent requirement. It gives considerable importance to the phenomenological dimension. But what does it mean to violate such a principle? It means acknowledging that our perceptual system is *massively illusory* in regards to the location of sound. Such a discovery could radically change our representation of the sonic realm. To illustrate: the librarian's voice announcing the closure of the library spreads through the air and hits us on the surface of our body. This voice is now somewhere above us, among the bookshelves, yet we have the illusion that it is close to the loudspeaker, a few metres from us.

But if that is the case, does any theory that locates sound in the medium rather than at the source always contest the RFA? Not necessarily. In response, medialists have attempted to counter this trend in two ways: (1) by showing that medialism is *sometimes compatible* with RFA, or (2) by showing the *inconsistency* of the principle of locatedness itself.

Starting from (2), one possibility is to say that the way sounds come from their sources only provides information about their *direction*, not their *position*. This is the case with tactile and olfactory perception: we can feel the breeze coming from somewhere without receiving any information about its source. This attenuated version of locatedness even seems to be consistent with medialism, since the direction of the sound heard coincides with the position of the listener in relation to the propagating radius. Once again, however, this does not seem to be the way in which we perceive sound.

The other strategy (1) is to say that a particular medial theory is consistent with RCA. This is one of the arguments presented by Kwok in [39]. He maintains that even though sounds are compression waves, we may not be able to perceive their propagation. In order to reach this conclusion, he reports an argument that is formally analogous to the one used to reject medialism:

1. Water flows in the river;

2. if we look a the river from afar, we cannot see anything as moving in the way water flows;

3. therefore, if we look at the river from afar we cannot see the water in the river.

The conclusion of the argument is obviously false. What we can't see is not the water, but the flow of the water. In the same way, we could reasonably say that the displacement of waves

[15]Here's how the argument unfolds: if sounds were moving objects for nature, our representational apparatus would surely have described them in this way, since the mind has the power to represent motion for sound (as in the situation of a moving sounding source). However, this is not the case. Thus sounds, unlike waves, are not moving entities.

does not entail any relevant change in perception. Sounds are still moving and are perceived as stationary, it's just that the journey doesn't enter into the perception[16]. But what if we move closer to the river? perhaps the movement of the water will slowly appear to our eyes. Is listening the same as looking from a distance?

Whether convincing or not, this stance involves a diplomatic effort, both in affirming a medial position and negating a significant misperception in audition. Of course, this trend does not extend to the entire medial landscape. Many researchers also reject RCA. They argue that sound is made up primarily of tiny vibrations in the air, and secondarily of representations in our minds. How these representations are located in aural space depends on physiological and biological factors.

Overall, what can really lead us to assume the importance of phenomenology in the description of perceptual events is the confidence we place in one metaphysical address rather than another. For example, if we think that individuals are ontologically prior to the theoretical entities of physics, then we may adopt a *descriptive* ontological address. For a descriptive ontology, the conceptual scheme by which we explore and talk about the world must provide the elementary parts of our ontology. Otherwise, a *realist* (or revisionist) approach to metaphysics does not regard our cognition as privileged, but as one of many possible experiences. A scientific realist theory must capture the nature that lies beyond any possible experience. As I see it, the dispute between the medial and distal positions rests on the fact that sounds ground a descriptive as well as revisionist explanations. Each telling a piece of the story.

So far, we have seen some situations in which the idea of a location arising from a medial position on sounds is challenged. In addition, we have introduced some salient aspects of distal formulations, as the importance of experiecial data. In what follows we will concentrate on these aspects and on the main variants of distalism. Note that the medialist's point of strength as explanatory power regarding reflections, echoes, motion effects, etc. will return to challenge distalists in turn.

2.4 Distal theories

The last song of the A-side 'Stairway to the stars' has just finished. The headshell automatically returns to its resting place patiently awaiting our next move: flipping the vinyl record over. It was at this very moment that our mobile phone began to ring loudly. We scramble to recall where we

[16]This is only the first part of Kwok's position; the other half consists in directly challenging the distal assumption that sounds are heard as co-located with their sources [39](p.140).

left it, but thankfully, the ringtone guides us. We rise from our seats and traverse the room to find it on the corner of the table, precisely where the ringtone is coming from. Is this merely a fortuitous coincidence?

A distal theory holds that sounds are entities that *occupy the same position as their sources*. This philosophical attitude, like Strawson's aspatial theory, can't be traced back to any direct ancestor, but rather arose in response to the need for an account that would both give sounds a seat among the things that exist in the world and be consistent with the *requirement of fidelity to auditory content*. As previously outlined, theories attributing a location to sounds invariably ascribe particular ontological characteristics to them. Proximal theories, which locate sounds in the head or on the body surface of the listener, count them either as mental objects or as private events. On the other hand, medial theories, which consider sounds as identical or supervening on compression waves, treat them as relational or physical events. Similarly, the notion of sounds being located at their source yields distinct ontological implications. Distal theories can be broadly categorized into two groups: those equating sounds with *properties* of their sources and those regarding sounds as localized or relational *events*.

2.4.1 The property view

The concept of sound as a property of its source is rooted, in part, in Locke's epistemological doctrine. In his 1690 work, "An Essay concerning Human Understanding", he argued that sound, like colour and odours, belongs to the category of *secondary properties*: that is, the class of qualities that have the power to evoke a particular sensory change in the observer. Just as we are affected by the shades of red of the cabriolet driving along the road, e similarly respond to the sounds emanating from its engine. While these impressions are stimulated by the car, they are not inherent to its essence.

This position is in line with that advocated by Pasnau in [53], which belongs to the distal property view[17]. He posits that *sounds are properties of objects*, and objects *have* sounds just as they have colours. Therefore, the properties of the machine motor that produce a terrible noise are not certainly to be located in the medium, but probably in the carrier itself.

Several objections can be raised against this initial formulation. First, the ambiguity surrounding the ontological status of secondary qualities may ultimately lead to an undesired mentalist interpre-

[17]Even if scholars used to refer to this theory as the property view, in his article Panau oscilate in considering sound as *produced* by the source, as *belonging* to the source, or as identical to the vibration itself. More recently he has clarified his position by adopting the distal event view.

tation. It is not clear whether secondary properties are inherent in objects or not, or whether they are mind-dependent or not—and we are not going to go any further. Without this clarification, the property view is in danger of collapsing into a proximal picture. Additionally, the analogy between sound and color isn't entirely convincing. Commonly, we say that an object "emits" a sound rather than "possesses" it, contrasting with how we describe color, which appears to be possessed rather than emitted. Moreover, from a phenomenological standpoint, sounds are distinguished by temporal aspects that are absent in the case of colors, as Di Bona and Santarcangelo point out in [17].

A more structured variant of the property view is presented by Kulvicki in [38]. He suggests that sounds should be considered as a manifestation or display of the *dispositions* of objects. Unlike mere property, a disposition is a tendency to act or react characteristically in certain situations. For example, the fragility of glass signifies its propensity to shatter when appropriately struck.

Kulvicki analysis takes seriously the question of whether objects really possess sounds as much as colours. As he argue, the epistemological difference between colour and sound lies in their stability. While colours are relatively permanent—wall surfaces don't usually change colour unless repainted, and bananas in the fridge take a whole week to go from green to brown—sounds are ephemeral, lasting but a moment. But are colours really that stable? At second glance, our ocular impressions often change according to the quality of the light; a mountain seen at dawn is different from the same mountain seen at midday, the streets of Paris turn gold at sunset and yellow at night, in the absence of light, everything turns black.

Surfaces reflect when stimulated by light, but even in the absence of light we are inclined to say that a given object is still red, purple or grey. The stability of colours in our representation is not endangered by light or impression modifications. From this fact follows that colours can be considered as the *disposition* of bodies to reflect light in a certain way instead of individual reflection of light. Similarly, even if an object must be stimulated to produce a sound, it is far from saying it has no sounds as much as it has no colours in the dark. As Kulvicki affirms:

> Objects are disposed to vibrate in particular ways when thwacked. Interestingly enough, these vibratory disposition remain stable as modes of stimulation and contexts change. Any elastic object tends to damp out certain frequencies that stimulate it and resonate with others to differing degrees [38](p.4).

In support of this view, he points to a number of other analogies with the realm of light. The occurrence of 'thwacked' here assumes a technically restricted meaning: a *thwack* is in fact a short

physical stimulus that has the same intensity over the whole range of frequencies, i.e. a white noise. For this reason, a thwack may be more suitable for revealing the natural mode of resonance of objects than any other kind of stimulus, assuming a *normative significance*[18]. The sound an objects *has* is thus the one revealed when thwacked, even if the same objects can *produce* a lot more other sounds (just think to a loudspeaker).

Furthermore, he argues that our auditory system excels at recognizing similarities between signals. We can readily identify the timbre of human voices, even when factors such as phonemes, volume, and pitch differ substantially. As he states "auditory perception concerns not how some objects happens to be vibrating ant any given moment but rather how that object is disposed to vibrate across modes of stimulation" [38](p.11). There is at least a phenomenological interest in the dispositional view.

Possible objections to this theory arise from the same notion of the normative meaning of thwack. Objects can be stimulated in many (perhaps infinite) ways, each of which produces a different sound. Should we say that the objects have as many dispositions? Moreover, while light is the sole "thwaker" for colours, sounds are usually revealed by a miscellany of stimuli. Casati and Dokic refer to this argument as the *many-disposition problem* [11](p.42).

On another front, Di Bona and Santarcangelo, in [17](p. 99), state that within the current framework, we cannot explain what kind of entities sounds become once the disposition is fully realized. Kulvicki distinguishes between *audible events* and *individual sounds*, asserting that only the former really exists. But if audible events coincide with our perception of sounds, it follows that unheard sounds do not exist. In the end, this might aligns with the mentalist view that the existence of sounds is contingent upon their being perceived.

Notwithstanding the problems associated with the dispositional account, I think that Kulvicki is in the right spirit when he emphasises the recursiveness of sound events. Day in, day out, we encounter an enormous number of sounds, many of which we have already experienced and stored in our memory; and when it happens—to encounter a sound we already know—we recognise it in the same way as we recognise a person or an object. Aren't we familiar with the voices of our parents or the sound of a cembalo playing? Recognition plays an essential role in the perception of sound, and this is particularly evident when we look at recordings. But for now, let us move on to the second account of the distal position.

[18]The same normative significance regarding colour is ascribed to daylight, which approximates full-spectrum illumination.

2.4.2 Sounds as distal events

As we have seen above, those theories that consider sounds as properties of sounding objects do not seem to preserve the temporal aspect that is constitutive in audition—even though they capture a distal intuition. To address this, some philosophers, such as Casati, Dokic, and O'Callaghan opt to include sounds in the metaphysical category of *events*. We have already encountered events when talking about sound-waves, and we have said that a sound wave is a disturbance that *happens* in the medium. In fact, events, differently form objects, are said to *happen*, *occur* or *take place* rather than *exist*; they are said to have vague spatial boundaries but sharp temporal boundaries (beginning and end) and to be composed by temporal parts (at least one). Once an event has finished, we cannot encounter it again. Furthermore, as previously mentioned, distal theorists, aligning with Dretske's argument, maintain that events are static and cannot move. Common examples of events are sports games, parties, historical episodes, dinners, and perhaps sounds.

The earlier account of the distal theory of events is that proposed by Casati and Dokic in [10]. In the literature this theory is referred to as *The Located Event Theory* (LET). According to this theory, sounds are events that are located at their source and are identical to, or at least supervene on, the vibration in the source. LET seems to give a positive answer to many of the dilemmas discussed above. As far as RFA is concerned, this account preserves the idea, derived from an untrained auditory experience, that sounds do not propagate from objects. The loud ringing sound seems to be coming from the cell phone because it is exactly right there. The only situation in which a sound is said to move is when the sounding objects move, yet the event remains stationary with respect to its sources.

While a medium is necessary for vibratory events to reach our ears, it is not *essential* to their existence. If both the listener and the source are immersed in the air—as they used to be—the sound is *revealed* in auditory perception, otherwise the sound is not revealed but still exists. This theory is in favour of a kind of strong realism for sounds, preserving them in many situations where they wouldn't be audible. On the other hand, it could lead to the formulation of very exotic and controversial scenarios, such as the one described in [11]:

> Imagine a vacuum jar which has the property of immediately creating a vacuum upon closing the lid, and of immediately recalling air upon opening the lid. Take now a tuning fork at 440Hz and have it to vibrate, supposing that the vibration fades and become inaudible after 10 seconds [...] now place the tuning fork inside the jar and

> have it to vibrate as before, and open-and-close the lid of the jar, say once a bit less than a second. What do you ear? [11](p.30).

This formulation is structured variant of the *argument of the void*, often used as a testing ground for the tolerance of ontological positions on sound. Clearly, medialists and proximalists in such a situation will reasonably count as many sounds as medial disturbances or auditory impressions. The argument continues:

> ...you may have the feeling that a few short sounds, each feebler than the preceding one, comes into existence and pass away. But you may have as well the impression of a sound that is *revealed* by the opening and closing of the jar. Indeed the fading of the sound should be audible from each "window" to the next, implying that there is a single sound that fades (*ibid.*).

According to Casati and Dokic, the present scenario wouldn't force us to admit that multiple events have occurred, but rather that *multiple revelations* of temporal slices belonging to the same whole event have been heard. Both insights can be supported or debated, and the choice of party is obviously left to the sensibility of the lector.

A possible alternative that retains both the distal position and the status of events, but rejects the possibility of void-proof sounds, is that developed by O'Callaghan in [50]. Whether Casati and Dokic consider monadic events, such as the single vibration of the source[19], O'Callaghan thinks that the category of *relational events* is more tailored to represent sounds. In fact, we will refer to this theory as *Relational Event Theory* (RET), where a relational event that produces an audible impression is said to occur when a moving object disturbs a surrounding medium, causing the propagation of mechanical waves. Think, for example, of the melody produced by an open grand piano in a concert hall. Neither the vibration of the strings stroked by the pianist nor the waves filling the room are the sound itself; instead, the author says, the sound is identical to the transfer of energy that takes place between the surfaces of the chords and the portion of air around them.

For RET, the existence of a medium is a necessary condition, and this is the reason why O'Callaghan's argument with the jar cannot stand that the tuning fork still produces sounds with the lid closed. As relational, a disturbance requires both its *relata*: the sound source and the medium. In the absence of one of these, no sound is said to occur. The class of events with the property

[19]The choice to consider sounds as identical to the event source isn't included in the original version of LET, instead it has recently been adopted in [9].

'being a sound' selected by the relational theory is restricted with respect to that of the LET, which means that RET is less tolerant but perhaps more precise[20].

One might legitimately wonder: if the sounds heard do not correspond to the vibrations of the sources, what is the relationship between the two? Causality and instantiation are certainly the most intuitive choices. "However, O'Callaghan's purpose in [51] is to appeal to mereology and consider the sounds heard as the *constituent parts* of a whole, which are audible events. This choice is motivated by the circumstance that in multi-modal experiences sounds are perceived as parts of broader episodes:

> Suppose you stomp your foot on the floor in air. Your foot and the floor collide. Because there is a medium, they do so loudly and percussively. The sound. the sound you hear is particular event-like individual characterized by its audible features over time. Its part of a more complex occurrence that has aspects of features that may be inaudible. You may hear the more complex occurrence which include a foot stomping, a release of energy, vibration, a painful sensation, and a sound [51](p.396)

According to this view, sounds are heard as bound to their sources, together with an extended list of other non-audible events. To oppose this scenario, Casati, Dokic and Di Bona proposed a parsimonious variant of LET aimed at reducing the number of events at stake [9]. In this detailed version audible events are considered as monadic events, and although they may appear structurally complex, sounds are heard as unified entities, rather than shattered or fragmented. This sense of unity is better explained in terms of *identity*.

In conclusion this second account of eventive theory (relational events) enjoys all the positive aspects previously noted for the first: it still respects RFA and provides a reliable location of sound in the external world; it is still realist, but does not count inaccessible events as valid candidates for sounds. It furthermore takes into account some aspects that are relevant to the medial position. We will now consider some of the more interesting difficulties that both RET and LET would run into.

2.4.3 Challenging Distalism

The burden of the distal theorist is not only to show the correspondence between audition and external event features, but also to protect the accuracy of the theory from unusual cases.

[20]Note that this does not exclude the possibility for both variants that the medium and the source may coincide, as in the case of thunder or sounds produced by some brass instruments such as flutes.

The generation of *echo* is a first scenario that challenges the distal position. Suppose we are at the top of a mountain with a beautiful rocky landscape in front of us. We pump our lungs and shout "Hello!" and suddenly the greeting is returned by the mountains, which are, after all, very polite. The echo is the reflection of a wave from a surface or object, so that a weaker version of it is detected shortly after the original. The delay between the two gives an indication of the distance of the reflecting surface. Since the original sound and the weaker reflection appear to be heard at different locations, the case of echoes adds fresh fuel to the idea *that sound travels*.

For what concern LET, it is said that sometimes audition may fail in locating sounds as much as view may fail in locating object reflected by mirror. Both in the case mirror and in the echoes the responsibility for the perceptual mislocation resides in the deviant path of waves [11](p.25).

For O'Callaghan, the echo argument poses an additional obstacle for eventive theorists, namely the apparent re-encounter of the same events in experiencing reflection. According to the standard theory, vents occur only once, whereas objects can be encountered many times[21]. If sounds can be re-encountered, then they are not events. But are the original sound and the reflected sound truly identical? O'Callaghan argues that, despite the differences between the original sound and the reflected one, experiencing an echo is equivalent to hearing the original sound, albeit with distortions in time, place, and attributes:

> Hearing an echo, therefore, is reencountering a sound with the help of a special trick;the reflection of sound waves from a surface serves as the occasion for the trick. you perceive the primary sound event because its traces rebound and return [50](p.125).

It follows that the distinction between primary and reflected occurrence is entirely rooted in our perceptual apparatus and depends on the propagation of waves.

Surprisingly, the echo argument can also turn out to be positive for distalists. An interesting aspect is the seemingly casual relationship that accompanies echo experiences. This leads to the conclusion that both sounds (the original and its repetition) have the same source, which is very often true. Notably, the auditory system seems to be very efficient in identifying material sources that cause sounds, and reflection represents another positive case. Echoes are a subclass of the more common reflections, such as reverberation. Typical listening environments are usually rich in reflections—think for contrast of semi-anechoic environments or acoustically fixed recording studios. However, auditory system systematically masks spatial information of reflection occurrences.

[21] We are referring to Particularism. This framework will be explored in depth in 3.2.

In psychoacoustics this feature is referred to as the *precedence effect*: when a principal sound followed by short delayed repetitions passes through the ears, *no more than one event is perceived*, and the location of the subsequent repetition tends to be dominated by the first of the series. Instead, if the repetition arrives more than 50 milliseconds later, the delayed sound is perceived as an echo. Whether or not this feature can be considered a flaw in our system, for Pasnau the precedence effect stands in the way of accurate sound detection [53](p.313).

Another possible objection to distalism concerns the *Doppler effect* presented in section 2.3.2. As we said, this effect consists in a frequency shift due to the movement of either the observer or the sound source. Medialists simply explain this phenomenon by appealing to the notion of sound-waves. How can someone who denies that sound moves explain this change in pitch? According to both RET and LET, we are again faced with a *misperception* due to *physical interference*. The sound produced by the ambulance siren does not change in pitch as it passes us; it remains consistent with the way the ambulance driver perceives it. Such perceptual distortions do not reflect genuine differences in the world of sound.

Away from the sound effects, distalism may face a different kind of problem. To recall, we said that both for relational and located event theories, sounds are said to depend on (or be identical to) environmental events: a shattering sound depends on a glass or pot breaking into a thousand pieces, a thud depends on a huge body falling to the ground, and so on. But what happens when such sounds occur without the presence of the corresponding environmental event? As Nudds argues in [49], if one agrees that a thud can occur without a thud event, then there would be no need to consider event sources as carriers of sound properties. Rather, there would be something, the sound, that resides in the medium. Nudds support a version of medialism, namely *emanative phenomenology* [29], which affirms that the diffusion of the sound wave enters the perceptual experience of the listener. According to this idea, sounds are not only heard as coming from their sources, but they are perceived as emanating from and penetrating a region in which the source is included. To be persuasive, he presents the reader with two auditory scenarios that serve as premises:

S1 a collision occurs and we hear the sounds it produces;

S2 we still hear a collision, but no *ordinary* environmental event occurs.

