

---

# The Early Stage of Perception of Contemporary Art Music: A matter of time

---

VINCENZO SANTARCANGELO and RICCARDO WANKE

University of Bergamo, Polytechnic University of Turin, Italy. Email: [vincenzo.santarcangelo@unibg.it](mailto:vincenzo.santarcangelo@unibg.it)  
CESEM – Centre for the Study of the Sociology and Aesthetics of Music, NOVA University, Lisbon, Portugal.  
Email: [riccardowanke@gmail.com](mailto:riccardowanke@gmail.com)

**By way of a multidisciplinary approach, this article advances the idea that our listening to certain practices of contemporary art music (electroacoustic, classical contemporary, and electronic music) relies on precise connections to the early stage of perception. These styles of music are characterised by essential sound configurations that evolve in time, thus eliciting a sensorial impact which transcends features regarding sound sources and affective responses. Listeners grasp what Scruton calls ‘pure events’ in a ‘world of sound’, being able to distinguish, separate and sort acoustic stimuli. The article establishes a key parallel among seminal works of Bregman, McAdams, Kubovy, Bayle and other authors, highlighting a fundamental agreement of perceptual studies in psychology, neurophysiology and musicology for the understanding of the early stage of sound perception. Music practices typical of this perspective develop certain sound configurations, such as figure/ground arrangements, recurrent elements and morphological distinction, that closely mirror our innate mechanisms of prediction in perception. A parallel is made between studies in the philosophy of perception and the neurophysiology which allows us to postulate the idea that these styles of music are essentially based on pure temporal proto-objects.**

## 1. INTRODUCTION

In this article, we aim to discuss the early stage of perception in relation to contemporary art music practices.<sup>1</sup> In particular, we examine how temporality of certain styles of music is closely intertwined with the corresponding sound configurations and with predictability, which is taken as a key principle of sound perception, and hence of time perception.

We believe that the study of the fundamental stages of perception could be useful in understanding how to approach today’s experimental music, and in particular those practices that make use of a large sonic palette within non-linear structures (e.g., electroacoustic music, post-spectralism, glitch electronica). We argue that the study of the early stage of perception in relation to these

<sup>1</sup>We use the term ‘contemporary art music’ to refer to a large set of current genres and aesthetics, both institutional and underground, that are generally included within current experimental music practices. This is an impartial term that takes the place of other common terms such as ‘new music’, ‘contemporary and experimental music’ or ‘avant-garde music’ (Grebosz-Haring and Weichbold 2020).

kinds of music practices should take into consideration the complete set of characteristics elicited in our auditory system. The auditory system processes stimuli focus, in a first phase, on frequency and pitch distinction (Hyde, Peretz and Zatorre 2008), spatio-temporal indications (Schönwiesner and Zatorre 2009) and music/speech recognition (Peretz, Vuvan, Lagrois and Armony 2015).<sup>2</sup> The focus on the physical characteristics of sound is a key feature of many genres of contemporary and electroacoustic music practices, and the perception of this type of music may operate differently in so far as it confronts the listener with a wide range of sound types as against the more restricted set of conventional musical sounds. Extended temporal forms and frequency ranges, aural stimuli at the threshold of our perception, or real-world sounds: all these elements stretch our perceptual system, thus affecting how we apprehend this music – unlike more conventional forms (e.g., stimulating unexpected regions of our auditory cortex).

We are conscious that many of these features appear to be central aspects of how we perceive music in general (e.g., long held pedal points and staccato; harmonic and melodic figures). However, because of the specific aural characteristics of certain styles within electroacoustic music, it is reasonable to suppose that, when faced with a various range of contemporary art music practices, our perception is engaged with a series of acoustic stimuli that involve a complex set of mechanisms over and above those of conventional music listening, and covering the entire process of sound perception. Our ability to make the primary distinction between high and low frequency, and transient and continuous sounds, is such a fundamental aspect of

<sup>2</sup>We refer in particular to the work of Norman-Haignere, Kanwisher and McDermott (2015), which describes six different regions of our auditory cortex reacting to decisive aspects of acoustic information. It appears that two primal regions of our auditory perception are based on a specific sensitivity to low and high frequencies of sound. Second, regions three and four respond to particular aspects of the spectro-temporality of sound. Finally, specific regions in distinct non-primary areas are dedicated to speech (foreign and native languages) and music (vocal and absolute music), thus indicating that the auditory cortex has clear anatomical pathways for the reception of music and language.

how we parse sound stimuli that needs to be given a primary role in approaching contemporary art music.

Through an exploration of perceptual aspects and the temporal features of this music, we aim to find the key features that may elucidate the mechanisms of the appreciation of it.

First, analysing different ways of listening, we suggest that, apart from emotional and cultural perspectives, a peculiar characteristic of these types of music is to exhibit a series of temporal *sound configurations* deployed in our perception. A cross-disciplinary overview of studies in music and sound perception confirms that psychologists, neuroscientists and music theorists agree on the principal pathways of our early stage perception, and their content, claiming both functional and anatomical validity.

Second, we give a series of examples taken from a large variety of today's music practices that indicate how primal sound configurations are the essential elements of contemporary art music, and are constituted of primitive acoustic articulations arranged in a domain of frequency against time.

Finally, we correlate these musical analyses with a set of studies in the philosophy of perception and neuroscience, demonstrating that these (apparently) different perspectives converge into a unique definition of a *pure temporal* domain of perception. This definition matches considerations of sensorial, musical and neurological aspects of perception in relation to certain practices of contemporary art music, confirming that these styles possess specific temporal traits that call for a different approach from the one found with traditional styles of music. This article sets out a multidisciplinary framework that allows us to postulate a novel description of the first stages of the perceptual experience of these styles of music.