If both S1 and S2 are correct, *then* eventive distalism does not capture the totality of sounds. The argument hinges on the notion of ordinary environmental events. For Nudds, collisions, thumps, clangs and rattles are ordinary phenomena, but the sounds produced by the loudspeaker are not.

Thus S2 would be a scenario that requires no hallucinatory experience, just a pair of speakers. He then extends the same reasoning to colour and presents an analogous pair of situations: first a red cube is seen (**S3**), and the same colour pattern is perceived without the presence of the host (**S4**). As he concludes, while S4 makes the conclusion false for colours—since we cannot see a red cube in the absence of a material object that reflects light—we can hear collision in the absence of ordinary environmental events. Thus the inference holds for sounds.

On the other hand, this argument relies on a vague use of the term 'environmental event'. Does this notion exclude loudspeakers as a standard circumstance of sound production? Not necessarily. The association between the sound of a collision and the physical vibration of two colliding bodies represents an epistemological fact. In reality, there are no structural distinctions between this vibration and the vibration of a cone driven by a magnet. It follows that in S2 there is still an environmental event, and both eventive accounts are preserved.

To reach the same result as Nudds, i.e. to show the inconsistency of the distal position, Maedows in [45] uses a different argument. He asks the readers to imagine a situation in which there is an auditory experience, but in which there is no distal cause of the sound wave. The scenario consists of an infinitely extended pattern of sound wave, such that the wave is present at every moment and in every region. The situation depicted is a purely metaphysical *possibilia*, but for Meadows it still raises doubts about the validity of distalism in more standard situations.

Although other interesting diatribes on medialism and distalism would be worthy of our attention, we should stop and move on to the primary focus of this dissertation: the problem posed by records.

So far, we have introduced the notion of perception and external location of sounds, and the way in which scholars have attempted to arrange a reliable picture of sonic entities. Then we trained ourselves to organise the debate using three of the most common positions on on the argument: proximalism, medialism and distalism. For each of these, we tried to shed light on their strengths and weaknesses. How does this introduction to the topic serve the functions? As we shall see, a considerable part of the work on recorded sound has been born as an appendix to the more general doctrine presented above. Others, not bound to any particular metaphysical stance, still make use of the same vocabulary. Then a presentation of the debate on sound is not only useful, but obligatory.

Chapter 3

Replaying sounds

3.1 An ocean of (recorded) sounds

Suppose that we are going to attending the concert of one of our favourite artists, eager to immerse ourselves in the familiar tunes we love. Presenting our tickets to the theatre staff, we're guided to our seats, and the excitement intensifies. After a long waiting the venue of darkness stand for the imminent beginning of the show. As each musician takes the stage, a river of applause start heating the air. The performance has begun. Once it is finished, still enveloped in the afterglow, we decide to approach the man behind the mixing board. With the artist's blessing, we ask for a copy of the mix and he transfers the concert's audio, captured by the stage microphones and expertly mixed, onto our USB key. Later, at home, we decide to recreate the enchantment with our Hi-Fi system. The playback quality is nothing short of stunning. The recorded sounds mirror the live performance so accurately that it feels as if we're *reliving* the entire experience.

Up to now, we have introduced the salient points that the metaphysics of sounds has committed itself to in recent years. While we haven't covered every perspective, such as Strawson's thesis, it's worth noting that in the course of this chapter, we will introduce some alternative views that are strictly connected to the concept of sound reproduction.

To sum up, in the last chapter, we presented three main families of theories: *proximalism*, *medialism*, and *distalism*. Each of these theories takes a position on the location of sounds in space and, consequently, on their nature. Proximalism tends to consider sound as an object of the mind, strictly dependent on the perceiving subjects. Medialism, capturing the standpoint of mainstream acoustics, considers mechanical waves happening in the medium as the main candidates to bear acoustic qualities. Lastly, distalism considers distal sources as the object of audition, as it is the only way to give credibility to a phenomenological account of perception.

We have also seen that this division is compatible with several philosophical stances. For

example, both *localized* or *relational events*, *secondary qualities*, and *dispositions* are possible interpretations that share the distal intuition, just as medialism is not necessarily involved in considering *waves* as genuine events. However, as I hope to have conveyed, events seem to be the main candidates for capturing the intrinsic *temporality* of audition. From the observations of the physical sciences, which consider sound entities to be mostly measurable and describable in terms of duration, periodicity, and behaviour in time, to empirical and phenomenological evidence, we have good reason to resist the emancipation of sound from temporality [17](ch.4). Any attempt to the contrary is faced with the task of somehow justifying this emancipation by proposing a model of epistemic access to sounds in which temporal factors are *subordinate*. From this perspective, the prompt of the following argumentation would be compatible with *any theory that assumes sounds under the label of events*, whether distal, medial, or proximal.

On the other hand, alternative perspectives direct our focus toward specific phenomena that emerge in auditory experiences, challenging the compatibility of standard eventive interpretations. As demonstrated in the case of *echoes*, and as we will explore further in the context of recorded sounds, these scenarios play a crucial role in elucidating the ontological status of sonic entities.

Recordings——If we pause for a moment and reflect on how much these things shape our everyday lives, we will easily discover their ubiquity. In fact, audio playback is one of those things with which people living in the twenty-first century are very familiar. From the stunning experience of reliving a musical performance in our living room to the annoying ringtone of the digital alarm loudly plying every morning, both cases are the fruits of a human technology. Since the invention of the first sound reproduction system, often attributed to Edison's phonograph (1879 ca.), those "artificial" phenomena have been slowly creeping in the worldly sonic front as much as today we may have difficulties to recognize for any sounds though is a genuine environmental noise. The impact of this invention has been enormous in many sound-related fields, first and foremost the *music*. To borrow the words of Kania:

> The advent of broadcasting and recording technology brought a sea-change in the standard situation in which music was heard. Where, before, music was rarely heard in the absence of musicians producing it live, now one could listen in one's living room to a performance that was actually going on thousands of miles away, or, stranger still, one that was already finished, and one could listen to the latter kind over and over again [32](p.22).

But the importance of recordings is not limited to the aesthetic realm. Human activities benefit

from this type of technology in many different ways. As a tool for learning new languages, for preserving historical evidence over time, for communicating with people on the other side of the world, for tracking the acoustic wellness state of wild ecosystems[1], and so on. In short, a world without recordings surely would be less noisy, and perhaps less interesting.

3.1.1 Philosophically speaking

From a philosophical standpoint, the very existence of this type of object raises several questions. For example, one may ponder the nature of sound recordings. Generally, a recording is the *capturing* of an audible signal for subsequent reproduction. In everyday language practice, we use the word 'recordings' in a dual manner, to refer both to something that *can be heard* (an audible phenomena) and to something that *cannot be heard* (Discs, tapes and Digital files). To disentangle this ambiguity, we should divide the *product of recordings*—namely, the specific pattern of sounds reproduced by a loudspeaker—from the *medium of transmission*, which is the object encoding the sound (namely, track or sample)[2]. While only the former can be considered the proper bearer of audible qualities, the latter constitutes a distinct type of object, more akin to a musical score or mathematical graph.

Naturally, the product and the medium of transmission in recordings are closely connected. For a recording, in the sense of a product, a medium of transmission is necessary, but it does not require a specific type of encoding or format. In other words, the same pattern of sounds can be inscribed on a potentially infinite number of supports. Moreover, the process of encoding, which depends on the tool used, is often reliable in capturing the relevant information from the original event. In one word, recordings are faithful.

In this chapter we focus on discussing the primary meaning of recordings: the *reproduction* of specific *sound events*. In accordance with this initial meaning, listening to recording generates auditory impressions. Similar to environmental scenarios, sounds heard in recordings can be characterized by audible qualities. For instance, when the sound of our morning alarm is played, it possesses specific attributes such as volume, a structured pitch progression, and appears to came from a particular position in the space. Many account of audition could align with this descriptive model. Simultaneously, from a certain perspective this experience might be thought as *illusive*.

[1] I'm referring to the work of Bernie Krause, one of the pioneers in the study of ecosystems from a sonic perspective. This discipline, known as *Soundscape Ecology*, establishes a taxonomy of sound sources (biophony, geophony, and anthropophony) and provides a detailed analysis of the impact of artificial noises on natural environments.

[2] This distinction is discussed in [31](p.403). It is worth noting that the medium of transmission should not be confused with the medium of propagation. The latter, introduced in 2.3, indicates the elastic material in which mechanical waves propagate and is not related to audio reproduction devices.

Consider the initial scenario as a paradigmatic case. When we reproduce the concert in the comfort of our house, undeniably, a new sonic event takes place. However, our experience suggests that the sounds produced by our loudspeakers are the same as those that occurred in the theatre earlier. Remarkably, even though the performance originally took place in the past, we are currently listening to it, and, while it once unfolded in a concert hall several kilometers away, it now takes place precisely in our living room. Despite our familiarity with this phenomenon (we may not be as astounded as people in the early twentieth century), it remains inherently fascinating. In the experience of recordings, lending an ear to the past, we feel like encountering the *same individual sounds once again*.

This becomes even more evident when examining how we typically articulate identity statements in the context of recordings. For example, in the field of music production, engineers readily acknowledge listening to *the exact same tracks* repeatedly. In crowded environments, such as when we receive a call, recognizing the distinct sound of the caller's voice may pose difficulties. Nevertheless, despite challenges in intelligibility, we have no doubts that we are indeed hearing *the very voice* of the person speaking. On the other hand, it's reasonable to suggest that when discussing recordings, we refer to 'identical' sounds in a somewhat loose manner, akin to how we might consider an original canvas and its forgeries as the 'same' picture. Relying solely on linguistic evidence does not adequately address the concept of sound recurrence. Therefore, a more substantial exploration of sound properties is necessary.

Transposing this intuition into the field of sound studies, recordings exhibit a familiarity with a previously analysed acoustic phenomenon: the *echo* (2.4.3). Echo experiences present a challenge for both LET and RET, not only in suggesting the idea that sounds travel in space but, as noted by O'Callaghan, in the belief that echoes entail the *re-encountering* of the same sound.

From this perspective, recordings and echoes share similarities, yet they differ in the following ways: (*i*) while echoes are *natural* environmental phenomena whose existence depend on the conformation of the space in which the wave propagates, recordings are the result of human technologies which not require such condition; (*ii*) in the case of echoes, we may perceive a *causal relation* between the original event and the secondary reflection, a connection apparently absent in the case of recordings[3]. In support of this last point, we can argue that in the case of records, the original and reproduced sound may be separated by a very long period of time. This is the case

[3]For an account of the perception of causal relations in audition, see [4]. For the sake of argument, we should note that in this article, Di Bona defends the existence of a causal connection entering auditory experience, with the relata being the sound and the material source. However, the way in which she presents this in the diachronic case, in my opinion, is extendable to the case of echoes and short reflections as well.

when a tapes containing raw tracks of famous rock band are found in someone's cellar. However it's essential to note that the apparently lack of a causal wire doesn't exclude the existence of a whatsoever connection between the original signal and subsequent repetitions. Based on these considerations, we can assume that, regarding the possibility of encountering the same sound twice, recordings present a *strengthened version* of the argument derived from echoes.

As seen in the case of echoes, the possibility of the same whole event occurring more than once is ruled out by a standard metaphysical account of events. Events are usually individuated by *temporal indexicals* (like "yesterday" or "two hours ago"), *terms* (such as "the 22nd of February"), and *temporal windows* ("from 3 to 5 p.m."), in a manner similar to how physical objects are individuated by referring to the portions of space they occupy. Furthermore, boundaries, like "end" or "beginning", behave precisely like physical edges for objects such as bricks or bottles. A football match exists during its proper temporal path, neither before nor after.

If events are construed as *particulars* or *substances*, they inherently lack the capacity for *recurrence*; in other words, they cannot transpire within a temporal framework different from that defined by their boundaries. To elucidate, a singular concert, for instance, cannot unfold more than once, despite the occurrence of numerous remarkably similar concerts before and after. In essence, this uncertainty is what characterizes our intuitive sense of sound re-encountering when listening to recordings.

3.1.2 The same once again

Whenever particulars, such as events or objects, exactly occupy two disjointed spatiotemporal regions, a certain phenomenon—namely, *multilocation*—is at stake. In essence, since recordings suggest the possibility of encountering the same particular sound more than once in a manner much stronger than in the case of echoes, *mutatis mutandis*, they introduce a degree of inconsistency into the eventive account of sounds. Some authors, like Martin, advanced precisely this claim:

> Auditory reproduction gives us a connection to the past event, not just through using a mechanism to produce a recognizably similar sound, but by managing to give us the original sound once more [44](p.445).

Taking seriously the possibility for sound to return into existence, Martin challenges the ontological category of events as unrepeatable individual occurrences in representing the sonic world. Instead, sounds might be likened to more *abstract entities*, such as fictional characters or works of art, for which the possibility of recurring is allowed.

On their part, event theorists should not be entirely unprepared to face this challenge. The idea of sounds as repeatable entities hinges on the assertion that both encounters with the original and its reproduction *constitute experiences of the same particular*. Supporters of the event theory, however, might argue that, beyond our initial intuition, recordings do not entail genuine cases of identity. To substantiate this, they can readily point to an evergreen piece of evidence: two sonic events are at least *numerically different*. Furthermore, they might argue that recording playback resemble original sounds only in an impoverished way. For instance, when listening to the concert's recordings, despite its similarity with the original performance, the former seems to lack reverberation due to the close position of the microphones to the musicians. They could also note that during the live performance, a change in the listener's position results in a shift in perception, which does not occur in the case of recordings. In a nutshell, they might try to reduce the number of *relevant attributes* shared by those particulars.

In defense of the repeatability account, one might emphasize that highlighting differences between the original and reproduced sounds could be an inefficient strategy. Consider again the initial scenario of a recorded concert. Since the original sound and the reproduced one constitute two distinct sonic events, how can we justify the idea of *acquiring information about the former* simply by listening to the latter? For those who believe that sounds recur, our epistemic access to the original signal is preserved. Indeed, every auditory experience that evokes the idea of recurrence is grounded in the process of individuating and re-individuating the same object twice. However, we can distinguish at least two ways of understanding the phenomenon of recurrence: a *narrow* or strong one and a *broad* or weak one. Regarding the strong account of recurrence, for two sonic events separated in time to be identical, it entails that they share every acoustic or audible property. This could be the case, for example, in a certain sequence of chess moves, such as an instance of a French Defense, which could be easily executed many times while preserving all properties relevant in the context of a chess match. Similarly, the reproduced sound is *indiscernible* from the original in a strict sense.

Nonetheless, this first formulation appears hardly attainable for sounds for a couple of reasons. First, from a purely acoustical standpoint, the possibility of having every factor under control during the process of sound production is nothing more than an ideal scenario that rarely meets reality. For instance, the behavior of a vibrating source is strictly variable depending on temperature and other chaotic factors. Brass instruments, in general, change their timbre when heated; guitar bodies expand in humid environments, resulting in a change in tuning. On their part, loudspeakers

respond differently to signals based on their temperature. It follows that every attempt at reproducing vibrating processes results in a sort of *approximation* to the original.

Additionally, from an epistemic perspective, the way in which we acknowledge recognizing sounds occurring more than once seems to exclude such a narrow idea of recurrence. Indeed, sound individuation is more akin to the way in which we recognize a person rather than a chess opening. An illustration of this could be the following.

Mendez-Martinez, in his exploration of the possibility of recurrence for sonic events, introduced the concept of a paradoxical account of individuation namely *Funesian individuation*. Inspired by Borges' "Funes, the Memorious", he envisions a subject with the ability to memorize every minimal detail of each object and happening encountered in his experience. However, this extraordinary power is balanced by the impossibility of formulating identity statements due to the vast number of variations observed in each encounter. Such a subject could be "unable to tell that the dog seen at 3:15 and the dog seen at 3:30 were the same" [46](p.178).

In an auditory context, this example argues in favor of a broader account of recurrence. That is, despite certain *audible variances* between the original and the reproduction, we often recognize the very same sound occurring once more. For instance, even if during phone calls speeches assume a hollow and tinny timbre, one could genuinely recognize the very same voice of the caller. Yet, Funes, the character depicted by Borges, would not. Negating this intuitive evidence might come at the cost of ascribing a significant amount of deception in cognition[4].

Assuming this concept of *recurrence* for sounds, merely emphasizing the differences between the original and reproduced sound would become trivial and lose its point. Instead, what becomes significant in this narrative about recordings is the revelation of elements that are genuinely preserved during the processes of reproduction, evoking the sense of experiencing the same once again.

Returning to the starting point, one might rightly ask: what is at stake in statements about re-encountering the same individual sound? Which factors contribute to sound recognition? To what extent can a sonic entity undergo change? Are there other modalities of cognitive access that do not necessitate encountering the very same sound once more? If a supporter of the event theory wishes to sidestep the pitfalls of a Funesian individuation, as is advisable, they must furnish a precise account of which audible attributes are relevant for sounds to be deemed the same and under what conditions such properties are preserved. In short, a law of identity for sonic events is

[4]A move similar to the one against the medial position in 2.3.3.

crucial.

Departing from this naive framework, defenders of the event view have explored alternative paths. This is partially driven by the fact that the literature already offers well-known, despite possibly controversial, laws of identity for events, such as the one proposed by Davidson [15]. Since sounds are considered as events, one might question the need for a different set of conditions.

Indeed, rather than establishing conditions for original and reproduced sounds to be identical, authors like Dokic and O'Callaghan have attempted to solve the puzzle of recurrence by introducing the notion of *mediated epistemic access to past events*. They propose encountering the very same sound with *distortions of space and time* and justifying the resemblance in terms of *representation*, drawing parallels between recordings and images or video.

In the following part, we will delve into theories that have attempted to address the problem of repeatability within a standard account of events. The next section will present in detail the general motivations that lead to an ontology of events and introduce some contending approaches aimed at explaining the phenomenon of recurrence. These approaches, departing from the notion of the particular, are almost universally endorsed today by those who posit the possibility of re-encountering sounds. In this context, we will provide a detailed presentation of the formulation of recurrence for sound proposed by Martin and explore the questions it raises in the context of sounds. Section 3.3 will focus on reasons stressed by supporters of the distal event theory for considering recording experiences as consistent with the framework. With Dokic, we will explore the possibility for sound to be represented or depicted, in a way similar to how sculpture and photographs depict their subjects. Then, with O'Callaghan, we will introduce a solution to the puzzle of recordings based on perception. By the end of the chapter, we will have acquired a structured idea of sound recurrence, which will be sufficient for introducing further investigation.

3.2 Toward An Ontology of recurring sounds

The existence of recordings introduces a unique problem in the current landscape by suggesting the *possibility of recurrence* for sonic entities. This has notable consequences not only in the way we represent audition but also in the way we conceptualize the very *object of audition*. We initiated this discussion by questioning the nature of sounds and determining the best ontological category to describe them. After considering various perspectives, we particularly focused on the notion of event, shared by both medial and distal perspectives, as an objects that unfold in and takes time. However, events, for compelling reasons, are conventionally thought of as individual occurrences that do not repeat. Mixing all these elements, we might find ourselves goaded to reconsider the entire view to make it consistent with the experience of recordings. Unexpectedly, vinyl, groups' albums, radio broadcasting and MP3 files become intertwined with a serious philosophical problem——the problem of recurring events.

Events have always been a matter of interest, but only in the second half of the last century did they receive a substantial discussion in metaphysical terms. The account we are referring to as the 'standard account of events' is the widely accepted view known as *particularism*. Particularism is a general position in metaphysics, and when applied to events, it produces the claim that statements about kissing and strolling should be paraphrased into propositions quantifying on unrepeatable entities.