## 2. VARIETIES OF LISTENING

Usually, in contemporary art music there is a lack of consensual syntax, and listeners end up mostly talking in terms of technical aspects (as often occurs with electroacoustic multichannel performances), genre-based features (faced with new pieces, we tend to refer to similar authors or stylistic references) or emotional impact (Dibben 2001).

In our previous work (Wanke 2019), a listening survey over a range of pieces of contemporary music (post-spectralism, mixed-source ensembles and glitch electronica) supports the idea that some contemporary music practices favour our sensorial approach to the most basic and universal aural forms. This survey – administered as a sorting task – reveals that listeners tend to approach this kind of music with three varieties of listening: (1) an *analytic* listening that allows the

listener to identify contextual elements (i.e., styles and instrumentation); (2) an *affective* listening which encompasses a second set of sorting criteria based on emotional associations resulting from aural stimuli; and (3) a third mode, called *immanent* listening, which consists in a sort of music description through an act of direct apprehension of musical forms and shapes, sound morphologies, and evolutions that develop over time. This third type of listening seems to honour a non-physicalist account of sound such as Roger Scruton's, according to which 'the acousmatic experience of sound is precisely what is exploited by the art of music' (Scruton 1997: 3). When we hear sounds as music, Scruton claims, we (can) hear them as *secondary objects* and *pure events* detached 'from the circumstances of [their] production' (ibid.). From a philosophical standpoint, then, sounds are 'objects in their own right, bearers of properties ... identifiable separately both from the things that emit them and from the places they are located ... things that happen but don't happen to anything' (Scruton 2009: 50).

As we will clarify, in the very first stages of perception, primitive sound segregation is a grouping process based on frequency, time and intensity rather than on a direct source recognition (Bregman 1990). Different tones of the violin, for instance, are sorted together by our auditory system not because of their common origin, but because of their spectro-temporal similarities. In this sense, listening to music and being receptive to its aesthetically significant features requires us not to listen to violins, horns or brushes on snare drums, but rather to hear sounds and grasp them regardless of their common sources. *Immanent* listening, in particular, focuses on the immanent aspect of sound, using natural schemes of recognition and formal categories that are in some way general. Listeners draw from the natural tendency to organise sensorial data in time, shapes ('punctual pulsations'), temporal evolution ('pulses periodically repeated') and inner spatial dimension ('in a figure/background relationship'). These elements are actually time-spans, or 'temporal *Gestalt*-units' – as they are defined by James Tenney – 'whose *perceptual* boundaries are largely determined by the nature of the sounds and sound-configurations' (Tenney and Polansky 1980: 205).<sup>3</sup>

Referring to musical examples derived mainly from the traditional classical repertoire, Eric Clarke

<sup>3</sup>A parallel with the neo-Gestaltic approach of Michael Kubovy and David van Valkenburg (2001) can be drawn: a perceptual object is what stands out as such on a background, in compliance with the *Gestalt* principle of the figure/ground segregation. A perceptual object is anything that can undergo the phenomenon of figure/ground segregation. Obviously, modes of figure/ground segregation vary in the case of an auditory object compared to those of an object that we claim to see: 'auditory objects are formed in pitch-time, whereas visual objects are formed in space-time' (2001: 106).

evidences that '[w]hen perception proceeds in an unproblematic way, we are usually unaware of the sensory aspect of the stimulus information, and are only attuned to the events that are specified by stimulus structure. But when that relationship is problematic, the stimulus structure itself can become more evident' (Clarke 2005: 32).

Here, we argue that it is precisely this kind of 'problematic perception' of today's experimental music that allows some works in the genre to trigger the 'stimulus structure': the music of authors such as Georg Friedrich Haas, Pierre Alexandre Tremblay or Pan Sonic is based precisely on unconventional sound configurations that exploit sonic effects embedded within musical episodes, eliciting a sensorial impact (cf. Kendall's *phenomenal qualities* (Kendall 2014: 200)). This argument allows us to hypothesise that our listening to this type of music extracts – at a primal stage – a set of frames in our perception that operate as time-spans, prototypical examples of time perception: primal sonic shapes that evolve in along axes of frequency against time. These styles of music, perhaps more than others, focus their effects on the early stage of perception, where primitive segregation processes begin.

### 3. THE EARLY STAGE OF SOUND PERCEPTION: A CROSS-DISCIPLINARY PERSPECTIVE

The perceptual aspect of sound has been the subject of an immense range of studies in the cognitive sciences, from psychology (Susini, Lemaitre and McAdams 2011) to neuroscience (Blood, Zatorre, Bermudez and Evans 1999; Griffiths and Warren 2004; for a review, see Santarcangelo 2015 and Koelsch and Siebel 2005).<sup>4</sup> These studies span human reactions to simple stimuli all the way up to complex feelings and emotions (Juslin 2013). Acoustics, psychoacoustics and neurophysiologists, as well as philosophers, musicologists and composers, have discussed the mechanism of perception for a given aural stimulus. As Figure 1 shows, some core fundamentals of sound perception – in particular, the temporal succession of events – are shared among the different scientific domains.