Particularism is a common ground for both *fine-grained* and *coarse-grained* principles of individuation. The former tends to consider events as unified as ordinary objects, while the latter counts as many non-reducible entities as possible descriptions of the same events [12](p.20). However, particularism does not *automatically necessitate* the introduction of event-like entities in one's ontology; it merely imposes a condition of *unrepeatability* that must be satisfied by terms in propositions representing *prima facie* event-words. Theories like those of Goldman or Kim are particularist in exactly this sense.

A noteworthy account that combines both the condition of unrepeatability and the claim that events constitute *bona-fide* entities is the one proposed by Davidson [14][15]. Davidson argues that there are several compelling reasons for introducing events into the catalogue of being, primarily to avoid fallacies in *natural language analysis*. Let's briefly reformulate his position. We are all familiar with statements like:

1. "John kissed Mary"

2. “the alarm rang at 7:30 this morning”

Apparently, these statements don’t force us to assume the existence of more entities than John, Mary, and the alarm. For instance, (1) could be analysed by taking a two-place predicate *K*, namely ‘kissing’, such that:

3. K(*John*, *Mary*)

However, Davidson resists this strategy because it does not preserve the possibility for certain event-like statements to be logical consequences of one another. For example, if one applies the same analysis to (2), the result is a monadic predicate *R* which stands for the property ‘having rung at 7:30 this morning’, such that the alarm falls under this predicate. What immediately stands out in the latter analysis is that the predicate *R* is “richer in information” than *K*. It not only describes an happening——the ringing of the alarm——but it provides further details, namely specific temporal restrictions on when it occurred, narrowing the scope of the predicate. This increase in richness is known as *adverbial modifications*. The intuition behind adverbial modifications, as Davidson argues, is that any occurrence of a predicate like ‘having rung at 7:30 in the morning’ or ‘having rung loudly’ necessarily entails the non-modified occurrence of ‘having rung’ *simpliciter*. However, by adopting an analysis like (3), such entailment can only be preserved by considering predicates within the domain of quantification——second-order logic.

The strategy proposed by Davidson is to consider event-like statements as propositions *counting* on actual events rather than ascribing properties to participating objects. Following this line, statements like (2) express the existence of an event e, having the properties of happening at 7:30 this morning, in which the alarm participates.

4. $\exists e(e \text{ is a ringing} \wedge \text{participate}(alarm, e) \wedge e \text{ happens at 7:30 this morning})$

This shift from specifying predicates to attributes of events is precisely what preserves the intuition of co-extensionality of adverbial modification. Moreover, it does not require nothing but first order logic plus identity. Applying the schema in (4), the non-modified statement like “the alarm rang” results as a logical consequence of (2), despite any further modification. Reasons for defending an account of event entities primarily derive from the ontological commitments of natural language.

3.2.1 Sameness and similarity

Regarding the adoption of the unrepeatability condition, Davidson and others argue that particularism should be preferred for a variety of reasons. They first account for the pre-analytic intuition that events resemble physical objects in being *linguistically individuated*:

> We can speak about the button, as well as a button, about there being many buttons; and we can, standardly, count buttons. Similarly, we can speak about the button pressing, as well as a button pressing, about there being many button pressings, and we can, standardly, count them. We assign proper names to physical objects, conventionally to all persons, but also to other physical objects, names such as 'The Empire State Building,' 'The Brooklyn Bridge,' and 'Mt. Everest'. Similarly, we assign proper names to events, names such as 'The Battle of Waterloo,' 'The Civil War,' 'Typhoon Shirley,' and 'Superbowl VII' [5](p.134).

On the other hands, particularism might be desirable even for its *ontological parsimony*. While accounts that consider recurring events as instantiations of properties introduce a high-level distinction between individuals and universals, along with a relation of instantiation from the first to the second, particularists need only individuals. In the case of Davidson's theory, these individual events are sufficiently distinguished from enduring things by their non-physical nature[5].

However, when dealing with statements about recurring happenings, there may be only the illusion of simplicity. For instance, it can be argued that ontological parsimony comes at the cost of the loss of intelligibility. To grasp this point, we need to observe how Davidson proposes to analyze statements in which the same action is performed *more than once*. Let's introduce a detail into statement (1):

1.1. "John kissed Mary three times"

As in preceding cases, we may think that this statement too expresses an occurrence of adverbial modification. Thus, the best analysis should convey something like, "There is an event, namely the kissing, in which John and Mary participate, and this event has the properties of having occurred three times". Instead, as the unrepeatability condition imposes, individuals cannot occur more than once; then the property of 'having occurred three times', referred to events, expresses an *empty*

[5]Furthermore, events are said to be, to a certain extent, dependent or posterior (on a scale of fundamentality) with respect to physical objects. See [41].

predicate, and the whole proposition is made false. Since we don't think that propositions like (1.1) refer to either paradoxical or vague states of affairs, we should desire an account capable of displaying their truth-condition.

To overcome this difficulty, Davidson's purpose is to express the number of occurrences of similar events by introducing *as many existential quantifiers*, each one pointing to exactly one individual event. This sentence-schema is to be understood as entailing that there is a property that x has but that y lacks, that is, that x and y are distinct. In fact, as he thought, recurrence can be better explained in terms of *resemblance* rather than identity, and resemblance is nothing more than belonging to the same *event-type*. Even though each one of the kisses is a particular, unrepeatable happening, they all are tokens of the same kissing-type.

Turning back to ontological parsimony, one might note that even particularism points to a very crowded domain. Think about statements on frequently happening events like "the sun rises every morning". For a particularist, it takes a boundless number of variables to represent such a proposition, whereas endorsing a universalist account of events only one would be sufficient. Davidson, replying to critics regarding recurrence advanced by Chisholm, expresses his mistrust toward this very criterion of parsimony:

> Clarity is desirable, but parsimony may or may not make for clarity. Of course, once one has a viable theory, it is interesting to learn that part only of the ontology one thought was needed will suffice; but such reduction must come after the provision of a working theory. One may also have intuitions that suggest what entities it is appropriate to call on the interpretation of a given stretch of linguistic territory. I feel myself that to summon up classes or universal to explain 'I did it again' is using a cannon to shot a mouse [14](p.27).

What looms in the foreground of these sharp and bitter responses is that economy cannot be the motive to give up a dense and uniform theory of events. Parsimony still is a criterion, but only when the choice happens between theories with the same explanatory power, soundness, and richness.

Aside from ontological parsimony, Davidson's account of recurrence has been criticized from two distinct fronts. The first is an applicative one, focusing on whether this kind of analysis will work in all cases. Davidson himself considered some hard cases for his theory[6]. For example, statements in which two subjects are said to perform the *same action* (e.g., "John shouted and Smith

[6]Some of them are collected in [6] pp.97-98.

did the same thing") require special treatment. A careless analysis might fall into the fact that both subjects *cooperate* in performing an individual action, which is not what often is meant.

The trick proposed by Davidson is to distinguish all the relevant aspects in which events may be similar (for example, whether John shouted at his boss, Smith did it to his mother). However, this is not always so simple. Consider the following sentence:

5. "Sarah danced the waltz eight times"

Davidson suggests paraphrasing it into a proposition committed to eight variable ranging over individual events, each of which is a token of the waltz-type performed by the dancer [13]. But the truth condition of this paraphrase seems to be different from the one of (5). In fact, the former remains true even if Sarah danced *eight different* waltzes, while the latter seems to count only the number of similar dances. The difficulty here consists in individuating those aspects that ground similarity. In the words of Brandl:

> When are two particular dances similar enough so we can count them as dances of the same waltz? The similarity here might consist in nothing else but the intention of the dancer to do the same thing again. The particularist owes us an explanation of what intending to do the same thing means here [6](p.98).

This consideration leads us to the second order of critics, that is, what it means for two events to belong to the same type. As argued in [5], the problem of type instantiation in events can be unfolded following the same problem in the domain of objects.

Generally, types or sorts are *collections of objects* having *more than one member*. They are called 'natural kinds' when referred to non-artificially class of things. Examples of natural kinds are plants, metals, or physical particles. Types are evidently distinguished from sets: they cannot be singletons or null; furthermore, they don't have an identity defined in terms of membership[7]. Types are even different from *properties*. Like sets, property can have one or zero instances. In addition, some properties, like the one of being self-identical, seem to be universally distributed. In the case of sorts, no universal instantiation occurs.

Given such a definition, the challenge is to furnish an account of sameness for particulars based on membership in the same type. A strategy could consist of stipulating all the sorts instantiated

[7]They can gain or lose members while remaining the same. Furthermore, sorts with the same extension can be distinguished. For instance, the sort unicorn and the sort elf, though not containing any actual members, refer to different things.

by certain objects. As is widely understood, things often belong to more than one type at the same time. However, for two objects to be similar in a respect, they just have to share one *narrow sort*, and the same is true for events ([8]. For instance, in (5), all the waltzes were similar in belonging to the narrow sort, or type, 'waltz of flowers' from Tchaikovsky's ballet "The Nutcracker." Yet, we have not enough information to indicate that each dance instantiates the same 'waltz of flowers' type. Maybe the first was slower than the second, the fourth clumsy, and the last surprisingly precise. In conclusion, one can object that the dancer has not performed the *very same action eight times*.

Brand's perspective implies that the sameness of type is a *context-dependent* notion, requiring judgment based on the information available to us. This introduces a potential element of vagueness and may not ensure soundness to Davidson's account of resemblance. While achieving context-independence is not impossible, it would demand an extensive set of information, diverging from our ordinary way of evaluating such matters.

3.2.2 Abstract sounds

Returning to the realm of sounds, everything discussed about the standard account of events also applies, with necessary adjustments, to recordings. When experiencing a recording, we might feel like we are encountering the very same sonic event once again. Statements like "my ringtones sound the same as yours" or "I've listened to Bill Withers' album 'Just as I am' three times today", rooted in those experience, refer, according to a Davidsonian perspective, to the occurrence of numerous *unrepeatable events*, each instantiating the same, possibly complex, type.

However, the experience of acoustical reproduction, under certain circumstances, can be so *faithful* that we cannot distinguish between the original and the reproduced event, for not mention several reproduction of the same media. In many cases, signal capturing and reproduction systems have been developed to achieve this result. This aligns with the initial idea behind the cult of High Fidelity streaming—the possibility of reproducing the exact sounds of a concert through our speakers, not missing a single note. Such experiences reinforce the notion of re-encountering sounds in our minds which might play a role against the particularist approach on sounds.

In light of these considerations, some scholars have proposed alternative views to align with the experience of sound recurrence. These theories notably diverge from the predominant eventive

[8]Being narrow is the property of being a non-reducible sort. Given two sorts F and G, "F is narrower than G iff, for any x at time t, if x is of the sort F during t, then x is of the sort G during t, but not conversely" [5](p.141). For clarity, the sort 'olive' is narrower than the sort of 'plant' because the second is a specification of the first.

conception, including located and relational event theories, by regarding reproduced sounds as non-eventive repeatable entities.

As we seen, a viable approach to reconcile with the phenomenon of recurrence involves *conceptualizing sounds as properties of things*. This perspective has been explored extensively within distal accounts (2.4.1). Scholars like Pasnau and Kulvicki have proposed that sounds are manifestations of properties or dispositions, emphasizing that audition reveals aspects inherent to the source rather than the medium of propagation. This view aligns with statements expressing the possession of audible features by resonating objects.

This perspective can also be examined in light of Montague's standpoint about repetitive happenings [47]. Montague posits that certain events, occurring frequently, are better understood as *properties of time* rather than particulars. He introduces the concept of *instantaneous generic events*, events happening at a specific moment, such as the sunrise[9]. These events can be considered as properties of specific moments or short time interval. Properties, as universals, are supposed to be exemplified more than once, making them capable of recurring. However, the difficulty in accepting such a view resides in precisely *locating* these events. While properties inherently belong to objects, determining whether they are located in those objects poses a complex question. The conflict arises from the idea that properties, as platonic universals, may exist *separately* from spatiotemporal reality, akin to mathematical entities and concepts. This perspective challenges our intuition that events are located where their participants are.

A more recording-oriented proposal comes from Martin, also endorsed by Nudds. This perspective involves considering sounds as *abstract individuals* rather than particular events. As he think recordings connect us with original events, and since this connection is made by producing once more the same sound, it is not a mediated or deceptive source of knowledge.

In [44] Martin explores the possibility of creating the auditory equivalent of *visual images*[10]. According to his view, the concept of images itself belongs to the visual domain. Indeed, as he argues, vision stands out among the senses by presenting material things as objects of awareness, a characteristic not mirrored in audition. While instances of pure *visibilia* exist——representing visual experiences of features like color or shapes not mediated by physical objects, *e.g.* rainbows or shadows——these are not the primary objects of vision. Physical surfaces remain the proper

[9]Instantaneous generic events are distinct from *instantaneous individual events*. However it seems that for Montague it is not always clear how to individuate such individual events and how to chose for an events for its being individual or general. See [6] p.99.

[10]the concept of sonic image will be adequately introduced in section 3.3.1.

focus of what we ordinarily see. Unlike vision, Martin argues, sounds are the sole carriers of sonic features, and material sources only enter audition secondarily[11].

The possibility for visual images to represent something lies in their presentation of properties they do not *exemplify*. This condition, often referred to as the twofold condition, contrasts with sounds. For a sound to convey the appearance of having the properties of another sound, it means exemplifying those properties and violate this twofoldness. Consequently, sounds lack a correlate to visual images.

In Martin's view, the playback of a recording is not an image of the original sound. Yet, he claim that listening to a concert in our living room *provides us access to the original acoustic phenomenon*. The connection to past events in audition does not involve representation in the way photos or videos do; thus, the connection must be carried in a different way. Recordings do not represent a similar reproduction of the original because they give us the very original sounds once more. In light of these thoughts, Martin proposes considering sounds as *abstract particulars*:

> First, if sounds as individuals can be reproduced, then they fall in the category which P. F. Strawson labelled 'abstract particulars'. Unlike concrete particulars, there is no requirement that they have a unique spatial location at a time, nor yet that they trace a continuous path through space over time. The sounds that Florence Nightingale made in July 1890 were recorded on that occasion, but did not persist through time until reproduced by users of the Wellcome site in recent years. So the sounds have an intermittent existence. If reproduction is reproduction of the original sound, then that sound is multiply located wherever the sound is reproduced [44](p.345).

Martin's argument centers on the idea that abstract individuals, unlike universal properties, *maintain their identity through reproduction*. However, the manner in which a certain sound is considered a copy of the original may be questionable. Sounds, in his view, seem to exemplify seemingly contradictory properties of being particular and being multiply instantiated. In essence, reproduced signals posses a structured or layered identity.

He observes that sounds rarely exhibit qualitative indiscernibility. Even in reproduction, certain properties belong to the actual sounds rather than the original physical events. For example, a playback of a speech preserved in the grooves of an old vinyl may sound crackly and distorted,

[11] This idea is often referred to as the *sonicist* account of sounds, primarily presented by Scruton. In this perspective, sounds are considered still as events but detached from any material source. To characterize this idea, Scruton defines audible events as pure events or secondary objects. Recorded experiences are exactly those in which the material source is completely irrelevant in defining the particular sounds (see [55]).

unlike the original voice. Moreover, Martin suggests that through multiple reproductions of the same copies, "distinctness of individuals can be grounded" [44](p.346), implying more than mere numerical distinctiveness. On the other hand, the identity of sounds as abstract depends on a *suitable chain of reproduction*. Two seemingly identical songs do not instantiate the same abstract identity unless they are copies of the same original template or copies of each other in the appropriate way.

According to Martin, this duality is the basis for our perception of recurrence in the case of recordings. He acknowledges that we can shift our attention back and forth between original sounds and their reproductions. However, the exact *relation* between these two identities—whether abstract identity supervenes on individual identity—is left open-ended.

Martin's perspective also introduces the intriguing notion that sounds exist in our world in a *scattered* manner. When we play back a record, we encounter an abstract individual, whose other occurrences might be distantly spread across the past. In the case of our example, the sounds performed at the concert, of which we possess a recording, follow a peculiar path: *composed of two temporally disjointed parts, one occurring in the theatre and another in our living room*.

Even if it seems odd when applied to sounds, this idea holds relevance for events in general. Davidson proposed an account of recurrence based on *temporal discontinuity*:

> Events have parts that are events, and the part may be discontinuous temporally or spatially (think of chess tournament , an argument, a war). Thus the sum of all my droppings of saucers of mud is a particular event, one of whose parts (which was a dropping of a saucer of mud by me) occurred last night; another such part occurred tonight [14](p.28).

However, Davidson himself was dubious about this strategy. As he thought, this strange event-sum is not what we ordinarily refer to with statements about recurrence. Following Brandl, this idea of temporal disconnected event parts could be compared with the category of concrete universals[12]. Concrete universals refer to masses, like water or gold, which are non-individual—and often irregularly distributed—portions of matter. A mereological peculiarity is that when regular objects are divided into parts, masses are divided by quantity. Certain events, namely *homogeneous* events, might recall concrete universals. Imagine a process such as the increase in temperature; this event doesn't have to be continuous in time, yet each part of it is itself an increase in temperature. In this way, homogeneous events behave like masses, i.e, they represent measurable process and their

[12]See [6] p.101.

identity exist only at the level of sum. Nevertheless, even though this account of recurrence is correct, it applies to a very small number of cases. Sounds, while retaining a certain degree of homogeneity—-for instance, any silence is a part of a sound—, they enter our experience with sharp temporal boundaries. They are present in our experience as wholes. Envelope, transient's shape, internal structure, volume decreasing, are among those factors that determine when a sound has begun or ended.

Our intuition urges caution towards an ontology of scattered sounds. Although this concept is more plausible in describing the ontology of music—where a Western musical work is instantiated each time someone plays the score—sounds appear to exist differently, and perhaps recordings alone may not trigger a substantial shift in our intuitive understanding of audition. However, this mere fragment of intuitive or pre-analytic understanding, as I suppose, is insufficient to dismiss Martin's argument. The soundness of his proposition depends on two underlying assumptions:

1. In the act of listening to recordings, our awareness of the original sound emerges through a *natural perceptual process*.

2. Reproduced sounds do not *mediate* our perception in the same way visual pictures do.

From these two assumptions thogether, Martin derives that our intuition of re-encountering depends on having the original sounds, such as the voice of Florence Nightingale, once more in our ears. The lack of her physical presence becomes irrelevant for the object sounds to enter our auditory experience.

For our standard eventive ontology of sounds to remain consistent with recurrence, it is imperative to present additional reasons for resisting Martin's notion of abstracting the sonic world. The most logical approach, as the laws of reasoning suggest, is to challenge its premises. Specifically, either recordings release non-veridical perceptual content (negation of 1), or in audition, our perception might be mediated (negation of 2). As we will soon discover, these responses are largely put forth by distalists.

This far, we've presented the standard account on events, considering them as unrepeatable particulars, a stance shared by many philosophical account on sounds. Delving into Davidsonian theory, we've recognized linguistic evidence supporting the inclusion of events in our ontology. The subsequent issue of recurrence led particularists to consider statements involving event identity in terms of several similar event occurrences, using type-token strategies. However, it could be challenging to precisely determine the respects in which two occurrences resemble each other.

Conversely, for those who seriously entertain the possibility of recurrence for sounds, as evident in the experience of recordings playback, alternative ontological perspectives have been proposed. Some philosophers argue that sounds, as repeatable entities, are better understood as universals or, more plausibly, as abstract individuals. Having presented the challenge of repeatability for events, we now turn to explore specific responses from distalists aimed at defending event theory from the problem of recurrence.

3.3 In Defense of particularism

The experiences of recordings ground in listeners the idea of re-encountering the same sound more than once. From the standpoint of those who consider sounds as genuine happenings occurring in the same physical world that surrounds us, this poses a challenge. Especially in the case of recordings, the type-token explanation seems to be hindered by the extremely close resemblance of the two sonic events in audition: the original and the playback. Gathering these considerations together, endorsers of the eventive account have tried to delve into those factors that are relevant in such experiences.

Departing from the general ontology of events and natural language analysis, Dokic and O'Callaghan have sought to solve the problem posed by recording for sounds by looking at the modality in which we usually experience recordings. Since the idea of recurrence arises in audition, a theory of perception could disclose relevant distinctions between environmental and artificial experiences. While Dokic has analyzed accurately those considerations that would require an overhaul of our theory of particular events, O'Callaghan focused on the notion of natural perceptual awareness and whether recordings can represent reliable sources of information.

3.3.1 Two Ontologies of sound

Dokic is one of the proponents of the *Located Event Theory*, a distal account that considers sounds as *monadic events* occurring at their source. Being monadic events means that the vibration of an elastic body (within a frequency of 20 to 20.000 Hz) does not simply cause the sounds; instead, the vibration itself is the bearer of acoustic information. The existence of a medium for the propagation of the vibration, while necessary for auditory awareness of this event, does not pose a condition for the existence of sonic entities. Consequently, sounds can exist in the void.

In [18] Dokic proposes a comparison between two ontological stances: one that models sounds as *unrepeatable events* (UE) and one that considers sounds *repeatable objects* (RO). This second

perspective partially aligns with the the Martin's perspective presented above. The comparison is developed with respect to some particular sound-related scenarios, including recording experiences. The author aims to evaluate the impossibility for UE to be consistent in each of these cases.

For Dokic, the RO view is committed to the following claims:

1. sounds are particulars, not universal.
2. Sounds do not occupy physical space in the same manner as tangible objects; however, they can still be located in a specific region.
3. Sounds are autonomous entities regarding their sources.
4. Sounds are repeatable entities, capable of existing as a whole in more than one place and time.

While the first two claims are compatible with UE, the latter two deviate from a standard account of events. Regarding (3), in LET, sounds happen at their sources and are not autonomous. Instead, (4) challenges the condition of unrepeatability. Most importantly, what counts for two sounds to be identical in the sense of RO is their causal history:

> Two instances of the same auditory type are the same sound only if they both casually derive from the same sources. A qualitative identical sound made by another material object would count as a distinct sound [18](p.392).

This mirrors, in some sense, what Martin states about the process of instantiation for abstract individuals—that two playbacks of a recording share identity only if they are copies of the same particulars. In a similar spirit, Dokic considers RO as an instance of the ontology of *repeatable artwork*. The idea behind this theory traces back to the work of Van Inwagen in [67]. In this noteworthy article, the American philosopher, drawing from linguistic evidence, presents the idea, often referred to as *Meinongianism*, that fictional characters should be included in the catalogue of being. These entities, like Sarah Gamp in Charles Dickens' "Martin Chuzzlewit", exist in a manner different from material objects—akin to numbers, novels, and entities postulated by physics. Creatures of fiction are intangible, abstract things.

Moving on to musical entities, this "Platonist" account has played a relevant role in interpreting the Western tradition[13]. Symphonies, like novels and poems (as opposed to tangible copies of novels

[13]For an introduction to this theory see [54].

and poems), are usually considered to exist in the same *abstract manner* as fictional characters. Furthermore, they are *instantiated* through performances but exist continuously from their creation. In summary, construed as repeatable objects, sounds are similar to symphonies, except for the fact that they only exist when instantiated by environmental happenings.

In articulating the RO conception, Dokic delves into scenarios where this perspective better captures our ontological understanding of sounds in comparison to UE. Here, we will specifically focus on two of these scenarios: first, the case of reduced listening, and then the already familiar experience of recording.

The first scenario in which RO could be more accurate than UE is the musical encounter. Some argue that theories like LET, while accurately representing how audition works in *natural environments*, do not equally capture the aesthetic practice of listening. Perceptual awareness and recognition of sounding sources, which is presumably the natural task of audition, does not solely dictate the use of ears. The former, we can call it *ecological audition*, is about whatever causes the sounds instead of the sounds itself. Conversely, more "aristocratic" modalities of listening, like musical listening, would have as their objects the very sonic entities, completely abstracted from material sources. Regardless of this idea of multiple modalities in audition reflecting ontological differences, sounds as repeatable objects could be more accurate in describing aesthetic perception.

One of the most famous theorists proposing the idea that sounds exist separately from their physical sources was Schaeffer. Known as the father of *musique concréte*, an experimental practice producing new musical works from environmental noises, Schaeffer was also a prominent intellectual who extensively investigated the nature of sounds and recording tools. His works, along with the German avant-garde of *Elektronische Musik*, which focused on the method of producing sounds in electronic instruments, laid the foundation for what is now the broad field of electronic music. Both of these revolutions were made possible by a rush of technological development in the first half of the 20th century, with many finding their aesthetic applications after previous military or industrial usage. Together, these inventions piqued the curiosity of many scholars, and, as often happens, these pioneers were more theoretically aware of what kind of changes these instruments represent and how they conform to the current practice of Western music.

In providing the theoretical apparatus that grounds the composing practice of musique concrète, Schaeffer defines the notion of "objet sonore". As he states in his experimental diary[14], this is

[14]Schaeffer was also one of the greatest theorists of the usage of recording devices in musical practice. What we know today about his theoretical explorations derives from a diary from 1948—a period during which Schaeffer spent time conducting experiments in the studios of Radiodiffusion Française. This diary was published in 1952 under the

precisely the pure object of hearing, emerging from the multiple repetitions of bits of recorded sound—a sound completely detached from its original material source. The objet sonore is the very minimal unit of sound upon which Schaeffer grounds his project of a symphony of noises. This concept, as many have argued, derives from an interest in Husserlian *phenomenology*. In such perspectives, objects, intended as transcendental unities, are the product of mental synthesis—an act which Husserl refers to as "noesis". The material upon which consciousness operates is multiple experiences of perspectival acquaintance with external sources. Despite involving different modes of intentionality regarding the physical world, the objet sonore is also thought of as a mental construct. Just as material bodies are the result of the stream of perspectival views (the series of adumbrations), individual sounds derive from multiple acts of listening by a single subject[15]. It's easy to imagine how recordings could have influenced this vision.

Another notion that Schaeffer derives from Husserlian phenomenology is the intriguing concept of reduced listening or *acousmatic experience*[16]. Termed as such, it refers to the unique encounter by which the object sonore reveals itself to the listener. Drawing parallels with Husserl's notion of *epoché*, which involves suspending judgment on the external world to experience it anew, reduced listening similarly interrupts the natural standpoint. In doing so, it unveils the conditions that make subjective experiences possible. This practice involves hearing sound for its own sake, transcending its typical association with the physical world and the meaning it may convey (its network of cross-references).

The concept of reduced listening, originating from musique concréte, introduced the idea of layered audition, with lasting influence. Firstly, it provided the foundation for exploring rich *musical listening practices*, exemplified by concepts like radical listening associated with ambient music production[17]. Secondly, the notion of sound shaped by Schaeffer has inspired neuroscientific discussions on perception. In 2004, neuroscientists Griffiths and Warren proposed that the human brain represents auditory impressions in two-dimensional spaces—pitch and time—treating sounds as individual, spatially occupying objects [27]. Thirdly, philosophically, the idea of sounds detached from their material sources has been endorsed by Scruton [56]. In certain cases, Scruton suggests considering sounds as secondary objects or *perceptual ephemera*—borderline objects of perception,

name "À la recherche d'une musique concrète".

[15]For the Schaeffer's interest in phenomenology see [30].

[16]'Acusmatic' is a term deriving from the ancient Greek, often referred to the Patagonian practice of listening the voice of the master behind a veil. Schaeffer transposed this metaphor in audition, referring to the way of listening to sound abstracting from their sources.

[17]As reported in the Toop's book [65].

similar to rainbows and shadows in the case of sight. While Scruton acknowledges the importance of the temporal dimension for sounds, he argues that, at times, sounds enter perception as pure events without direct relations to objects as their participants.

Assuming this perspective, RO appears to more accurately capture the notion of sound portrayed as a cognitive object-like construct or detached from its material causes. However, Dokic argues that these phenomenological premises alone aren't sufficient reasons to shift from standard accounts like LET. While a focused practice of listening may reveal inner mechanisms and biases inherent in aesthetic perception—sometimes overlooked in ecological audition—it ultimately results in a mere change in the listener's attention:

> The sounds that one ears in reduced listening do not sound very different from the sounds that one hears in ecological audition. The shift from the latter to the former is a shift of attention rather than an ontological conversion. It is analogous to Gestalt switch in the visual domain where, as Wittgenstein pointed out, we see that the object itself remains unchanged [18](p.394).

The second scenario outlined to justify the RO perspective is the experience of audio reproduction. Dokic provides clear reasons from the outset for resisting the identification of the original event with its later reproduction. For UE, the two events are distinctly separate, not just numerically; they stand in a *causal relation*, and this alone excludes identity. However, to demonstrate the same degree of explanatory power as RO, UE should also furnish convincing reasons for why we are often inclined to recognize later acoustic encounters as the very same individual sound.

Differently from Martin, Dokic views recordings as *images* of the original sounds. Similar to how a photograph represents its subject, like a woman's face, in the sonic realm, the playback of a concert's recording portrays an audible event. While the original subject and its representations (photographs) are distinct—whether the first is made of meat and bones, the second is made of paper, or whether the second is a two-dimensional figure and the first is a three-dimensional body with causal powers—they are similar enough for us to recognize the woman in the picture as a certain existing woman. This analogy with visual pictures helps articulate the problem of sameness for sounds in terms of representation, offering a step toward solving the puzzles. However, it's crucial to acknowledge that this analogy might have some discrepancies. To establish credibility in this position, Dokic's argument proceeds by posing two questions: (i) What are the conditions for an image to be an image? (ii) Do these conditions also apply to recording playback?

Regarding the first question, Dokic recalls two necessary conditions for images:

1. the *non-transparency* condition—the subject and medium of representation must have perceptible differences.

2. The *twofoldness* condition—the experience of a picture must be twofold.

The non-transparency condition it simply requires the medium to be different from the subject. For instance, no real pipe is an *image* of a pipe. Conversely, the famous painting by René Magritte, "C'est ne pas une Pipe", as the title suggests, it is. Unlike any real pipe, the painted one is flat and presumably non-functional for smoking, and this difference is sufficient for it to be non-transparent. The twofoldness condition, though, requires being aware of both the particular arrangement of shapes, points, and shades in a picture and the subjects represented by means of such arrangements. This notion can also be compared with the dichotomous couple of *seeing-in/seeing* as coined by Wollheim in [69]: if one sees a pipe in Magritte's painting, one does not see the painting as a pipe itself.

Regarding the second question—whether these two conditions apply in the case of sounds—some problems arise. At first glance, recordings playback seems to be either transparent and not twofold. First, "when a sound attempts to depict another sound, the medium becomes transparent to the subject" [18](p.397). Second, the notion of seeing-in seems to lack any auditory correlate in the form of *hearing-in*. Omitting the assumption that sounds can carry non-acoustic information, such as meanings or emotions, anything one hears in a sound, he also hears it as a sound.

However, for Dokic, this is nothing but an apparent hindrance. If one enlarges the analogy with other non-depictive forms of representation, like sculpture or trompe-l'oeil, it becomes clear that not every instance of representation strongly *obeys* the twofoldness condition. Wax sculptures, like those collected in the Madame Tussaud museum in London, represent their subjects as if they were really there. The huge astonishment provoked by these sculptures resides in their presenting their subjects in a very realistic manner, letting us see, as in the case of sounds, the statue as the very subject it represents. Of course, perceptual differences between the representation and the original exist, and after an initial dismay, we become aware of them. Nevertheless, those differences, grounding a failure in transparency, characterize our experiences as twofold only in a loose sense.

Similarly, recordings are accompanied by properties that are not shared with the original sound event. For instance, vinyl crackles, the bit-crushed timbre of mp3 compression, coming from a loudspeaker—these are all properties belonging only to the playback. Consequently, audio representations predominantly function as non-depictive images. Moreover, as Dokic suggests, in

specific instances, audio pictures might qualify as *images* in a narrower sense. In scenarios like phone calls, the concept of hearing-in holds ground: "you hear a loud sound in another, soft sound, but you certainly do not hear the latter as a loud sound" [18](p.399). The potential to represent audible properties in signals lacking these features is linked to *shortcuts* in our recognition abilities. Perhaps, a voice's timbre carries enough information to distinguish between shouting (loudly) and whispering (softly). Leveraging these shortcuts allows for the creation of depictive audio representations that respect both non-transparency and twofoldness conditions.

The possibility for two sounds to be one the representation of the other provides an intuitive solution to the problem of recurrence. This strategy, in turn, justifies the UE position by conceding that subsequent encounters of a sound event do not have the same epistemological status as the original encounter, and our impression of hearing the same thing more than once is grounded in the *efficiency of the representational process* (such as recording and playback). Neither the practice of reduced listening nor the experience of recordings force us to set aside our temporal understanding of sonic events. For Dokic, once these presumed weaknesses of UE are resolved, the construction of sounds as repeatable objects would lose its point.

3.3.2 Hearing the past

For O'Callaghan, sounds are events occurring when a vibrating source meets an elastic medium, a perspective introduced in 2.4.2 as *Relational Event Theory* (RET). His 2007 work [50] focuses on characterizing the perceptual parameters of audition that justify our ontological stances. Furthermore he dedicates many pages to explaining acoustic phenomena that clash with the eventive perspective, including recordings [50](ch. 10).

According to RET, the problem of recurrence, as manifest in recordings, is similar to that of echoes, supporting a comparable explanation in terms of partially *deceptive experiences*. In his introduction to the topic, O'Callaghan points to three aspects in which the puzzle present itself. The first involves cross-modal connections between the playback of audio signals and visual representation. He suggests that the problem of perceptual reproduction doesn't only apply to sounds, partially aligning with Dokic's point: "live video, film, and even photographs enlist technologies that ground information-preserving causal connections between ordinary objects or events and visual experiences" ([50](p.143).

The second concern is how recordings interface with perception. Just as film requires a precise set of tools to capture and reproduce scenes, in the case of recording, the perception

of the original sounds flows through artificial equipment. This intermediacy, as someone might argue, interrupting, altering, or even extending the *natural causal chain* of perceptual awareness, compromise the possibility of being aware of the original events. In this sense, recordings behave like a sort of prosthetic extension of our perceptual apparatus, raising questions about their reliability as sources of information.

The third aspect concerns differences among the phenomena batched under the class 'recording'. For O'Callaghan, it is important to differentiate between *real-time*, *non-real-time*, and *musical* recording experiences. For instance, tuning the radio to the football channel to attend a match involves listening to a sonic event—the commentator's speech—happening in (quite) the same moment but in a distant location. The same could be applied to video recording, such as watching the same match transmitted on our television. Conversely, tapes containing speeches and video recordings from the olden days, when played back, inform us about something that happened in a certain region of the past time. Lastly, musical performances' playback, which can occur either real-time or non-real-time, is characterized by an additional aesthetic value.

This last point requires clarification. One might not completely agree to consider real-time broadcast as genuine cases of recordings because of their immediateness. These transmissions lack the essential quality of being recordings—their enduring nature, capable of producing multiple experiences of the same sort. This is partly the point raised by Uidhir in [66] in supposing the existence of self-destructive tapes such that they cannot be reproduced more than once. However, as Kania argued, this argument rests on a confusion between actual repeatability and *potential* or *possible repeatability*, where only the latter is a relevant property ([32](p.25). For considering real-time playback experiences as genuine cases of recordings, we have to look at the way those tokens are produced. In my view, whenever an audio signal is captured and "translated" into a *preservable format*, like a sequence of digital bits, in the grooves of a vinyl or in the alignment of magnetic dipoles within a tape, we have a case of recording. The fact that such translated sound is then reproduced once, multiple times, or never has nothing to do with the transcription itself. In an opposite manner, direct amplifications of audio sources that don't require any transcription in such a format, as could be the case of an electric guitar, being captured by means of microphones and loudspeakers, don't count as recordings. In essence, if, in real-time broadcasting, sounds pass through the right process (as is the case with Digital Audio Broadcast nowadays), then we shall consider it as an instance of recordings.

Starting precisely from such real-time experiences, O'Callaghan wonders what makes being at

the football game different from seeing it on our television. As someone might argue, attending the game through an artificial device could count as an interruption of the causal process involved in ordinary perception. For many philosophers, including Strawson, the preservation of a natural causal chain is a necessary condition for a perception to happen under normal circumstances [62]. Lacking this condition results in the failure of perception. We will refer to this stance as *natural perceptual awareness* (NPA). Assuming the existence of the NPA process, many mediated phenomena cannot be considered as happening in the proper way, and in certain cases, this condition seems to be too strong:

> virtual reality glasses and surgically implanted vision aids of the future, as well as present-day hearing aids and cochlear implants, qualify as making possible genuine perceptual experience. None, however, preserves the normal causal process [50](p.144).

For O'Callaghan what is relevant in considering an experience veridical is not its casual process, rather the *preserving* of the perceptual content. Radio and television broadcasting, by preserving relevant information about originating events make us perceptually aware of speeches and sports matches, yet, pace Strawson, they do it in a mediated and impoverished way. But in what does this mediateness consist, and how does mediated awareness differ from natural awareness?

The answer reside in the *perspectival content* of experiences. In vision, physical objects are presented as having a certain shape, colour or texture, and these properties enter as contents of our perception. In addition, they unleash some other properties, namely relational properties, which depend on our perspective on the environment. For instance, when looking at the Rubik Cube on our shelf, we become aware of the chromatic combination of some faces, but we also know that it occupies a certain region of our visual field, that it is a certain distance from us, and that, from this standpoint, it presents with a peculiar reflection of light rays. All these pieces of information display the intersection between objects of awareness and the egological dimension.

Such a perspectival content, according to O'Callaghan, is a crucial aspect of perception. However, mediated visual experiences, such as television transmissions, fall short in preserving this feature. They are *deceptive* in maintaining the egocentric location of objects they depict in relation to the viewer:

> When, for example, a subject views a televised image of a speech, the perspectival characteristic of the experience do not depend in the right counterfactual supporting

> way upon the spatial relationship of the viewer to the speaker. The president could move from side to side while occupying the same relative position on the television screen. You can move from the living room and view nearly the same image of the speech. The televised image does not change in the subtle ways that perceptual content does as you move your head and eyes to explore the scene [50](p.146)

As with vision, auditory perception also presents perspectival content. The position one occupies in relation to a sounding source is relevant for defining the proper ecological location, the location from which the sounds originate. Furthermore, non-perspectival properties like pitch or amplitude of the signal might change depending on whether the listener is stationary or moving relative to the source (as in the case of the Doppler effect). Following the same line of reasoning, real-time audio transmission should be considered as failing to provide a natural awareness of the original sound.

However, for O'Callaghan, stating that hearing a relayed speech in telephonic conversation doesn't count as a veridical experience is an overly radical consequence of philosophical theorizing. The crucial point here is that sounds contain different information depending on the qualitative type they instantiate—he distinguishes between spoken words, environmental noises, or musical sounds—which are preserved in the playback. Through this process, we produce a new sound of the same type that has roughly the same informational value as the original. For this reason, experiencing the reproduction of a sonic event counts as being aware of the original signal.

Regarding perspectival content, as we mentioned earlier, this feature is crucial in distinguishing between *perceiving* and *failing to perceive*. An experience lacking such content cannot be considered an experience at all. Despite being mediated, the perception of a sound through its playback does not lack perspectival content. Instead, such content is determined by the mediating devices, which, in the case of audio representation, include the acquisition systems (e.g., microphones). The same principle could be applied to visual analogues:

> Despite its impoverished veridical content, an experience of a transmitted sound or image constitutes a form of perceptual contact with the original source in virtue of our adopting a point of view that we do not in fact occupy. Since the experience invokes the perspective of the camera or microphone, that content suffices, along with the veridical qualitative and non-perspectival content to grasp the original source as an autonomous of perceptual acquaintance [50](p.155).