Stephen McAdams, Tim Griffiths and their colleagues (Griffiths and Warren 2004; Egermann and McAdams 2013) isolate a first step of *encoding* or *transduction* in which our system reacts within a psychophysical dimension and operates a primal grouping of sounds identified by Albert Bregman as *primitive segregation* (1990), and by Michael Kubovy as *early perceptual stage* (Kubovy and van Valkenburg 2001). This is the intuitive phase of

perceiving with the ear (*ouïr*: Bayle 1993: 93–109). Once the sonic message is segregated, the next steps correspond – following François Bayle's labels – to the processes of listening (*écouter*) and hearing (*entendre*).<sup>5</sup> At this stage, the attention leads to a figure/ground segregation (Kubovy and van Valkenburg 2001), thus allowing the activation of *abstract knowledge structures* (McAdams 1996) and the emergence of contextual expectations (Konečni 2008; Kendall 2014). This is the perceptual stage at which induced emotive mechanisms are activated, as noted by some psychoacousticians (Juslin, Liljeström, Västfjäll and Lundqvist 2010; Egermann and McAdams 2013).

These studies show that the features on which our auditory system bases its sensorial processes has to do with both material aspects (e.g., temporal and frequency-based) and other more complex criteria, such as melodic progression, timbre or articulation, which rely on cultural and environmental traits. What is peculiar to the majority of these studies is that they either consider a set of many kinds of short, simple aural stimuli, or draw on a musical database of works from Western classical, World and popular music repertoires.

We shall attempt to discern how our neurological and emotive response works when music does not use tonal constructions, traditional narrative structures or linear time perception – instead being composed, of noises, real-world sounds, acoustic and electronic sounds, or sounds close to the threshold of perception, all potentially within non-linear temporal structures. How does musical temporality function for this type of music?

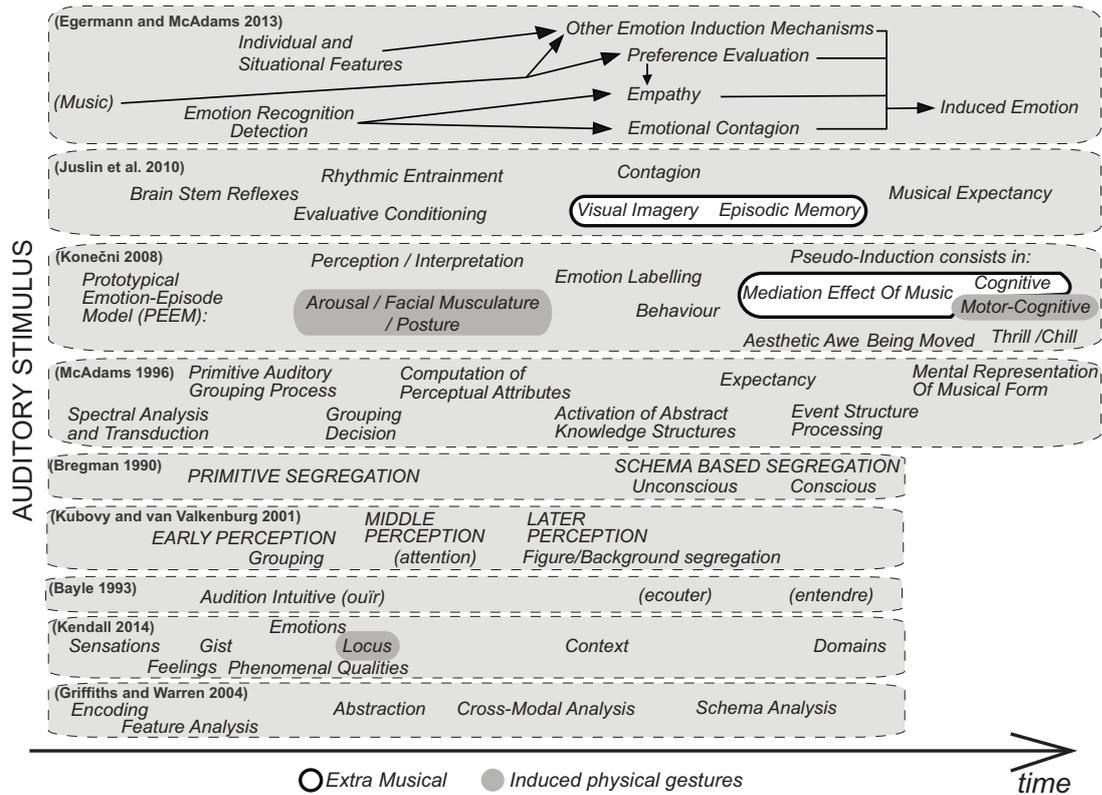
### 4. SOME EXAMPLES

Consider, for instance, compositions such Alvin Lucier's *Two Circles*, or *Solo IV* for double-bass by Jacob Ullman, or even various pieces by Ryoji Ikeda, Salvatore Sciarrino, Luc Ferrari or Richard Chartier: these works spill across the limits of our perception. They exceed the form of a musical (or sound) object. They do not possess clear edges or boundaries (being potentially confused within environmental sound or creating acoustic challenges in perception). They puzzle our perceptual segregation by shifting the intensity of morphologically similar aural stimuli or by mixing different sounds into a coherent *perceptual chain* (see section 5), challenging our expectation of future musical events.

Several pieces of electroacoustic music present musical syntaxes that figure with specific temporal evolutions, often structured as interlaced events in the spectral domain as well as in spatial and temporal

<sup>4</sup>For an updated review of the philosophical literature, see Casati, Dokic and Di Bona (2020).

<sup>5</sup>We follow John Dack's translation of the Schaefferian terms (Chion [1983] 2009).



**Figure 1.** Schematic representation of the primary process of perception (horizontal temporal line) for different studies. This parallel aims to synchronise perceptual and cognitive steps from different perspective, nevertheless it is common among psychologists (‘emotivists’) to examine coexisting cognitive states while neuroscientists use to develop sequential frameworks. Since our main focus is not addressed to emotional response to acoustic stimuli, we did not indicate the vast neuroscientific literature examining the role of the amygdala in the emotional experience of music.

dimensions. For instance, a 360° multichannel piece could construct a virtual scenery in which each sound, by reverberating with a specific decay and character, delineates the geometry and the dimensions of that suggested space. In this case, the listener is immersed in a simulated reality with a corresponding time frame, and predictability goes together with a coherent set of primal segregation of sound events. In particular, perhaps more than other compositions, pieces that make rich use of references and signifiers to sounds in the real world involve the activation of those perceptual areas dedicated to real-world sounds (i.e., the main four regions of our auditory system that respond to pitch and temporal cues). These sounds activate a richer cognitive response compared with non-referential sounds, but at an early perceptual stage, their temporal segregation and predictability function in a similar way, especially when they show recurrent motives (e.g., sound of waves, birds) or continuous and metric streams (e.g., industrial sounds, waterfalls).

When, on the other hand, the temporal simulation – made either by artificial or natural sonic features – is disturbed by sounds of a different nature, possessing, for instance, different delay times, the perceiver is

puzzled and moved towards different temporal dimensions. In *Etude Élastique* of *De Natura Sonorum* by Bernard Parmegiani, listeners are immersed in a multidimensional sonic environment in which temporality appears to progress at different rates. When the succession of different sonic events is very fast, the listener’s immersion within a coherent sonic space is challenging, with perception of the temporal flow becoming more confused and the segregation of sonic stimuli being more uncertain. The listener is then not immersed in a single aural landscape, and the succession of finite events, placed in combination within a non-coherent simulation, usually pushes the listener to experience them in chronometric time: objects are arranged in mutual relations and they are perceived distinctly, finitely, pushing the listener out of an immersive virtual reconstruction.

Other types of temporality exist in those works that make use of recurrent structures to create hypnotic reiterations (e.g., *in vain* by Haas), slow evolutions (e.g., *Lichtbogen* by Kaija Saariaho), periodic profiles (e.g., *DW8* by Bernhard Lang) and continuous shapes (e.g., *Stria* by John Chowning). Recurrent structures play a crucial role in predictability – as

we shall see in section 5 – and these temporalities approximate to Jonathan Kramer’s definition of *vertical* time (Kramer 1988). According to Kramer, vertical temporality is characterised by the lack of music directionality and a consistency of expectation, but it is worth noticing that parallels between, for instance, the lack of teleology and the absence of organisation and hierarchy does not always hold. Many works may contain long musical passages that are substantially undifferentiated and continuous, but their development follows a strong dramaturgy and creates musical forms that are ordered and organised, thus opening up to a more structured predictability. Giacinto Scelsi’s *Quattro Pezzi*, for example, shows a sense of dramatic construction (e.g., using the ‘tension/resolution’ schema) typical of the Western musical tradition, together with unconventional sonic metres commonly found in Eastern practices (Menesson 2008). In such pieces, there is a sort of expectation and structural hierarchy: the musical time is not linear but, nevertheless, a progression of sound flow is discernible. There is a narrative (Andean 2016), but not a teleological profile (i.e., directionality): the listener is absorbed (or immersed) within a sonic discourse.

The central hypothesis we propose here is that a part of contemporary art music has developed starting from a set of primal sound configurations. The very first encounter with this music reproduces implicit forms shaped by temporal boundaries (Kubovy and van Valkenburg 2001) as, for instance:

- figure/ground segregation;
- complex multiple plane arrangements;
- clashes between pulses vs sustained stimuli;
- time-based frames eliciting causal chains in perception (i.e., tension/resolution);
- unfolding of contrasting profiles (e.g., ascending/descending).

Electroacoustic pieces frequently display simultaneous but morphologically different sounds. These events are sorted in various ways based on simple rules which are dependent on time and frequency (Bregman 1990) and on what comes before. As we shall see shortly, our capacity to store in memory single or multiple acoustic stimuli may in fact favour the activation of one of these sounds, thus enabling a specific kind of segregation (e.g., pitch and timbre features trigger distinct areas in memory (Jeong and Hokyoung 2016)). These concurrent sonic events can be then segregated (1) as figure/ground, if one has similar characteristics to previous sounds or shows an extended form (i.e., based on ADSR parameters); (2) as a unique episode, if their spectral morphologies have several characteristics in common; or (3) as distinct sounds happening at the same time, if they all appear as ‘new’ (i.e., when no elements stored in memory help

a particular recognition), they differ enough in frequency and ADSR features, but not too much to be considered within a figure/ground setting. These simultaneous sounds may be seen as an atemporal (and spatial) instance within musical episodes that put less emphasis on sound articulation in time. They are experienced as isolated events and are processed as ‘single points’ within linear time (Kramer 1988) by our auditory system (Bregman 1990).

A piece such as *Dulle Griet* (2009–10) by Giovanni Verrando exemplifies how a complex composition may be perceived, at a primal level, as an interlaced combination of sound configurations. Verrando (2013) constructs a complex morphology-based interplay between harmonic and inharmonic musical configurations and between acoustic and electric sounds (Figure 2). Constituted by ascending and descending lines, this ever-changing sonic texture challenges predictability, but a succession of the similar sonic aggregates works as a backbone for a sort of morphological recognition (see Figure 2 spectral areas labelled as I, II, I’, II’). This arrangement, as we shall see, has an important effect on expectancy and sonic identification.