The notion of 'impoverished veridical content' directly connects to our previous discussion on illusions and hallucinations in the context of proximal theories (2.2). Real-time audio representations are impoverished in terms of their perspectival content, whereas any genuine experience, like *Natural Perceptual Awareness* (NPA), ensures appropriate perspectival access to perceptual objects. However, mediated experiences still provide perspectival acquaintance, albeit with an illusive content. The illusive perspectival content involves perceiving a the original source as being located in a position it does not actually occupy, for example, the radio's loudspeaker. It's important to note that an experience that is illusive to some extent does not qualify as hallucinatory, meaning it is not non-perceptual. In conclusion, real-time recordings' experiences are *bona fide* mediated experiences that ensure acquaintance with original sounds but are illusive regarding the location of these sounds in relation to the listener.

Moving in the realm of *non-real-time* replays of recordings, the same kind of argument applies. Consider a tape containing, for instance, a speech from the 80s recorded with the iconic Sony Walkman. When played back, it produces an auditory impression that, according to O'Callaghan, ensures genuine awareness of the original sound event. This holds true for films and photographs as well. While this position might seem paradoxical, one might rightly object that the existence of the perceived objects or events at the time of perception is a minimum requirement for awareness.

In defending its position, O'Callaghan proposes considering a *temporal outlook* among the perspectival contents included in our experience. The act of listening to someone's discourse is *apperceived* as something taking place in the present. Even if one could attribute duration to sounds articulating words, such information intuitively is nothing but inferred from multiple instants of listening, each one occurring in the immediateness of the present. As he says, any appeal to presentism regarding perceptual awareness, confusing a failure in perception as evidence for non-existence, collapses into a mistaken account of perception. Things of the past could still enter our senses through appropriate devices of reproduction; nonetheless, in doing this, they don't preserve their original temporal perspective:

> Hearing a recorded sound thus is a way of listening into the past. You hear the sounds and events that occurred a the time of the recording, but experience present those sounds and events as taking place during the time of listening [50](p.159)

In the case of recordings of past events, similar to real-time experiences, what we perceive is an *impoverished veridical content*. However, this content is illusive, beyond its spatial information,

in presenting the *original temporal perspective*. Our impression of attending a speech from 1980 taking place right now in the comfort of our living room is an illusion supported by our adopting the temporal perspective of the microphone used in capturing it. Nonetheless, perceptual acquaintance with the original events is preserved. Drawing a parallel with vision, the light emitted by stars in the night sky, which we often observe, carries information about a source that has already ceased to exist. The luminous signal traveling through space brings nothing but information about the past.

For what concerns the last kind of recordings, namely *musical performances*, O'Callaghan invites us in making a distinction between our listening to a song and our attending to a past performance. Starting from the former, imagining listening to the live version of 'Harvest Moon' in Neil Young's album "Unplugged" from 1993. As a long-standing listener might know, the name 'unplugged', featured in a variety of live albums, refers to the projects begun in 1989 from MTV (Music television) to bring rock artists on the stage in an acoustical vest. In this case, the album collects some of the songs performed during the shows.

What guarantees the possibility for 'Harvest Moon' to be multiply reproduced over time resides in the capacity of our stereo to produce each time the same song, that is particular sounds of the same type as the one performed in 1993 on the stage. In general, as O'Callaghan says: "a pattern of sounds counts as an instance of a song just in case it includes sounds of the appropriate pitch and duration, arranged according to the right time" [50](p.161). Thus, any kind of perspectival impoverishment occurs in our listening to a token of 'Harvest Moon' with our stereo.

Things go differently concerning our *perceiving a performance* through hearing a song. Listening to 'Harvest Moon' is certainly a way to relive the experience of being among the seats at Universal Studio (LA) that night. However, our being aware of the performance is illusive for its temporal and spatial content. Furthermore, occurring in a single modality (audition), the performance lacks its original aesthetic value—a blending of cross-modal experiences essential in live show or recording studio experiences. Therefore we hear the performance only in a loose sense.

Upon a comprehensive examination of O'Callaghan's argument, a potential source of confusion emerges regarding its consistency with the particularism claim of uniqueness for individuals. O'Callaghan asserts that listening to a recording allows us to be aware of and, to some extent, relive the very same sounds from the past. This raises a question: doesn't particularism rule out encountering the same event twice? While O'Callaghan's account doesn't explicitly endorse multilocation for events, it suggests that perception *extends* over the actual existence of the original sounds.

The mechanism of reproduction plays a crucial role in extending the front of our experiences which is not strictly bound to the present time but extend at least to the past. Yet, the awareness of past events always involves specific mechanisms, some "special tricks", such as reflection in the case of echoes, aiding sounds to rebound or return. Despite this, being aware of something that happened in the past does not, in reality, add new disconnected parts to its original spatiotemporal path. While the sound produced by our stereo is a new event, it still makes us perceptually aware of the original one. According to O'Callaghan, this view remains consistent with particularism.

Until this point, we've outlined the primary responses offered by proponents of the distal conception to address the issue of recurrence as it surfaces in recording experiences, integrating it into an eventive perspective. Both of these views contrast with Martin's argument by challenging its underlying assumptions. Despite their diverse viewpoints, these responses share a common goal: to furnish not only adequate justification for distinguishing the original from subsequent playbacks of sounds but also compelling explanations for the precise foundation of our intuition regarding sameness. Let's attempt a concise retracing of these elements.

According to Dokic, there are compelling reasons to uphold the eventive perspective. The diversification in listening attitudes and recording experiences doesn't offer strong evidence to abandon our naive ontological account of sound. Specifically, the sense of re-encountering the same individual events while listening to a recording is evoked by the fact that later reproductions serve as *representations* of the original. Audio recording systems, akin to cameras, align with other devices in the process of representing and fabricating *images* of their subjects. While some audio images, like telephonic transmission, are purely depictive, adhering to conditions of twofoldness and transparency, others, such as high-fidelity music playback, are non-depictive. These non-depictive forms, akin to realist visual representations like sculpture, create the illusion of presenting the subject once more.

Consequently, the reproduction and the "prototype", while sharing some perceptual properties, don't hold the same epistemological status. Moreover, they exist in a causal relation, sufficient for excluding identity. In essence, just as we can distinguish someone from a picture of him, we can equally discern the original performance of a concert from its later reproduction.

O'Callaghan partially aligns with the analogy between forms of perceptual reproduction. Furthermore, he asserts that playbacks or recordings offer *genuine acquaintance with the original events*, establishing a foundation for attributing strong significance to the idea of sameness in terms of perceptions. The justification for our impression of hearing the same alarm's beeping again lies

in the actual perception of those beeping once more. Notably, this perspective effectively preserves ontological soundness while acknowledging the original sense associated with recording experiences. It demonstrates that the concept of recurrence, when interpreted through a non-presentist account extending perceptual awareness to past time, poses no threat to particularism.

To uphold the notion of perceiving past events, O'Callaghan characterizes recording experiences as *impoverished* and *illusory* in terms of spatio-temporal perspective and aesthetic value. However, reluctance may arise when equating these experiences with pure auditory illusions like the Shepard scale or Doppler effect. Soteriou, advocating for a pluralist ontological account of sounds, rejects describing radio broadcasts and music playback as deceptive experiences [58]. Citing critics of Austin's move to a standard account of perception and considering the veridical/illusory dichotomy as "bogus", Soteriou contends that what philosophers might label as deceptive, the plain man, with a less minimal idea of perceptual datum, may not[18]. Therefore, an account of sounds seeking to embrace the widespread idea of recurrence should revise either the conditions imposed by particularism or the deceptive nature attributed to recordings and echoes.

In my view, considering the listening of a musical record as somewhat illusory is not an alarming departure. Playback experiences are deceptive not only concerning temporal and spatial factors. Examining the process reveals a significant issue: a considerable amount of acoustic information unleashed by a sounding source gets lost due to the limitations of our devices. While microphones are undoubtedly splendid and useful tools, akin to the original blueprint—the human ear—they only capture a *partial representation* of the entire vibrating process of a body. Using a hundred different microphones would yield a hundred different samples of the same event. Which of these samples is the truly loyal and veridical representation in a strong sense? Probably none.

Setting aside illusions, I believe that both arguments proposed by Dokic and O'Callaghan offer positive responses to the question of whether a satisfactory explanation of the sense of recurrence in recording experiences can be provided within the framework of particularism for events. However, an exploration of the ever-evolving landscape of recorded music, shaped by technological advancements and transformative practices, continues to enrich our understanding of sound. Notably, as technologies like computers were introduced in the 90s, these ongoing modifications contribute to shaping the dynamic concept of recordings. Therefore, a closer examination of current practices in creating recorded sounds becomes crucial for a comprehensive exploration of the philosophical thought surrounding recurrence.

[18]For the critics of Austin see [1].

In the next chapter, I will introduce an expanded perspective on the existing framework, with a specific focus on the concept of *audio image*. Initially, we will define the role of recordings within the broader spectrum of representational systems, highlighting their structural similarities to photography. This will be followed by a comparison with other types of non-mimetic audio images. Subsequently, we will explore the nature of recordings, particularly what and which kinds of entities they can represent. Finally, the chapter will conclude with an examination of sound recognition in the context of listening to recordings.

Chapter 4

Sonic Pictures

4.1 Recordings and images

As we navigate the realm of sound, the phenomenon of recurrence appears to have assumed a less dangerous aspect. Recordings, while still warranting scrutiny regarding their nature, have grown more familiar since the beginning of our discourse. The perspective put forth by Martin and others, which posits sounds as abstract entities defined by their recurrence across various moments in time, maintains a certain fascination. Undeniably, the sensation of something returning during playback experiences persists as a distinctive characteristic of tape and CD reproductions. However, we now possess tools to redirect the essence of this experience from an *ontological* focus to a *perceptual* one, sparing us from unsettling our intuitive understanding of events.

Up until this point, we have presented how the experience of recording shapes our ontological understanding of sounds. Sound, being inherently ephemeral compared to physical matter, does not impose a specific metaphysical perspective. Rather, the sonic realm accommodates various descriptions, each selectively highlighting a relevant aspect of auditory perception. As argued extensively in this work, here we focus on phenomenological and perceptual considerations that prompt us to consider sounds as *unique events* occurring either at the source or into the medium of propagation.

While these philosophical inquiries have dedicated substantial efforts to describing environmental acoustic phenomena as events, recordings open up the intriguing possibility for sounds to transcend their inherent spatio-temporal boundaries. As previously articulated in 3.1, when we use the term ‘recording’, we are alluding to a sonic event: the playback of a sample. From a previously captured sound, we can indeed reproduce a signal with the same informational content as the one that has already faded from the forefront of the present time. Does this imply that the reproduced tokens and the original are the same particular sound? While Martin unequivocally affirms this,

Dokic and O'Callaghan suggest an alternative perspective, proposing that we can describe such recurrence in terms of resemblance, representation, and perceptual access.

What connects both Dokic and O'Callaghan's responses is the analogy they draw between *recordings* and *visual reproductions* such as photography and video. Although the concept of imagery traditionally resides within the visual domain, it aptly captures the nature of auditory individuation taking place when, for instance, we reproduce vocal messages from our cellular phone. However, this analogy comes with its nuances. As Dokic argues, recordings, in most cases, can be considered images only in a loose sense. They fall short of immediate opacity or non-transparency regarding their subjects. In a certain sense, recordings, among all forms of imagery, stand out as one of the most convincing and illusory representations.

Despite the plausibility of considering recordings as genuine images, we still lack a compelling philosophical theory of auditory representation. In the realm of visual things have unfolded differently and images quickly found their place in philosophical discourse. Whether paintings, drawings, or pictures, images represent objects by capturing their visual appearance. Simultaneously, these images induce a certain *vacillation* in the minds of perceivers who are tasked with describing them. On canvas, artists achieve perspective illusions on flat surfaces, similarly, pictures convey their subjects as three-dimensional on two-dimensional mediums[1]. Essentially, visual images prompt *conflicting descriptions*.

The pictorial feature often referred to as 'duality' has inspired philosophical interest, initiating a structured debate that took root in the sixties with the emergence of two seminal works, first by Gombrich in 1960 and then by Goodman in 1968. This research laid the groundwork for addressing questions such as: How do images represent? What psychological processes are involved in subject recognition? In what relation do subjects and representations stand? Do they resemble each other in some respect? Today, various accounts on images exist, including *resemblance* theories, which assert that pictorial representation, unlike linguistic representation, perceptually resembles the subject, psychological accounts focused on *recognition*, and more *formal* approaches, such as the one defended by Goodman.

In drawing an analogy between recordings and images, we take advantage of a wide range of studies and theories. If sound playbacks are rightly subject to the same general principles governing images, they should be classified among representational processes alongside photographs, languages, pictorial techniques, and diagrams, each with its distinct set of *rules* and features for

[1]For a detailed introduction to the topic see [36] Ch.1.

establishing whether a token adequately stands for something 'O'. In the linguistic domain, we find formal and semantic rules; in pictorial representation, perspective rules and the artist's intention play a role. While there may be disagreements on which set of rules is relevant for a particular system, we can converge on the shared intuition that the essence of representation—the act of standing in place of something else—is what *connects* every unique form of image production. Applying the same logic, recordings, as a distinct form of images, *should be explored using a similar investigative pattern to determine their rightful place among other forms of representation*.

While the preceding discussion has focused on clarifying scholars' opinions surrounding recordings, the forthcoming exploration delves into less-charted territory, transitioning from cabotage to offshore exploration. Considering that recordings, akin to pictures and other forms of images, provide insight into past events as a viable response to the issue of recurrence, the objective of this concluding chapter is to propose a suitable framework for sound images. By the end, we aim to address the following questions: *What* is a sound image? What *types* of sound images exist? What do recordings *can* represent as images, and lastly, which audible qualities are relevant in the process of *acoustic recognition*? The essence of this final part can be perceived as interrogative rather than explanatory, serving as my modest contribution to the topic.

Firstly, let's recapitulate our earlier discussion about sound images. Guided by Dokic's argument, we explored the analogies between sound playback and non-depictive forms of representation, such as sculpture. From an intuitive standpoint, these analogies seem more fitting. Sound images seemingly lack the capacity to generate *twofold experiences* in the listener, a characteristic often found in various forms of visual representation. On the other hand, while sculpture and realistic reproduction may give the impression of embodying these properties, they fall short of exhibiting the same level of *fidelity* as recording playbacks.

In this context, recordings risk being considered as either an exceptional and unique form of representation, relegated as a peripheral theme, or being considered images only in a loose sense, as we concluded earlier. Yet, fortunately, recordings can always find solace in their visual counterparts: photographs. Sound recording techniques and photographic techniques share a historical kinship, both having evolved in the industrial stream of the 19th century and continually improved into the digital era. Both brought about significant transformations in their respective domains, achieving widespread diffusion. Moreover, both photographs and recordings necessitate specialized mechanical equipment for their creation, a factor that replaces (or diminishes) the role of the *subject's mediateness*.

4.1.1 The family of recording processes

This last point is crucial for establishing a more structural connection between sound images and photographs. Kulvicki, in [37], echoing the spirit of Haugeland's concept of an *icon*, asserts that the property of being the product of a recording-based process is a common aspect shared by various forms of representations. In his work, the term 'recording' signifies a *standard method for producing images* of objects and events, applicable not exclusively to the sonic domain but also foundational for pictures, graphs, and diagrams.

As Kulvicki argues, while recordings possess a relational character, representations, in general, are *intentional*. Pictorial representations crafted before the advent of photography required mediation by the subject. In contrast, discrete cameras and microphones, for example, lack any *conscious state* with the subjects they represent. No form of *aboutness* is at play in the chemical formation of a photographic negative or a wax cylinder during the sound-capturing process—only a mechanical or causal process.

The relation between recording and the recorded is characterized as witless, being 'oblivious to content (if any) and ignorant of the world' [28](p.180). This does not imply that the technology underlying recording processes is itself witless, but rather that once these machines are realized and functioning correctly, the role of active agents takes a back seat.

Secondly, all recordings enable *playback*—an inverse witless process where stored files are reproduced, creating a replica of the original source. When the engine is turned, the wax cylinder begins to generate its own sound. In the digital domain, although the process may seem less cognitively accessible, the same witless process is at play: JPG files are converted into images, just as WAV files are transformed into pattern of positive and negative voltages that induce the movement of a magnet. Both forms of image production involve an *intermediate* phase in which the file is stored in a somewhat encrypted manner[2]. Conversely, for Kulvicki, when recording and playback coincide in a *unitary operation*, we are dealing with something akin to a *copying* process. This is evident in daguerreotype production—an obsolete photographic process invented in 1839, where a picture made on a silver surface sensitized with iodine was developed by exposure to mercury vapor.

Thirdly, in Kulvicki's sense, recordings have inherent limitations as they cannot *capture ev-*

[2]As discussed in the section on O'Challagna recordings (see 3.3.2), the format of recordings plays a crucial role in how sound is captured, stored, and reproduced. This encompasses not only the physical medium—such as vinyl, tape, or digital files—but also the technical specifications like bit rate, sampling frequency, and compression methods. These elements determine the fidelity and character of the recorded sound.

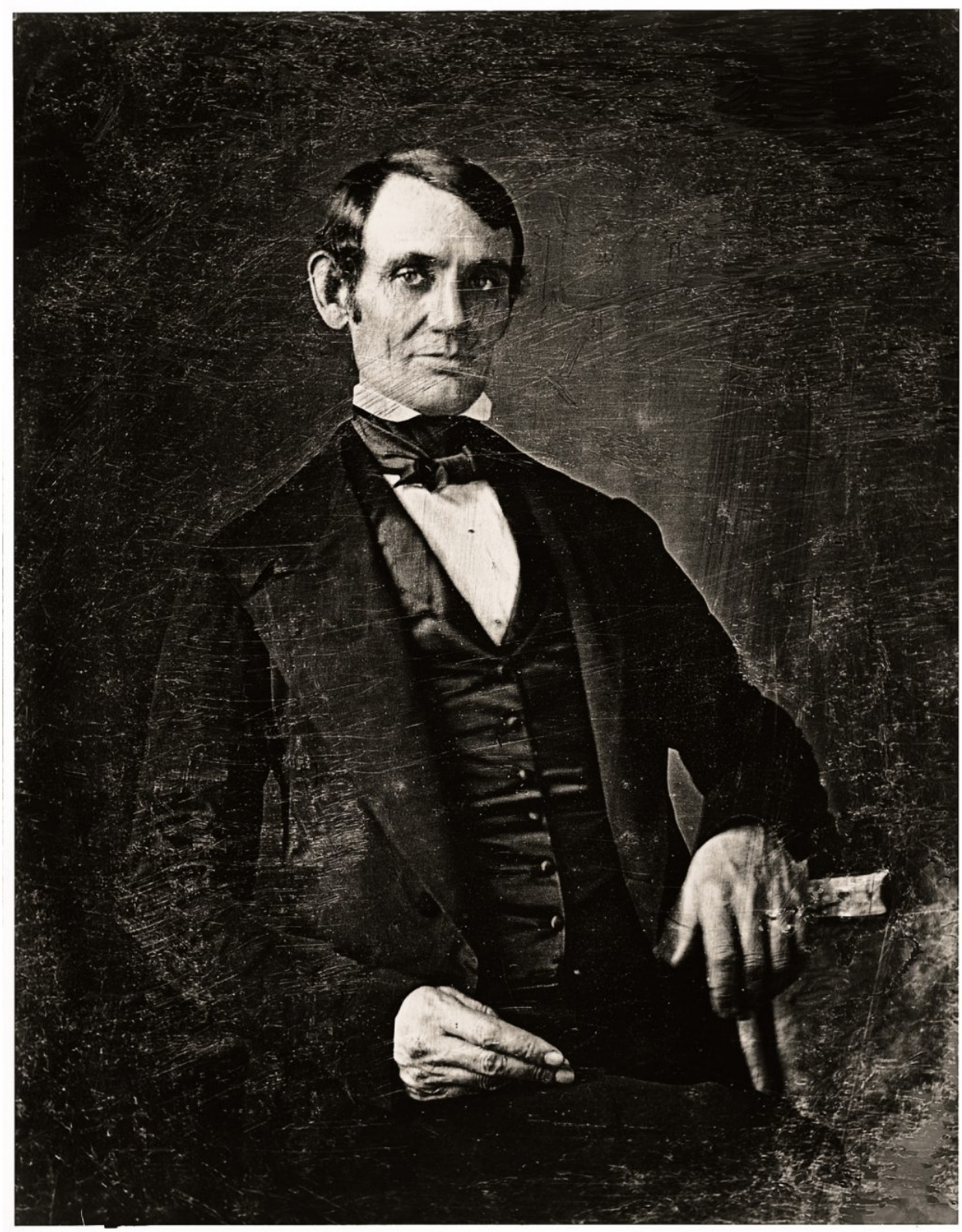

Figure 4.1: daguerreotype of Abram Lincoln from 1847.Source: https://www.whitehousehistory.org/.

erything, especially aspects that cannot be reproduced in a witless manner. Real individuals with flesh and bones like Charles de Gaulle cannot be faithfully reproduced, although certain attributes associated with a person can be recorded—such as voices, shapes, and internal images reflecting their mental activities. According to Kulvicki, the recording process is *restricted* to qualities and features that can causally enter into a witless process.