Similarly, the periodic (and specular) movements of pitch-based profiles in *Lautturi* by Pan Sonic or the intricate descending *glissando* traced by Haas in his String Quartet n°2 are what we call sound configurations (Figure 3). The arrangement of *Lautturi*, in its second part (from 02’25” (Pan Sonic 2004)), is perceived both as a steady sound made of a microtonal cluster of sounds and as an undulating profile that incorporates brilliance and ringing components (typical of sounds above 10 kHz, Figure 3, top left inserts). These high-pitched *glissandi* are characterised by symmetrical movements that have no natural harmonic correlation (a sound that goes from G8 and C#9 corresponds to a specular one from A#9 to C#10: Figure 3 top right insert). In his String Quartet n°2, Haas constructs an episode using repetitive *glissandi* played with the same dynamic structure (i.e., a middle note of the *glissando* is played *sforzato*) to build a larger configuration. The sonic cascade generated from Eb6 to G#3 is a movement that starts vertically (steep *glissandi*) and ends horizontally (with the use of sustained tones moving within a microtonal range of frequencies) and in which the interwoven construction of the four instruments describes a continual descent and the intensity of middle notes allows the perception of a linear slope (Figure 3 bottom).

Haas explores this idea also in *in vain*, in which a steep descent evolves into a flat and extended sound. In this case, the author composes these musical shapes at different scales: he uses simple musical elements to create new developments, creating a reiteration of descending notes to evoke a sort of large-scale falling

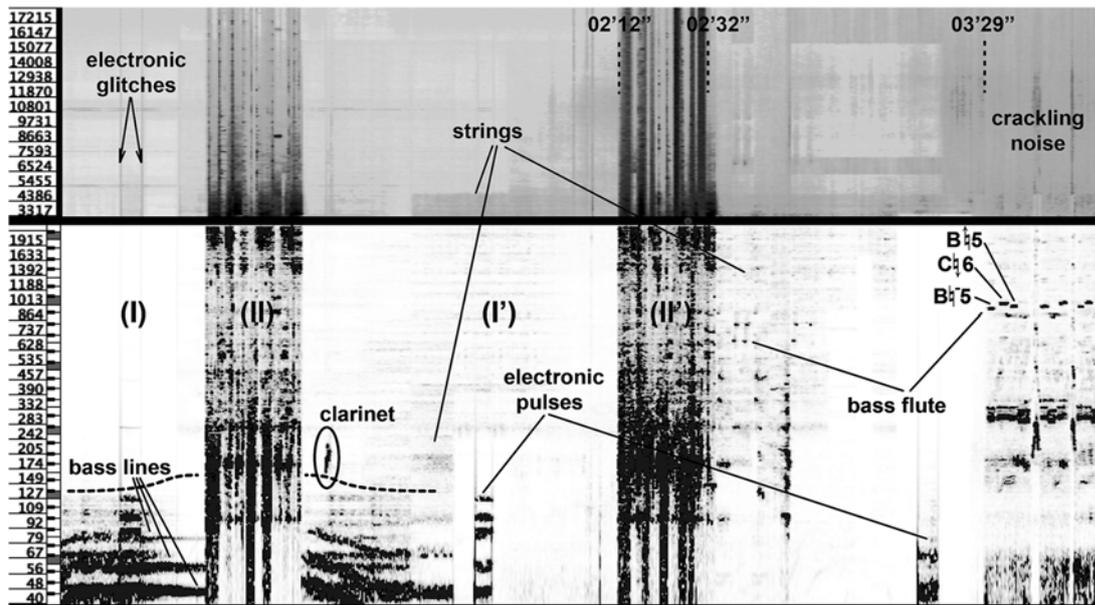


Figure 2. Spectrograms (top: 3.3–17.5 kHz; bottom: 40–2000 Hz / 00'00"–03'54") of *Dulle Griet* by Verrando (2013).