As discussed earlier (2.4.1), Kulvicki characterizes sounds as dispositions of objects to resonate in specific vibrational patterns, emphasizing their similarity to *universal properties* rather than *individual events*. Nonetheless, even adopting a particularist perspective on sounds, this restriction on the spectrum of recordings remains relevant. I think that both Charles de Gaulle and one of the speeches he delivered during the French Resistance from his London-based broadcaster were individual instances of a certain kind, somewhat related: the man who served as the chairman of the French National Committee during the Second World War and an individual sound event occurring just once in history. However, while only partial representations of the former are preserved today, the relevant auditory qualities of his speeches have been accurately stored[3]. *Every* audible aspect of a sound event can *causally enter a witless recording process*, leaving no additional unrecordable auditory content in its wake (except for contested properties like the original spatio-temporal location). This is why sound reproduction is so effective in creating tokens of the same type, unlike wax statues of French presidents, which merely mimic their subjects in a deceptive manner.

In simpler terms, what recordings capture of their subjects is the *skeletal content*—the maximum information furnished from a representation whose availability is not augmented or mediated in any way. This concept stands in opposition to the fleshed-out content, the overflow of inferential content that we might sometimes mistakenly attribute to the representation itself. For Kulvicki, the formula to uncover any bare-bones content is as follows:

> First, peel of the layers of implication, in its broadest sense, by imagining alternative, acceptable fleshed-out content. Then with these alternatives in mind, ask what they all have in common. The portrait depicts Lincoln, but it's possible that an exactly similar daguerreotype could have been the result of focusing on a wax model of Lincoln, an alien impostor, or even another daguerreotype of Lincoln. These disparate objects have in common only the way that they project onto a surface to form the picture. That common ground is the content's skeleton, that on which all the fleshings-out must build [37](p.275).

[3]Many of Charles de Gaulle's public speeches are compiled in the collection titled 'Charles De Gaulle – Discours Historiques 1940-1969', released in 1970 by the Institut Charles De Gaulle.

We can employ this method ourselves using the daguerreotype of Lincoln in fig. 4.1. Through this approach, we discover that the skeletal content is nothing more than a specific pattern of black and white stains on a flat surface. Employing a similar approach proves effective even in the case of de Gaulle's voice. Indeed, the same sound pattern could be recorded from a German spy with an excellent French accent or by other recordings; nevertheless, what we recognize during the reproduction is unmistakably de Gaulle's voice.

Compiling this identikit for reproductive processes—known as *new representational genera*—Kulvicki takes a step further by differentiating representational systems modeled on these processes that are not *bona fide* recordings from those that are both modeled on these processes and are fully-fledged recordings.

Within these former systems, we encounter the families of *graphs* and *diagrams*. These forms of representation are modeled on the recording process, yet they don't *instantiate* any of the features they represent. Consider, for instance, a graph depicting the changes in temperature over time. It is organized in two dimensions—discrete temperature values and time instants. However, at any given moment, the *graph's temperature* is irrelevant. Graphs and diagrams represent their subjects without sharing qualities with them. Instead, they resemble their subjects in terms of structural similarity[4].

On the opposite bank, we encounter a genus of representation that shares qualities with its subjects[5]. Much like graphs, these images are produced through recording processes. However, they generate their tokens by instantiating (either partially or entirely) the skeletal content of their subjects. Typically, tapes, when played back, produce sounding events whose acoustic qualities mirror those of their subjects. A recording of a clarinet solo doesn't represent it like a graph would; instead, it produces sounding tokens—a square wave unfolding with a certain frequency and volume pattern—in which *every auditory aspect* is relevant for determining whether it is a faithful representation of the original events.

This category is what Kulvicki elsewhere labels as *Pictorial* or *Mimetic* representation [35]. Furthermore, only forms belonging to this class, such as photographs, videos, or sounding tapes, can rightfully be called recordings. Moreover, photographic and sonic representation resemble each other from a formal perspective. This resemblance extends to the way these representative systems are organized in terms of syntactic structure, semantic richness, and transparency possibilities—all

[4]For an account diagrammatic representation see [64].

[5]A third class is the one in the middle of representation displaying *some* of the qualities of their subjects like fMRI scan of brains activities.

aspects that will be explained further ahead.

Returning to the initial questions, following this compelling argument, recordings occupy a specific rather than a vague position in relation to the families of representational systems. A sonic image is *a representation based on recordings, utilizing a witless process capable of producing multiple tokens, with the aim of capturing and reproducing the skeletal content of the original sound event*. Furthermore, the content is represented in a depictive or mimetic manner.

Aligning audio recordings with other mechanical processes like photography from their theoretical foundation brings an additional advantage: if sound playbacks are considered images only in a loose sense, then pictures *must be as well*. This conclusion may seem counterintuitive and one that, I believe, should be approached cautiously.

4.1.2 Graphic representations of sounds

Expanding on Kulvicki's framework, we may question whether recording playback is the sole possible form of representation for sounds. Indeed, while the condition of being *mimetic* is necessary for auditory images, not every recording-based system satisfies such a condition. What I propose is that, just as recording processes, as defined earlier, encompass a broad family of systems, the set of sound images can be further specified.

To begin with, the notion of representing sounds, or representing with sound, has historical roots. Classical Western music provides numerous examples in this sense. For instance, Bach's compositions were known for skillfully representing musically idiomatic religious scenes: "when Bach sets a text that mentions Christ descending from heaven, the theme descends, and when it mentions his ascending to heaven, the theme ascends" [34](p.190). In some sense Bach's music represents the passage. However, as Kivy suggests, the analogies connecting religious scenes within the musical phrase are nothing but abstract, unattainable without an external text or explanation.

More recently, the concept of Musical Images has been advocated by Leddington in [40]. According to his argument, certain *musical performances* can depict sounds in a more pertinent manner, without any further aid. This is exemplified in beatbox practice, a vocal technique for imitating the sounds of a drum, and cover songs. However, the process of recognizing, or hearing-in, the intended sound from the performance requires a significant quantity of fleshed-out content. Even though these practices may hold relevance in a musical sense, it remains dubious whether they truly capture the *skeletal core* of the original events, differentiating them from sound images in a strict sense.

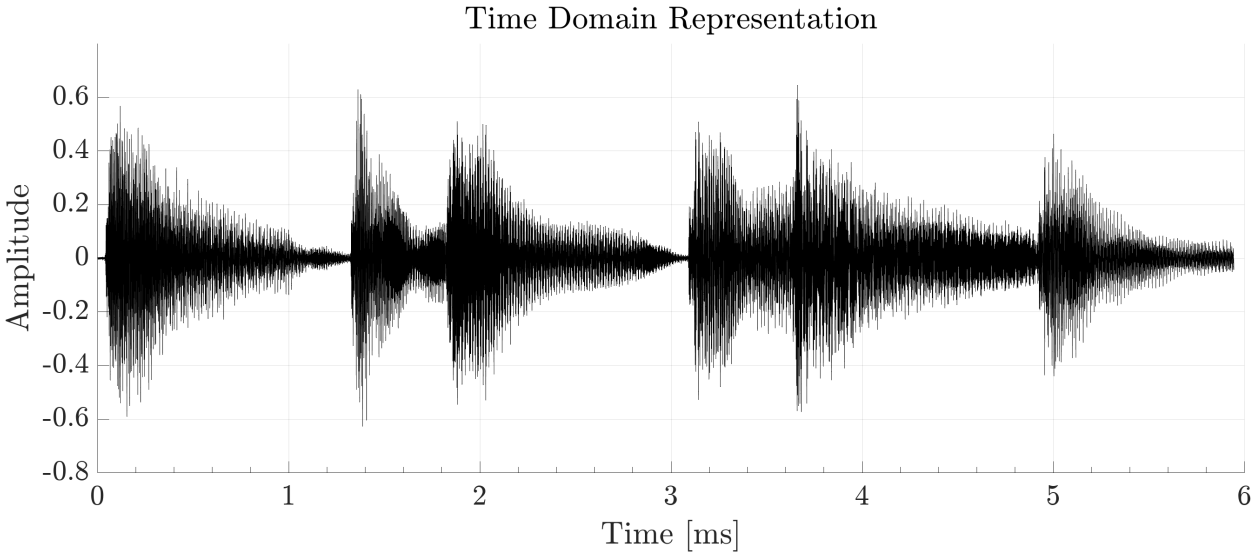

Figure 4.2: the waveform of the first six notes of Chopin's Nocturne No. 20 in C sharp minor.

Beyond musical images, another noteworthy form of sound representation within the present frameworks is the *graphic representation of acoustic signals*. Such representations are commonly encountered in sound design software interfaces, particularly so called digital audio workstations (DAW). These visualizations typically manifest as graphics where auditory qualities such as frequencies, amplitude, and length are depicted visually. The kind presentation can vary, emphasizing different aspects based on the visual guise chosen. We will explore at least three types of visually auditory images: the waveform, the spectrum, and the spectrogram.

The classical silhouette of a sonic event, known as the *waveform*, illustrates the trends of amplitude values in the time domain (fig. 4.2). By examining a waveform, one can easily identify the envelope of certain sounds, transients, and temporal evolutions of a track. However, frequency values are not immediately accessible.

From a historical perspective, *waveforms* emerged as the initial form of auditory images, predating the invention of the gramophone and playback technology by about twenty years. This innovation was made possible through an instrument called the *Phonautograph*, patented by the Frenchman Édouard-Léon Scott de Martinville in 1857. The Phonautograph had a unique ability to transform vibratory signals into a graphic representation—namely phonautogram—using a mechanical process inspired by the functioning of the human ear.

The Phonautograph in fig. 4.3 featured a wave-guide for channeling the propagation of the vibratory source (p), culminating in a thin, sensitive plate resembling a tympanum (m). A needle on the inner side (s) registered the vibrations on a moving roller or a memory (e). Martinville envisioned his invention as an innovation in communication, imagining a new form of stenography based on

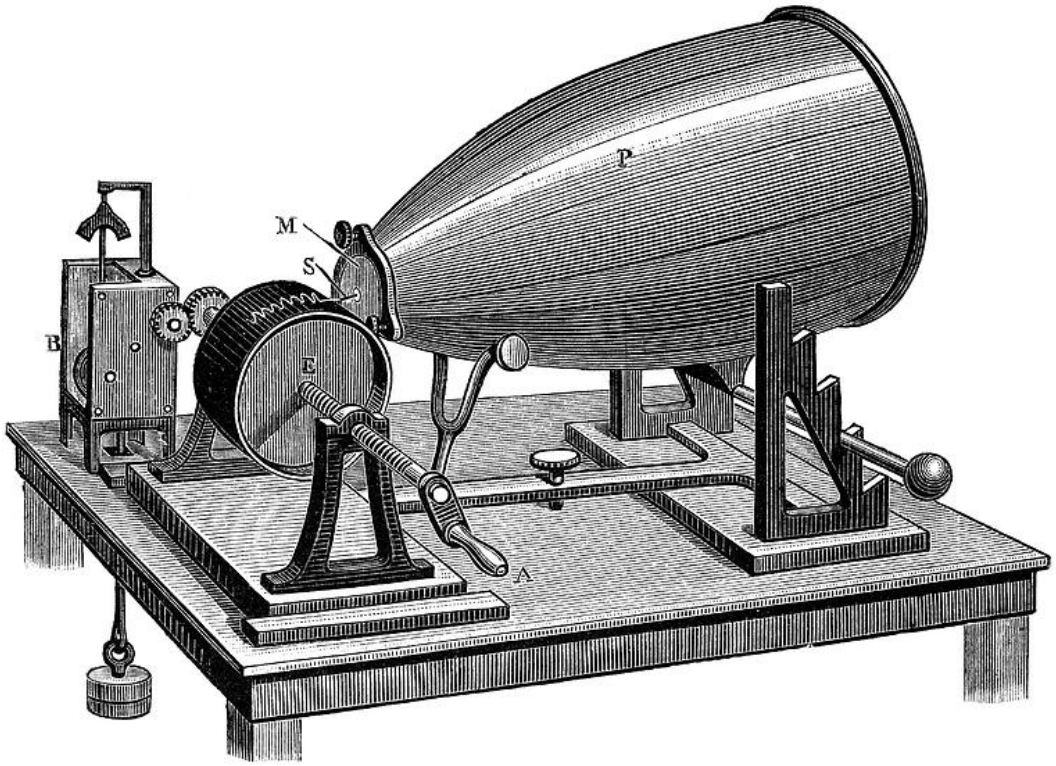

Figure 4.3: an illustration of a Phonautograph from 'Physique Populaire' by Emile Desbeaux, 1891. Source: https://sciencephotogallery.com/.

sound representation. However, the significance of this invention was ultimately overshadowed by Edison's phonograph, which allowed for the *reverse process* of sound reproduction.

On the other hand, the *spectrum* is a visualization of the waveform in the frequency domain, with the latter being obtained through Fourier transformation of the time domain signal (fig. 4.4). Common spectral representations illustrate the amplitude of the frequency components of a waveform versus frequency. This visualization is helpful for identifying the most relevant contributions in the frequency domain, determining if there are specific relationships between the most prominent components (such as harmonic ratios), or identifying significant broadband distribution of the waveform energy.

Lastly, the *spectrogram*, mixing the advantages of both waveform and spectrum, is a graphic representation of a signal. It exhibits time values on the x-axis, frequencies on the y-axis, and amplitude within the domain of color. While the one in fig. 4.5 is the most prevalent representation, numerous other visual forms of spectrograms can be devised.

Returning to Kulvicki's framework, let's now examine the rightful place of visual representations of signals within the realm of sound images. *Waveforms, spectrums, and spectrograms* are visual

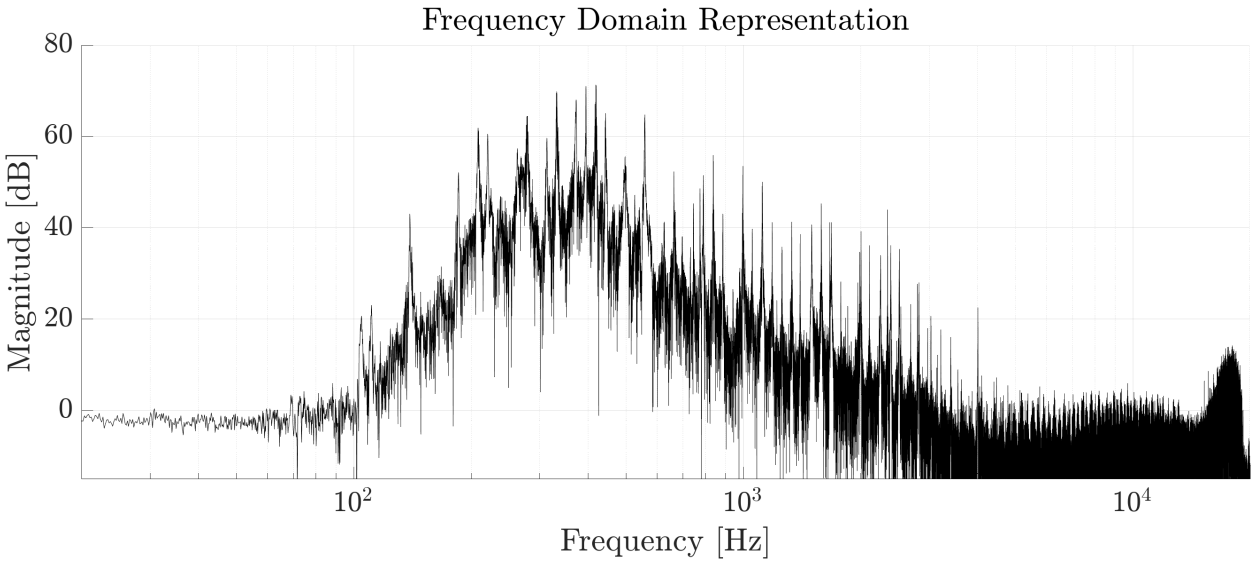

Figure 4.4: the spectrum derived from the waveform shown in fig. 4.2.

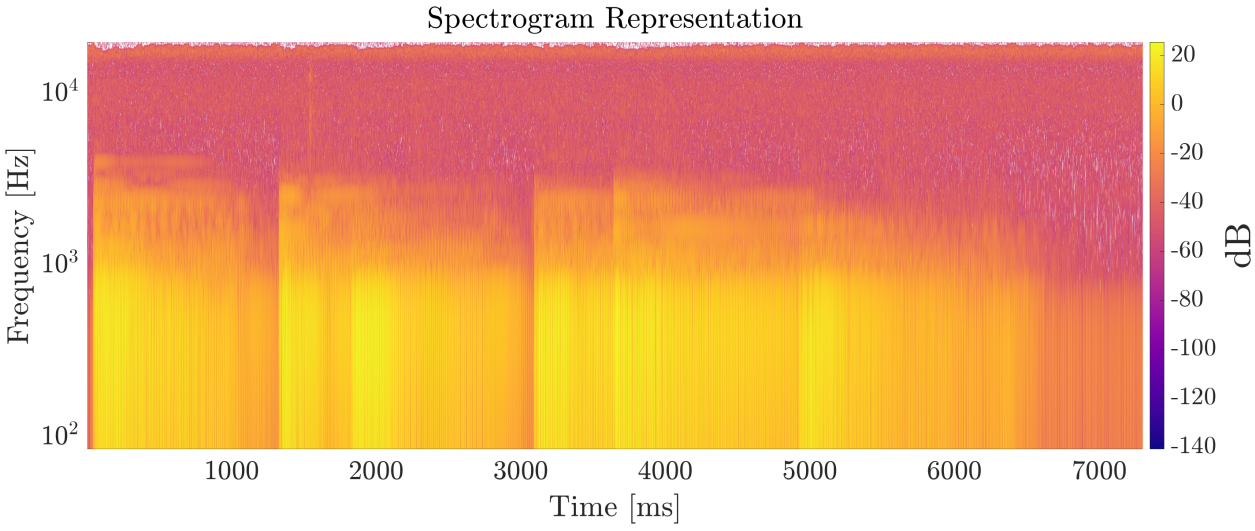

Figure 4.5: the spectrogram derived from the waveform shown in fig. 4.2.

images rooted in the recording process. They all originate from the same *skeletal content* at play in sound reproduction, and anything in the form of a waveform can be transformed into an audible signal, as evidenced by the reproduction of a Martinsville's phonautogram from 1960. However, these types of sound images, unlike audio reproduction tokens, *do not instantiate the acoustic qualities they represent*. Consequently, we should regard them similarly to graphs and diagrams, standing as non-depictive sound images.

Up to now, we have presented Kulvicki's theory of recording-based representational systems as a suitable framework for identifying those relevant aspects that distinguish recordings from other kinds of images. Furthermore, we expanded such a framework by proposing alternative, non-depictive forms of rendering.

4.2 A recording's content

What does a picture represent? Generally, pictures like photographs represent various scenes of the world around us. Often, they portray couples with long white dresses and black elegant suits, hands in hands; other times, they capture the magic of a beautiful landscape in the light of the sunset. Pictures must always be about something. As we've seen, some of them, precisely photographs, share structural similarities with audio recordings, and, as Kulvicki points out, "you don't have a recording absent something that it records" [37](p.271).

What a picture represents is called the *image-subject*, which is what we recognize by looking at the representation, or *image-object*. For example, when you stare at a photo of a young boy wearing vintage clothes and then recognize the person behind the glass—it's my father! Discovering a subject means determining the right chain that leads from the *signifier* to the *signified*. Photographs are supposed to point to a precise target: a boy in the spring of 1985. Other images like paintings, drawings, and various forms of pictorial art may point to vague or abstract subjects—if any.

Most importantly, images can often be *ordered* and *classified* based on the subjects they represent. When approaching the art of the camera, we can distinguish numerous types of subgenres such as portrait, landscape, macro, street, wildlife, and aerial photography, among others. How do we determine the kind of a picture? Simply by recognizing its subjects, we can ascertain, counterfactually, to which category it belongs. No landscape photo lacks a landscape, just as no wildlife photography lacks wild subjects.

A similar framework, I believe, can be applied to sound images, involving the classification of these images based on their represented content. We may recall encountering a categorization

of recordings proposed by O'Callaghan (3.3.2), which distinguishes between real-time, non-real-time, and musical recordings. These categories vary in terms of the aspects in which they exhibit deception: *spatial* deception, *spatiotemporal* deception, and *aesthetic* deception.

It is crucial to note that this classification doesn't necessarily indicate a *substantive difference* in the subject matter; rather, it highlights distinctions in perception. Firstly, anything audibles in real-time can also be reproduced later. Secondly, sounds can serve a musical role and carry aesthetic value depending on contextual factors. In addition, as we discussed in the introduction, the aesthetic halo is a concept that transcends *pure sounds*, operating on a more abstract level (the realm of art). Therefore, attempting to draw a strict distinction between musical and non-musical recordings may overlook differences at a more fundamental level.

This, however, does not dissuade us from recognizing the existence of distinctions at the aesthetic level that also unveil ontological differences at the sound level. This becomes apparent, in my view, in Edidin's approach to musical recordings [20], where he introduces a tripartition based on the kind of performances they reproduce. Let's delve into his theory.

Edidin's focus lies in exploring the extent of the impact of recordings on the realms of *aesthetics* and the *metaphysics of music*. The widespread dissemination of musical copies can be discerned in social and economic dimensions. However, from an artistic standpoint, the significant changes manifest in an unprecedented transformation of the listening experience: music can now be consumed by private and distracted ears, equipped with the ability to silence an entire orchestra at the touch of a button.

Conversely, recordings also reshapes the production of music, with musicians performing in warm and dry environments for an audience of microphones rather than people. For Edidin, this revolution provides a unique opportunity to illuminate "relations among musical activities and objects that antedate the technology of recording, particularly among performances, compositions, and the act of listening to both" [20](p.24). Within his work, he identifies three distinct types of recordings: recordings of performances, recordings of compositions, and recording artifacts. In the following sections, we will initially present these categories as originally outlined by Edidin and subsequently explore their potential significance in the realm of sound.

4.2.1 Musical recordings

The initial category of recordings describes by Edidin, namely *recordings of performances*, closely aligns with the original concept of sound playback. An illustrative example is presented in the

opening of chapter 3, where the file containing the concert performance falls precisely into this category. In a live music setting, such as a concert, every sound reaching the audience is generated in real-time during the unfolding of the show. The Neil Young album "Unplugged" also falls under this category.

In the musical context, the term 'performance' denotes the act of producing an audible rendition of a musical work. However, akin to sounds, performances are often perceived as unique occurrences, incapable of repetition. To borrow the words of Kania:

> A live performance of musical work is a *singular entity*, in the sense that although two performances might be similar in all sorts of way, they will remain two distinct events; two musical events cannot be single performance. A musical work, on the other hand, is a *multiple* entity, in the sense that two distinct performances may be performances of the *one and the same* work. To illustrate: you may attend two different performances of Brahms's violin concerto, but you can only ever hear one violin concerto by Brahms, because he only wrote one [32](p.23).

Edidin proposes a solution to the challenges posed by reproducing performances through recording playback, echoing O'Callaghan's perspective. A performance is a complex event, rich in aesthetic nuances perceivable in various ways, with a *proper part* being an acoustic event. Edidin refers to this acoustic event as the 'sounding' of a performance. While we can gain access to the sounding through recordings, the performance itself retains an *ephemeral nature*, existing only within the temporal window of its occurrence. Consequently, engaging with a recorded performance provides a somewhat circumscribed aesthetic experience.

The scenario shifts when it comes to *recordings of compositions*. This category of recordings allows the listener to experience a musical work, such as Mahler's Symphony No.1, independently of any specific performance. This capability stems from the fact that, for the most part, recordings of compositions are crafted by *assembling various performances or sections of performances* (referred to as 'takes') rather than capturing complete concerts. As Eisenberg aptly describes recordings of compositions, they are "pieced together from bits of actual events, they construct an ideal event . . . They are like the composite photograph of a minotaur" [21](p.109).

Building on Edidin's perspective, listening to a vinyl recording of Mahler's Symphony No.1 allows us to engage with the composition—an abstract entity instantiated each time a corresponding performance takes place. However, such an assertion raises certain concerns. Traditional musical

aesthetics theorists argue that the proper approach to a musical work is through performances. Consequently, since recordings are not performances, they may fall short in providing an *authentic listening experience*. This stance is rooted in *contextualism*, asserting that artworks are best appreciated within the context of their creation.

As Kania suggests, the potential resolution to this dilemma lies in viewing recordings as ephemeral as performances, aligning with Udhir's proposal[6]. In terms of our objectives, the *reproducibility* of recordings, facilitated by the translation of sounds into a *preservable format*, is precisely what characterizes them as sonic images. Referring briefly to Kulvicki's framework, this preservability can be seen as a form of non-depictive representation. Sound images, such as waveforms, exemplify this concept by illustrating the skeletal content of an auditory event without embodying any audible properties. This capability allows for the *preservation* of acoustic information, *sparing it from its inherent ephemeral nature*.

The acoustic information stored in this manner plays, concerning the reproduced token, the role of a 'template'. This template outlines the set of passages that a speaker should perform to reproduce a specific sound, akin to instructions for a Lego set or an architectural project for a building. Moreover, being a template implies possessing a sort of enduring nature. Reflecting on our earlier exploration of the semantic ambivalence of the term 'recordings', particularly in its secondary meaning, I am inclined to believe that what is commonly referred to as the sample might indeed be this template—*the stored skeletal content of a certain sound event*, allowing for the subsequent production of sounding tokens[7].

In addition to the questions stemming from Edidin's argument about listening to a composition through its playback, the third and final category of recordings he introduces is termed *recording artifacts*. In musical practices like jazz, recordings typically serve to make us aware of performances, and in the classical Western tradition, they help us appreciate compositions. However, the function of artifacts differs, as it is not intended to make us aware of something else. Simply put, a recording artifact *represents the sample itself.*

Recorded works of electronic music typically fall into this category. In electronic music, performances, involving sounds from various acoustic instruments and voices, often serve as the raw material for creating songs. However, the predominant part of their creation lies in the process of modifying and arranging these materials, ultimately giving form to a sample [8]. To illustrate:

[6]See 3.3.2.

[7]I credit this idea of a template, once again, to the work of Kania [32](p.25).

[8]Scholars such as Gracyk and Kania regard also Rock music, seen as an expansive artistic tradition intricately

> Using technology developed to record and reproduce the sounds of performances, they [electronic composers] begun creating tapes that when played back produced sound events that could not be considered an accurate record of any performance occurring in the studio in any sense. Any authentic copy of the master tape produced an authentic instance of the work when played back. In such "electronic music", the sound of the work, in an important sense, came straight from the composer without the mediation of a performing artist [31](p.401).

4.2.2 Three kinds of sonic pictures

Performances, compositions, and artifacts, according to Edidin, categorize all musical recordings under one or more of these three types. The question arises: does this partition extend to encompass any type of recordings, whether musical or non-musical? I contend that even when removed from its original *aesthetic context*, this classification can be applicable in the broader domain of sounds. To illustrate this point, let's reconsider each of these three categories, abstracting them from their initial aesthetic significance. From a purely auditory perspective, a musical performance is essentially an acoustic event—the occurrence of vibrations from a source. The awareness that what we are hearing is the rendition of a symphonic composition rather than environmental sounds *plays no role* in altering the nature of our auditory objects.

Firstly, a *recording of performances* in the ontology of sound delineates a representation of a *cohesive event*. The sound image, exemplified by the playback of the recording, serves to depict a unified auditory experience, closely resembling the one we would have encountered if we were physically present when the original event occurs. This is notably evident in situations like phone conversations, where the listener may perceive the interlocutor's voice emanating from the phone's speakers as a *seamless*, *unified flow*, without the awareness of multiple isolated takes containing individual words, compiled into phrases.

Similarly, a *recording of composition* aligns with its sonic counterpart in recordings of *disparate* (or scattered) *events*. In this case, various sonic events are juxtaposed and arranged to create the illusion of a cohesive auditory experience. This practice is not limited to music recordings but extends to other domains as well. In the realm of film, for instance, audio is frequently crafted by combining fragments of cinematic auditory effects, soundtracks, and at times actors' voiceovers, seamlessly blended with sounds recorded on the set.

---

reliant on recording technology, as a generator of recording artifacts rather than live performances. See [31] and [26].

Finally, *recording artifacts* represent recordings whose content does not strictly pertain to a specific environmental auditory occurrence. They fulfill the same role both in the ontology of music and in the framework of sonic studies. A recording artifact may originate from an acoustic source, influenced, as mentioned earlier, by the extensive application of distortion to the sample (synthetic reverb, delay, filter, compressor, and so forth). However, at this stage, we should refrain from imposing stringent conditions on what constitutes a recording artifact, as it proves challenging to pinpoint the extent to which a recorded sound ceases to reference its original archetype. However, an example of a genuine recording artifact, in my estimation, is a track generated from an AI program. For instance, when we use AI to generate voices from text, the resulting sound image has no direct referent in the world itself. Such recordings are *synthetic* in a strong sense.

An objection to this standpoint could be raised, arguing that a recording of an artifact might result in capturing a unified event—the vibration of the loudspeaker. This potential issue arises because two *twin sound events* could lead to one being categorized as an artifact and the other as a unified event, presenting a counterintuitive outcome. To some extent, it is accurate that a recording of an artifact results in a recording of a continuous signal, where neither multiple takes nor modifications occur during the process. However, one might contend that the relation 'O represents S', pivotal in depictive systems like sound reproduction, is *transitive*. In other words, if O represents S and Q represents O, then Q represents S, where S is an artifact. This holds true even when the same witless process is adopted, presenting *the same skeletal content captured and preserved*. From a more philosophical perspective, this property of representational systems reproducing the same token or content within multiple iterations of the process is what Kulvicki terms 'transparency'[9]:

> Roughly, given a representation of some object X, the result of iterating the process by which one gets the representation of X is another representation that is just like the representation of X one got in the first place, as far its syntactic identity goes. Imagine the case in which a clear, focused photograph is made of another picture. As long as the photographic plane and the plane of the picture are parallel to one another the resultant photograph is just like the original picture with respect to the determinate shapes and colors of all regions on the picture surface [35](p.331).

Once again, the affinity between sound recording and the photographic process becomes evident.

[9]With these terms, Kulvicki refers to the property of certain representational processes to generate the same token through iterative application, such as taking a photo of a photograph [35](p.330).

Kulvicki's concept of transparency, a property shared by members of the pictorial representation class, finds resonance in sound recording.

In summary, we've explored the possibility of establishing a classification for recordings based on their r*epresented subjects*. Drawing inspiration from Edidin's classification of musical recordings, which includes performances, compositions, and artifacts, we sought to extend this categorization to the broader domain of sounds, detached from its original aesthetic context. Our proposition suggests that recordings of performances generally represent *cohesive events*, recordings of compositions portray *scattered sound events* arranged together, and recording artifacts produce sounds that d*on't represent any specific event*. In the forthcoming conclusion, we will delve into further questions regarding the capacity of sounds to represent other sounds, especially in the presence of acoustic differences.

4.3 Individuating sounds

As we approach the conclusion of this exploration into recordings through the lens of sonic studies, our journey has taken us from the pure ontology of sounds and auditory perception to the formulation of a theory of *representation for sounds*—an unexpected outcome when embarking on this topic. Yet, as I continue to believe, the trajectory followed thus far offers the most comprehensive approach to investigating the entire spectrum of recordings experience. Like navigation, where we circumnavigate a land to delineate its proper boundaries on a map and determine its continent status, philosophical exploration necessitates approaching less-explored arguments by noting all the various facets in our diary.

Similar to how a series of white sandy shores followed by sharp rocky cliffs can represent two facets of the same land, the domain of audio recordings exists as a borderland encompassing various sciences and practices. I hope to have presented the argument in a lucid manner, providing future scholars with an organized, albeit potentially incomplete, navigation chart. Furthermore, in this concluding chapter, I've begun annotating certain thoughts and hypotheses that still harbor concerns and doubts regarding recordings, signaling areas for further exploration in the future.

In these final pages, I aim to delve into the concept of *subject recognition* when engaging with audio recordings. Similar to photographs, sound recordings can be viewed as images in the auditory domain, with the presented subject being a key consideration. As I previously suggested, there are three types of original sources in recordings: *cohesive events*, *disparate events*, or *recording artifacts*. Only in the case of cohesive events can we genuinely perceive the original event auditorily

and recognize it in the reproduction. Neither in the case of disparate events nor artifacts can we have experienced the subjects before their depiction—in the former because the original sounds did not occur simultaneously as presented now, and in the latter because artifacts represent nothing but their sample, and the sample remains inaudible except through playback.

Consequently, for the vast majority of recordings we listen to, we lack firsthand knowledge of the original sound events. Is this a problem for recognition? Not necessarily. Drawing a comparison with painting, we can recognize the archangel Gabriel in an Annunciation picture by observing the way he is portrayed (bowed, with a lily in his hands), even though we may not possess firsthand knowledge of him. Similarly, we can recognize a Beatles tune even though, unfortunately, we have never been with them in the studio.

The process of recognition in audio recordings involves a *matching of audible qualities*. When we recognize the sound of our morning alarm, we essentially identify a sonic event whose frequency pattern, envelope, duration, and volume match the regular sound produced by the alarm. From a phenomenological standpoint, recognition "may, at a minimum, consist in nothing more than a disposition to identify an object currently perceived as one encountered in the past, together with a feeling of familiarity with it" [42](p.137). Within recognition, two essential features should be considered: (a) the *reconfiguration*, i.e., how recognition affects experiences of things; (b) the *generative* capacity of recognition—from one act of recognition, we gain the ability to iterate it across a range of variations in appearance of the object.

As observed, depending on the context, we often demonstrate the ability to recognize sounds through playback even when they exhibit dissimilarities compared to the original events. A prime example is in telephonic conversations, where the reproduced voice may sound metallic and distorted, yet remains intelligible with respect to its message and relevant features for identifying the speaker. In a nutshell, we do not individuate sounds like Funes would do. Then how we recognize original sounds in recordings?

Firstly, it's important to note that we do not recognize sounds in their entirety. Generally, sounds—particularly the majority of non-synthetic sounds—present a *complex spectrum* to the ear, offering an immense amount of information to be processed and stored. However, given that our brain evolved to be functional rather than precise in representing phenomena, it engages in a series of unconscious operations to *compress* perceptual data into a sort of "fingerprint". A sound fingerprint can be understood as a set containing at least the minimum information necessary to recognize a sound. These fingerprints can vary among individuals; for instance, while a non-

musician may commonly possess a fingerprint of just a piano sound, an expert could distinguish between Bosendorfer, Steinway, or Fazioli pianos simply by listening to one of them being played.

In the process of compression, a portion of the captured information becomes redundant. Consequently, this kind of auditory surplus *remains masked*, failing to enter our mental representation of sound. This becomes evident, for instance, when we first practice music and learn to distinguish tones by isolating dominant frequencies. The field dedicated to studying this type of compression in sound reproduction systems is known as Perceptual Coding. It originated in the 1950s with the scientific exploration of the human ear, primarily in the context of telephonic industries aiming to optimize communication (AT&T laboratories). Perceptual coding relies on two fundamental concepts from psycho-acoustics, specifically the theories of *critical bands* and *masking*:

> Masking proposed that a louder sound could "hide" a quieter sound of a similar frequency content from the ear. The theory of critical bends proposed that masking effects could be conceived in terms of frequency regions somewhat like highways in the ears. With knowledge of critical bands masking response can be predicted [59](p.96).

As Sterne contends, these studies provided the foundation for the development of standard digital formats for sound reproduction, such as mp3 compression. Given that recording technology is primarily constrained by the *amount of information* it can store, both in temporal and quality dimensions, perceptual coding has played a crucial role in *optimizing* functionality. This optimization has, in turn, facilitated the hysterical dissemination of tracks that characterizes the age of the internet.

The challenge of sound recognition in recordings can also be reframed as a machine-to-machine problem, rather than merely machine-to-human. This shift in perspective involves the consideration of *algorithms*, specifically audio search engines, designed to identify tunes by recording brief segments of an audio event. One of the most familiar tools employing this approach is Shazam, which has provided people with a tunes recognition service since the early 2000s. Connected to an extensive library of samples, Shazam possesses knowledge of a far greater number of tunes than any human being. In carrying out its task, Shazam generates a spectrogram of the recorded source, distinguishing between noises and more likely segments of a song. Subsequently, it identifies a constellation of peaks, akin to a map or fingerprint, which is then compared with files stored in its memory. If the generated map can be overlaid onto one of the tunes in its memory, then the sought-after song is identified [68].

However, the astonishing aspect of sound recognition performed by humans lies in our ability, despite limited memory, to recognize subjects in recordings well beyond the capacity of any machine. For instance, Shazam cannot identify a tune if it is pitched up or down, a task we accomplish easily. Perhaps, with advancements in machine learning, more performative algorithms will emerge, yet they may never fully capture those instances in which a sonic image vaguely refers to a subject—instances characterized by a level of distortion in sound that would lead us to consider them as recording artifacts.

While hi-fi reproduction systems and binaural technology contribute to an increased awareness of sound recurrence and auditory deception, the emergence of distorted and poor-quality audio playback raises concerns within the framework of sound image, as we are discussing. These signals *underscore our extensive ability to recognize auditory subjects in sonic representation*. Drawing parallels once again to the world of art, identifying a woman in "The Girl With A Pearl Earring" by Vermeer is straightforward, but the real intrigue lies in recognizing a body in "Nude Descending a Staircase, No. 2" by Duchamp. To pose the question directly: How does an acoustic representation *maintain consistency* in presenting its subjects even with a certain degree of distortion? Specifically for recordings, if there is a limit, what threshold must be surpassed for the playback to cease representing its original subject?

4.3.1 Representational systems and their structures

In an attempt to address these questions, I will draw a parallel with what Kulvicki introduces as *projective transformation* in his work on images [35]. Within the formal theory of representational systems, he posits that, for a specific type of representation like a linear perspective picture, iterating the process with a certain degree of "distortion" still maintains the image identity, thereby producing the same syntactic type. Although he originally applies this argument exclusively to forms of visual depiction, I believe it holds potential application in the domain of sound. While I'll introduce this idea without complete development due to the vastness of the subject, delving into what he precisely means by image identity and syntactic type becomes crucial for making this comparison.