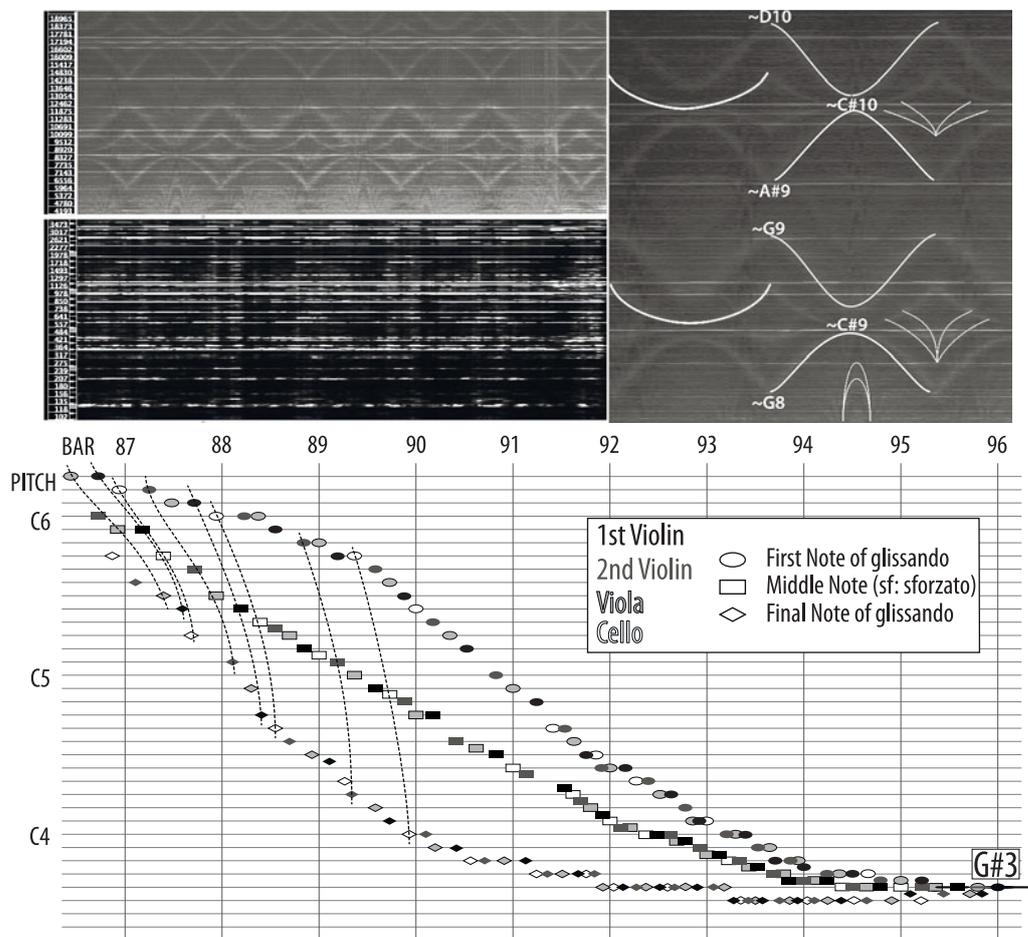


Figure 3. Top: spectrograms (left insert: bottom 102–3500 Hz; top 4100–19 kHz / 02'55"–03'55"). Right insert: 5000–20 k/02'41"–03'00") of *Lautturi* by Pan Sonic (2004). Bottom: pitch-based representation of the interlaced *glissandi* of bars 87–96 of *String Quartet n°2* by Haas.



Figure 4. Spectrogram (30–5700 Hz / bars 343–413) of *in vain* by Haas.

spiral that constantly ascends on the small-scale level. This main figure which opens and closes the piece (the opening figure being mirrored at the end), and which appears in a central episode, is a cyclical progression continuously changing in timbre, moving between percussion and piano to strings, giving a monumental global development. The descending lines which begin the figure are woven together in a dense sequence of notes and their fast repetition generates a sort of sustained and oscillating effect of suspended tritones. With each repetition, the lines converge on low pitches, mixing together into an indistinct mass of sound as the new ascending profile emerges. The initial sequence progressively slows down to a dilated succession of sustained notes.

This construction, when observed on a macro-scale, recalls the never-ending development of some acoustic effects such as the Shepard tone, which in Bregman (1990) are conceived as perceptual entities (*auditory streams*, in his terminology). Starting from a fast succession of short tones within a steep descending configuration, Haas progressively dilates the entry points of new notes and their durations, modifying the sequential overlapping of every descending figure (Figure 4, dashed arrows). He creates a pulsating temporal identity without using any percussive elements, composing a flexible sonic matter that is pulled out and extended.<sup>6</sup>

These energy-forms take shape in time and have the potential to elicit their profile as linear (i.e., pitch-based), textural (i.e., spectral-based) or consisting of a more complex combination of elements.

<sup>6</sup>It is worth emphasising that *in vain* contains lighting instructions for live performance, denoting shifts from darkness to full illumination, driving the audience's attention towards the corresponding perceptual aspects of sound. The processing of auditory stimuli in complete darkness differs from conventional acoustical listening: the lack of any spatial dimension affects the neurological response of our multimodal apprehension of the external world (Teder-Sälejärvi, Di Russo, McDonald and Hillyard 2005) and affects the localisation of sounds (Doucet, Guillemot, Lassonde, Gagné, Leclerc and Lepore 2005).

Like *in vain*, *Vortex Temporum* by Gérard Grisey is a complex case of multiple music temporalities (Hervé 2004). Musical time, and consequently the listener, are brought along at different rates: the succession of episodes with diverse profiles (e.g., circular, flat, accelerating), each one composed of repeated motifs, opens up to different temporalities and therefore to different perceptual dimensions. The basis of Grisey's research, in particular, is Henri Bergson's notion of duration (*durée*), according to which the music is an expression of time and the musical material is a manifestation of pure duration (*son-durée*): 'this principle of instantaneous generation, this matryoshka effect where sounds themselves generate one another at every given instant seems to me freighted with consequences . . . . The composer operates on the whole duration at once, at every possible instant' (Grisey [1982] 2008: 52).<sup>7</sup>

The shift from a real-time perception of music towards a post-experiential condition is characterised by the constant link between external and internal temporalities. Sound configurations unfold in music and, as we have seen, are efficaciously represented within spectrograms; even so they are perceived and processed by our auditory system. This is in constant movement, and in Grisey's music the innermost time is realised through a distortion of a sound's perspective in memory (Hervé 2004). Memory, 'through to its retroactivity', becomes the way in which we 'rebuild the trace described in time by what lasts' (Imberty 2005: 53).<sup>8</sup> Memory is the locus of the 'mental' resonance of sound: a sort of re-exposition of the trace of sound (Bonnet 2012: 29) allowing us to perceive this

<sup>7</sup>Ce principe de génération instantanée, cet effet de matriochka par lequel les sons s'engendrent les uns des autres à chaque instant donné me semble lourd de conséquences . . . . Le compositeur opère sur toute la durée à la fois, dans tous les instants possibles' (our translation).