In his work, Kulvicki proposes the possibility of distinguishing *pictorial* systems from *linguistic* and *diagrammatic* forms of representation by emphasizing their formal structure rather than relying on a theory of perception, focusing on how we perceive pictorial images. Rooted in the tradition of Goodman's work [25], this approach rejects any naive reliance on the relation of *similarity* and *resemblance*, which may encompass forms that evoke concerns and strong intuitions to the

contrary. Goodman's renowned argument against the use of intuitive resemblance in explaining how pictorial representation functions can be summarized as follows: 'representing' and 'resembling' are both two-place predicates—'x represents y' or 'x resembles y'—yet while the latter is *symmetric* and *reflexive*, the former is not. Indeed, everything resembles itself, and if something resembles something else, the same holds true even when the order of the terms is changed. To clarify, as Goodman asserted, "a girl is not a picture of her twin sister", even though both may resemble each other to the highest degree [24](p.437). Furthermore, what renders resemblance *paradoxical* in explaining depiction is that many objects can resemble other objects in very particular aspects:

> All objects resemble one another in indefinitely many respects. For example, I share qualities with inscription of my name "John". We both have mass, color and shape. We are almost the same distance from the sun. We are both vulnerable to fire, and composed of carbon. These similarities are uninteresting, but they do exist. You would not mistake me for my name, but then again you would not mistake me for my portrait either. Goodman thinks that it is uninteresting that pictures resemble their objects, since they resemble everything as well [36](p.53).

Embracing a Goodmanian perspective, Kulvicki advances a set of structural conditions (necessary and sufficient) that a representational system must satisfy to be considered pictorial. These conditions consist of *relative repleteness*, *relative syntactic sensitivity*, *semantic richness*, and *transparency*. However, before delving into these conditions, it is essential to take a step back and briefly introduce the general account of what a representational system comprises in terms of structure.

Firstly, every system encompasses a set of objects deemed representations, namely *tokens*, which are further categorized into *orthographic* types and *syntactic* types. For instance, the word 'car' can be instantiated into various orthographic types (block letters, capital letters, italics), but all occurrences of "car' refer to a precise string of letters defining its syntactic type. Additionally, syntactically identical tokens are also *semantically* identical, although the reverse does not necessarily hold true.

The properties of a token relevant to its classification into a particular type serve as a *supervenience base* for the emergence of the token's *identity*. In other words, we can alter the shape of a letter without changing its identity, but it is impossible to change the letter's identity without altering its shape. Providing a *full characterization* for a form of representation involves defining

the set of relevant properties, $\Phi = \{P_1, \ldots P_n\}$. To illustrate with another linguistic example, dimensions and chemical composition are not pertinent for determining which alphabet letter a certain sign represents. In contrast, "the set of determinate planar shape is a full characterization of the statistically relevant properties for an alphabet, since the identity of any token letter supervenes on its shape" [35](p.325). Assuming this framework, a depictive system's fully characterization must satisfying these four conditions:

1. **Relative repleteness**—System A is considered more replete than another system B if the *full characterization* of A includes that of B as its *proper subset*. A clear illustration can be drawn when comparing diagrams and pictures: more properties, such as the color of the background and the thickness of lines, are relevant to the identity of a picture's token in comparison to a diagram's token. Moreover, all the relevant properties of diagrams are encompassed in the full characterization of pictures. In Kulvicki's view, unlike Goodman's, the replete value can be achieved even when one set of properties is not *entirely included* in the other. Thus, we should regard A as the more replete system if its full characterization's cardinality, excluding the intersection with B, is greater than B, excluding the intersection with A.

2. **Syntactic sensitivity**—This condition stipulates that for depictive systems, a token's identity, supervening on a full characterization, is sufficiently *sensitive to changes*. The sensitivity value of two systems can only be derived from their intersecting properties. For example, many modifications in the shape of a letter do not jeopardize its identity. On the other hand, pictures are much less tolerant of changes in shape since even a small additional line could compromise their syntactic type. More formally, we can say that "one representational system is syntactically more sensitive than another if and only if the changes in syntactically relevant properties sufficient for a change in syntactic identity in the latter are properly included among the changes in syntactically relevant properties sufficient for a change in syntactic identity in the former" [35](p.327).

3. **Semantic richness**—This condition concerns how *syntactic types* are connected to *semantic types*. A pictorial system, which is both replete and sensitive, must combine its rich syntax with an equally rich set of denotations. Languages, even though not pictorial, satisfy this condition since "the number of possible denotations [meanings] is at least as great as the number of possible syntactic types [words]" [35](p.328).

4. **Transparency**—While repleteness, sensitivity, and richness are *necessary* conditions for a

system to be pictorial, these conditions alone are not sufficient to determine a system as purely pictorial. Introducing this additional condition, Kulvicki employs a paradigmatic example of a form of pictorial representation: *Linear perspective* (LP). LP is one of the most intuitive forms of perspective, obtained by determining a *projection point* from which emanating rays (or vanishing lines) extend toward the object one desires to represent and a *picture plan*e on which to realize the representation. Even color is relevant in LP, and thus, the representation and represented hues must match[10]. Correctly executed LP pictures produce very realistic tokens and photographs approximate exactly this process. Now, consider *iterating* the process by putting another picture plane in the space between the old plane and the projection point to realize a representation with the previous image as its subject. For simplicity, suppose that the new plane is parallel to the old one. What this new picture represents will be nothing but *the same object depicted by the previous picture*. With this in mind, transparency can be presented as follows: "a representational system S is transparent just in case for any token representation, R, in S, any representation of R in S is of the same syntactic type as R" [35](p.330).

Linear perspective is characterized by this kind of transparency. When we repeat the process multiple times, each iteration yields a representation that retains t*he same identity as the original*. This principle is not exclusive to linear perspective; it's also observable in other depictive systems. For example, a photograph of another photograph still represents the same subjects; similarly, a correctly made record of a vinyl produces an identical duplicate. However, it's important to note that not every iteration necessarily shares all relevant properties with its predecessors. For instance, as Kulvicki points out, in linear perspective, nothing mandates that each new picture plane *must be parallel* to the previous one. This can result in *distortions of the token's shape*. This raises a question: does a distorted picture still belong to the same syntactic type? This challenge is significant for transparency, as properties like shape and color are crucially important in pictorial representations.

Kulvicki suggests that we should accept this kind of transparency, even when distortions arise from the convergence of picture planes. This phenomenon, known in geometry as *projective transformation*, is a fundamental characteristic of linear perspective and perhaps of pictorial systems more broadly. While this may seem like a preanalytical assumption, there are arguments that support it. First, projective transformations adhere to strict geometrical rules that maintain a consistent

[10]I recommend using the illustrative images from Kulvicki's original paper to better visualize the design of Linear Perspective (LP).

mapping between points and lines across planes. Furthermore, viewing the product of an iteration as possessing a different identity, despite still being a projection of the original subject, seems untenable:

> It is certainly intuitive to call pictures that share *all* determinate shapes and shades alike syntactically and, hence, semantically, but pictures that differ in the ways that they can under arbitrary perspective maps do not intuitively share content. One motive for calling two possible tokens in a representational system syntactically *distinct* is that it is possible for them to be representations of different things: Syntactically identical representation *cannot* be representation of different things [35](p.332).

Including projective invariance in linear perspective can be problematic, as projections can occur in unique spaces like curved planes or multidimensional settings, influencing distortion in new tokens. Hence, Kulvicki argues for combining projective invariance with a clear notion of a picture's content to ensure transparency is maintained, even in cases of significant distortion.

4.3.2 Sonic projections

The concept of *projective invariance* in pictorial systems, as discussed by Kulvicki, might opens the door to understanding distortions in various representational systems, including audio. Kulvicki's alignment of sound playback with other pictorial systems under the conditions of repleteness, sensitivity, richness, and transparency, and its mimetic nature, raises an intriguing question: Is there an equivalent of *projective invariance* in the *audio domain*? This framework could offer insights into addressing sound distortion, such as the representation of original sounds in muffled or metallic phone conversations, and distinguishing between audio images and recording artifacts.

Further, this approach seems supported by the *digital treatment* of audio files. Audio files, which encode sonic information, are stored in digital formats that can be converted into temporal signals. Despite being laden with metadata and other codes, the essence of these files lies in the samples—a strings of quantized information representing the waveform, akin to collections of 1D arrays or vectors. Thus, digitalization provides a generalized algebraic tool applicable to various representational systems.

Firstly, it's important to note that every time we reproduce a recording, we're performing an *iteration* of the process. As stated earlier, the recording process, unlike the copying process in daguerreotype photography, consists of two phases: *recording* and *playback*. Conceptually, the

same witness process is at play in both phases. However, in the recording phase, the process converts a periodic signal into samples, while playback inversely converts samples back into periodic signals. Each correct playback of a vinyl record produces a sound token of the same syntactic type as the sound encoded in the vinyl's grooves. The information in these grooves acts like an identikit. Taken *as a whole*, the recording-playback process is transparent, in the sense that every subsequent recording will produce the same sounding token. However, it's also important to consider this as a c*omposite process*, recognizing that reproduction alone is an instance of iteration.

Secondly, assuming reproduction counts as iteration, it's worth suggesting that certain types of distortion in the reproduction process might not jeopardize the identity of the outgoing token. The most intuitive candidates for such distortions are those that modify sound representation's relevant properties, namely *amplitude*, *frequency*, and *duration*. These are rudimentary sound manipulation techniques, possibly as old as recording technology itself.

Modifying a signal's amplitude changes the perceived *volume*. For instance, adjusting the volume knob on an amplifier doesn't seem to alter the identity token significantly. This is especially true considering that, with the exception of the phonograph, signals through loudspeakers need amplification to be perceived. However, operations like *compression* or *limiting*, which change the relationship between transients and steady states in the waveform, might be more untactful. While I believe the identity of recording playbacks remains invariant to *"linear" modifications of amplitude levels*, I would be cautious about considering them invariant to operations like extreme compression.

The same framework applies when considering *pitch* and *duration*. Pitch is the perceivable aspect of frequency, determining a sound's height. Transposing a sound involves increasing its frequency by a certain factor. While a piano keyboard depicts pitch linearly, the true representation in terms of frequency is *logarithmic*. For example, for two notes an octave apart on a keyboard, the frequency of the lower note is half that of the higher one. Therefore, *doubling* the pitch value in a recording playback results in a sound token transposed one octave higher.

The relationship between duration and pitch is also significant. Normally, changes in pitch correspondingly alter the signal's duration. A common example is playing a vinyl LP at different speeds, such as switching from 33 rpm to 45 rpm. The resulting pitch distortion is not only a practical tool but also has aesthetic applications in electronic music. Despite these kinds of distortions in pitch and time values, we generally retain the ability to recognize recordings. A clear demonstration of this is our recognition of slowed-down versions of songs, a genre quite popular

in 2023, which are essentially altered instances of the original tune

However, the situation changes when modifications to pitch and time are applied non-linearly. Consider a *hypothetical machine* that alters the pitch of classical music recordings note by note, changing major keys to minor and vice versa. Such a machine, whether it exists or not, would likely fail to preserve the original identity of the sound token. Similar to the case of amplitude, recordings maintain their identity concerning linear modifications in pitch and duration.

What can we draw from this? It seems that the identity of a sound token does not solely depend on specific values of sound-relevant properties like pitch or volume. Instead, what appears crucial for a recording to be a faithful representation of another sound is the preservation of *proportions between elements that compose its progression*. It's not about a specific pitch or volume, but rather a *harmonious blend of auditory information*, interconnected by a structural 'backbone' that upholds the identity. However, this concept remains a preliminary hypothesis, and I intend to leave the question open for future, more thorough investigation.

Bibliography

[1] John Langshaw Austin and Geoffrey James Warnock. *Sense and sensibilia*. Vol. 83. Clarendon Press Oxford, 1962.

[2] Richard Beaudoin and Andrew Kania. "A musical photograph?" In: *The Journal of Aesthetics and Art Criticism* 70.1 (2012), pp. 115–127.

[3] Leo Leroy Beranek and Tim Mellow. *Acoustics: sound fields and transducers*. Academic Press, 2012.

[4] E. Di Bona. "The Method of Contrast and the Perception of Causality in Audition". In: *New Advances in Causation, Agency and Moral Responsibility*. Ed. by Fabio Bacchini at al. 2014, pp. 79–93.

[5] Myles Brand. "Particulars, events, and actions". In: *Action Theory: Proceedings of the Winnipeg Conference on Human Action, Held at Winnipeg, Manitoba, Canada, 9–11 May 1975*. Springer. 1976, pp. 133–157.

[6] J. Brandl. "Do Events Recur?" In: *Speaking of Events*. Ed. by Achille Varzi, James Higginbotham, and Fabio Pianesi. Oxford University Press, 2000, pp. 95–104.

[7] Albert S Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[8] Harold I Brown. "Direct realism, indirect realism, and epistemology". In: *Philosophy and Phenomenological Research* 52.2 (1992), pp. 341–363.

[9] Roberto Casati, Elvira Di Bona, and Jérôme Dokic. "The Ockhamization of the event sources of sound". In: *Analysis* 73.3 (2013), pp. 462–466.

[10] Roberto Casati and Jérôme Dokic. "La philosophie du son". In: (1994).

[11] Roberto Casati, Jerome Dokic, and Elvira Di Bona. "Sounds". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2020. Metaphysics Research Lab, Stanford University, 2020.

[12] Roberto Casati and Achille Varzi. "Events". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Fall 2023. Metaphysics Research Lab, Stanford University, 2023.

[13] Donald Davidson. "Eternal vs. ephemeral events". In: *Nous* (1971), pp. 335–349.

[14] Donald Davidson. "Events and particulars". In: *Noûs* (1970), pp. 25–32.

[15] Donald Davidson. "The individuation of events". In: *Essays in honor of Carl G. Hempel: A tribute on the occasion of his sixty-fifth birthday*. Springer, 1969, pp. 216–234.

[16] Elvira Di Bona. "Why space matters to an understanding of sound". In: *Spatial Senses*. Routledge, 2019, pp. 107–124.

[17] Elvira Di Bona, Vincenzo Santarcangelo, et al. *Il suono. L'esperienza uditiva ei suoi oggetti*. Raffaello Cortina Editore, 2018.

[18] Jérôme Dokic. "Two ontologies of sound". In: *The Monist* 90.3 (2007), pp. 391–402.

[19] Fred I Dretske. "Can events move?" In: *Mind* 76.304 (1967), pp. 479–492.

[20] Aron Edidin. "Three kinds of recording and the metaphysics of music". In: *The British Journal of Aesthetics* 39.1 (1999), pp. 24–39.

[21] Evan Eisenberg. *The recording angel: Music, records and culture from Aristotle to Zappa*. Yale University Press, 2005.

[22] Galileo Galilei and Sebastiano Timpanaro. *Dialoghi delle nuove Scienze. Il saggiatore*. Rizzoli, 1938.

[23] Stanley A Gelfand. *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2017.

[24] N Goodman. "Seven strictures on similarity". In: *Problems and projects* (1972).

[25] Nelson Goodman. "Languages of art". In: *Lexikon Schriften über Musik*. Springer, 2022, pp. 293–376.

[26] Theodore Gracyk. *Rhythm and noise: An aesthetics of rock*. Duke University Press, 2021.

[27] Timothy D Griffiths and Jason D Warren. "What is an auditory object?" In: *Nature Reviews Neuroscience* 5.11 (2004), pp. 887–892.

[28] John Haugeland. *Having thought: Essays in the metaphysics of mind*. Harvard University Press, 2000.

[29] Mark Eli Kalderon. *Sympathy in perception*. Cambridge University Press, 2017.

[30] Brian Kane. *Sound unseen: Acousmatic sound in theory and practice*. Oxford University Press, 2014.

[31] Andrew Kania. "Making tracks: The ontology of rock music". In: *The Journal of Aesthetics and Art Criticism* 64.4 (2006), pp. 401–414.

[32] Andrew Kania. "Musical recordings". In: *Philosophy Compass* 4.1 (2009), pp. 22–38.

[33] Anton Killin. "The polysemy theory of sound". In: *Erkenntnis* 87.2 (2022), pp. 435–458.

[34] Peter Kivy. *Introduction to a Philosophy of Music*. Oxford University Press, 2002.

[35] John Kulvicki. "Image structure". In: *The Journal of Aesthetics and Art Criticism* 61.4 (2003), pp. 323–340.

[36] John Kulvicki. *Images*. Routledge, 2013.

[37] John Kulvicki. "Recording and Representing, Analog and Digital". In: *Giving a Damn: Essays in Dialogue with John Haugeland* (2017), pp. 269–90.

[38] John Kulvicki. "The nature of noise". In: *Philosophers* 8 (2008).

[39] Calvin K. W. Kwok. "Hearing Waves: A Philosophy of Sound and Auditory Perception". PhD thesis. The University of Hong Kong, 2020.

[40] Jason P Leddington. "Sonic pictures". In: *The Journal of Aesthetics and Art Criticism* 79.3 (2021), pp. 354–365.

[41] Lawrence Brian Lombard. *Events: A metaphysical study*. Routledge, 2019.

[42] Dominic Lopes. *Understanding pictures*. Clarendon Press, 1996.

[43] D. L. C. Maclachlan. *Philosophy of Perception*. Cliffs Prentice-Hall, 1989.

[44] M. G. F. Martin. "Sounds and Images". In: *The British Journal of Aesthetics* 52.4 (2012), pp. 331–351.

[45] Phillip John Meadows. "In defense of medial theories of sound". In: *American Philosophical Quarterly* 55.3 (2018), pp. 293–302.

[46] Jorge Luis Méndez-Martínez. "What Counts as "a" Sound and How "to Count" a Sound, the Problems of Individuating and Identifying Sounds". In: *Synthesis Philosophica* 1.67 (2019), pp. 173–190.

[47] Richard Montague. "On the nature of certain philosophical entities". In: *The Monist* 53.2 (1969), pp. 159–194.

[48] Matthew Nudds. "Sounds and Space". In: *Sounds and Perception: New Philosophical Essays* (Jan. 2010).

[49] Matthew Nudds. "The Unitary Nature of Sounds". In: *Perceptual Ephemera*. Oxford University Press, June 2018.

[50] Casey O'Callaghan. *Sounds: a philosophical theory*. OUP Oxford, 2007.

[51] Casey O'callaghan. "XIII—Hearing Properties, Effects or Parts?" In: *Proceedings of the Aristotelian Society (Hardback)*. Vol. 111. 3pt3. Wiley Online Library. 2011, pp. 375–405.

[52] Brian O'Shaughnessy. *Consciousness and the World*. Oxford University Press, Jan. 2003.

[53] Robert Pasnau. "What is Sound?" In: *The Philosophical Quarterly* 49.196 (Jan. 2003), pp. 309–324.

[54] Kivy Peter. "Platonism in music: a kind of defense". In: *Grazer Philosophische Studien* 19 (1983), p. 109.

[55] Roger Scruton. "and Pure Events". In: *Sounds and Perception: New Philosophical Essays* (2009), p. 50.

[56] Roger Scruton. *The aesthetics of music*. Oxford University Press, 1997.

[57] Theodore Sider. *Four-dimensionalism: An ontology of persistence and time*. Oxford University Press, 2001.

[58] Matthew John Soteriou. "Sound and Illusion". In: *Oxford Scholarship Online* (2018).

[59] Jonathan Sterne. *MP3: The meaning of a format*. Duke University Press, 2012.

[60] S. S. Stevens and J. Volkmann. "The Relation of Pitch to Frequency: A Revised Scale". In: *The American Journal of Psychology* 53.3 (1940), pp. 329–353.

[61] Stanley Smith Stevens. "A scale for the measurement of a psychological magnitude: loudness." In: *Psychological Review* 43.5 (1936), p. 405.

[62] P. F. Strawson. "Causation in Perception". In: *Freedom and Resentment*. Methuen, 1974.

[63] Peter Frederick Strawson. *Individuals*. Taylor & Francis, 2002.

[64] Chris Swoyer. "Structural representation and surrogative reasoning". In: *Synthese* 87 (1991), pp. 449–508.

[65] David Toop. *Ocean of Sound: Ambient sound and radical listening in the age of communication*. Serpent's Tail, 2018.

[66] Christy Mag Uidhir. "Recordings as performances". In: *The British Journal of Aesthetics* 47.3 (2007), pp. 298–314.

[67] Peter Van Inwagen. “Creatures of fiction”. In: *American philosophical quarterly* 14.4 (1977), pp. 299–308.

[68] Avery Wang et al. “An industrial strength audio search algorithm.” In: *Ismir*. Vol. 2003. Washington, DC. 2003, pp. 7–13.

[69] Richard Wollheim. *Art and its Objects*. Cambridge University Press, 1980.