<sup>8</sup>[Q]ui, par sa rétroactivité, reconstruit la trace décrite dans le temps par ce qui dure' (our translation).

continuity as ‘movement that *is not associated to a cause*’ but consists ‘in a transformation where nothing changes’ (Bergson [1934] 1960, 164, italics ours).<sup>9</sup>

Memory mechanisms, as neuroscientists also confirm, ‘operate over a range of timescales and auditory processing stages, ranging from sensory memory traces in early auditory cortical areas ... to entry into long-term storage in higher-order areas’ (Griffiths and Warren 2004: 890–1). While ‘short-term memory’ generally refers to the simple (and passive) temporary storage of information, ‘working memory’ is intended to denote a specific type of short-term memory that consists in the maintenance and manipulation of information, taking on an active role and being characteristic of a mutual functioning between storage and processing (Baddeley 2012; Schulze and Koelsch 2012). Moreover, the manifest independent processing of spatial and non-spatial properties of acoustic stimuli in working memory (Alain, Arnott, Hevenor, Graham and Grady 2001; Furukawa, Washizawa, Ochi and Kashino 2013) is evidence that the identification of sound (pitch, spectral and temporal issues) and spatial location (based on interaural time differences) are retained and processed in separate streams.<sup>10</sup> At this primal level, then, temporal features appear to be detached from spatial ones and to be – together with frequency-based information – the essential dimension of sound configurations present in music and emergent in a sort of (un)stored domain of our memory.

## 5. THE (UN)STORAGED DOMAIN OF EARLY STAGE PERCEPTION

These foregoing musical examples illustrate the immense set of temporal possibilities in perception, and emphasise, in particular, how a number of styles within the experimental music scene (i.e., electro-acoustic, contemporary instrumental and electronic music) are characterised by a set of elemental aural forms that we have called ‘sound configurations’.

At the early stage of perception, the listener apprehends not just a sonic material, but also a potential physical energy form out of a spatial domain. These sound configurations are placed in a matrix of frequency against time, varying in dynamic, that is a pure temporal field where sound manifests itself as pure duration (following Bergson’s definition), or as

<sup>9</sup>[M]ouvement qui n’est pas attaché à un mobile, d’un changement sans rien qui change’ (our translation).

<sup>10</sup>It is still debated whether the component of the ‘Baddeley working memory model’ – the *phonological loop* – is specialised in processing only language or also other meaningful sounds, in particular music. According to a review by Schulze and Koelsch (2012), the *phonological loop* can be involved in music processing or even incorporate subsystems specialised in processing non-phonological information (i.e., music).

pure events (Scruton 1997, 2009), and to which Grisey himself alludes: ‘[the art of music] gives us to be perceived what Proust called “a little time in a pure state”’ (Grisey 1987: 269).

An alternative denomination for temporal field is, following Scruton (1997, 2009), ‘world of sound’: here, streaming ‘involves attributing to sounds an identity distinct from any process in their source, and involves the creation of a world of coherent sounds, rather than a world of coherent spatio-temporal objects’ (Scruton 2009: 58).

The ‘world of sound’ described by Scruton shares some similarities with the ‘no-space world’ depicted in a Peter Frederick Strawson’s thought experiment (Strawson 1959). In Chapter two of *Individuals*, entitled ‘Sounds’, Strawson invites the reader to think of specific kinds of individuals (understood as objective particulars) as having only temporal parts and features in spite of their being dependent on more basic entities that have also spatial features. Could such individuals exist entirely independently of any commitment to space? If we consider a subject whose sensory faculties are far more limited than ours, the spatial representations of the world accessible to his mind would be truncated, ‘and this would have ramifications for the general conception of the spatial world available to a mind with such meager sensory inputs’ (McGinn 1999: 44).

The ‘no-space world’, as Strawson conceives it, is a purely auditory world: its denizens are sounds (Santarcangelo and Terrone 2015). Strawson conceives of the ‘no-space world’ as a merely auditory world, which has a temporal but no spatial dimensions, treating auditory individuals as a paradigmatic case of entities that may appear to be merely temporal. The question is: can there be temporal entities or is it rather the case that, even when an entity seems to be merely temporal, there is some underlying spatial dimension that has to be made explicit?<sup>11</sup>

Even if an ontological commitment to higher-order level entities existing only in time (such as sound configurations) entails an ontological commitment to more basic entities having also a spatial, or at least spatial-like dimension, there can be a realistic account of purely temporal individuals. Since they seem to have primarily temporal but no spatial dimension, sounds become a paradigmatic case of entities that may appear to be purely temporal. If, as seems to be in the case in the ‘no-space world’, the only available *dimension* is time, then both a pure hearer’s experience and all the sounds at a certain moment will

<sup>11</sup>It is worth noting here that sounds are just the means to the end of investigating an ontological claim about space and time: Strawson is testing the metaphysical claim according to which the notion of space is necessary for any conception of objective reality. The ‘no-space world’ becomes nothing but ‘a device to help us to explore some of the relations between some of the fundamental concepts of our scheme of things’ (Strawson 1980: 282).

occur in one and the same *location*, namely the present time, and therefore there seems to be no room for an off-experience existence. In our spatiotemporal world, such a housing function is fulfilled by space. What could fulfil this function in a no-space world?

The real metaphysical requirement for the objective reality of sounds in the no-space world is their permanence over time, their ‘continued existence’, independently of their being experienced and therefore identified or re-identified. The key question of the thought experiment could thus be rephrased as follows: can we conceive of an objective time, distinct from subjective time, without postulating an objective space? How can objective time be objective without a spatial dimension in which to house unperceived entities? Where do sounds go when they exit the reach of experience?

It is a similar series of questions to what we repeatedly find in the neuroscientific literature about sound perceptions and auditory objects (Griffiths and Warren 2004; for a review, see Santarcangelo 2015; Di Bona and Santarcangelo 2018): when confronted with a mixture of sounds, the auditory system first seeks for continuations of previous sounds and removes these from the mixture, and then analyses the residue. But, being the mixture of simultaneously overlapping sounds, where does the residue of removed stimuli (the unperceived sounds) go when it exits the reach of experience?

A large number of authors propose that two key processing strategies occur in auditory perception: first, perceptual representations of incoming stimuli should be predictive (Friston 2005; Summerfield and Egner 2009), and second, perceptual decisions about what to listen to next should be flexible. It is likely that a specific purpose of our attention is to store each single sound separately when a new sound is coming to the system: ‘our auditory system applies all possible grouping heuristics to determine whether the new sound can be fitted to the stored representations of the preceding sounds’ (Denham and Winkler 2015). This method allows us to benefit from the greatest possible flexibility. ‘It includes a full reinterpretation of all (stored) data with each new sound event. The main disadvantages are high pre-processing and decision-making costs. That is, on the arrival of each new sound, all processing occurs as if we had no knowledge of the acoustic configuration of our surroundings’ (Winkler, Böhm, Mill, Bendixen and Denham 2012: 1002).

Furthermore, although it has been demonstrated that a large amount of detailed auditory information becomes no longer available after a very short period of time, some authors show that the same information is maintained in some way – to the extent that a single ‘reminder’ sound re-activates them, even after a long period of inactivity, as in the case of several musical episodes discussed above. Even when experimental

subjects were instructed to remember individual sounds, they seemed to be unable to separate them completely and quickly from the sounds that had preceded them (Cowan 2001). It may be that the memory tasks at the basis of our ability to identify sounds store them as part of longer perceptual chains. These sort of perceptual chains, as we have seen previously, typify a large number of electroacoustic pieces, which consist in complex sound aggregates, that is, families of morphologically similar sounds.

These perceptual chains can be considered to be a kind of representation of auditory objects. Storing these longer sound-chains, such as the repetitive *glissandi* found in Grisey’s and Haas’s pieces, gives us greater benefit than the ability to remember only individual sound events as separated from all the others: ‘besides reducing processing requirements and increasing perceptual stability, such a system can continuously produce predictions of future states of the acoustic environment, thus helping one to prepare for upcoming events without having to commit higher level cognitive processing resources, whose capacity is possibly more limited’ (Winkler et al. 2012: 1002).

According to this theory, the senses are intrinsically predictive: in auditory perception, to predict means to recognise what binds a series of discrete sounds and extrapolate information regarding these links for future perceptions.

In the classical auditory streaming example of a galloping pattern (van Noorden 1975), we are able to coherently and automatically group together sounds with other corresponding ones. Given a sound perceived at  $t$ , the sound one thinks is coming at, say,  $t'$ , is not a fully formed auditory object, but, as Winkler et al. (2012) propose to call it, a ‘proto-object’ – an ‘alternative grouping describing the sound sequence’. For every prediction, several proto-objects enter into competition for perceptual dominance. Just one of the competing proto-objects – one of the different solutions it hypothesises – becomes an auditory object:

evidence from numerous electrophysiological studies suggests that these types of predictive relations can be extracted from sound sequences outside the focus of attention and that the extracted information is turned into predictions about forthcoming stimuli. Two types of information are necessary for forming these predictive representations: sensory data (i.e., the actual features of the individual sounds) and link data (i.e., the relations between the individual sounds). (Winkler et al. 2012: 1002)

The system, thus, uses object representations in order to predict those that follow, but its *decisions* about which of the competing proto-objects will become the given auditory object are endless and continuous. These decisions are not based on entirely trustworthy knowledge of future sound events. Indeed, we cannot know at any given moment if the auditory object we

are now perceiving and have perceived so far will continue as it is. A new sound event may arrive and join the current sound pattern, or it may be the beginning of a new one: '[w]e suggest that it is in building and maintaining multiple alternative groupings in parallel that the auditory system manages to alleviate this problem, thus ensuring flexibility in perceptual decisions' (Winkler et al. 2012: 1003). When a new sound arrives, the system tries to link it to one of the representations of auditory objects it has already stored. Since these representations are inherently predictive, each incoming sound provides, by its mere presence, a validation test for these representations.

To summarise, the essential frames of our perceptual capacity are then grounded on a time-based extent where acoustic stimuli are processed in frequency. Together with the intensity dimension, this frequency-against-time extent appears as sort of *perceptual spectrogram*. It has been proposed, in fact, that since the early stage of auditory processing, 'the cochlea constructs sound images with dimensions to time and frequency domains' (Griffiths and Warren 2004: 888; Shamma 2001). Descending glissandi, periodic oscillations, metric and recurrent figures, layered continuous tones which vary in overtones are typically sound configurations which are codified by our primal sub-areas. These respond, as said, to pitch and spectro-temporal qualities of sound, being selective for both broadband sonic events that are rapidly changing over time, that is, discrete episodes, as well as all sonic features extended over time, such as continuous sounds with spectral variations, that is, harmonic and overtone modifications (Norman-Haignere et al. 2015). Thus, sound configurations pass, transfigured, from the external acoustic realm towards an inner storage in memory.

## 6. CONCLUSIONS

In considering the early stage of perception of certain practices of today's experimental music, we have touched upon the central cores of different disciplines. The parallels across neurophysiology, psychology, musicology and philosophy of perception point to the presence of a perceptual domain that is purely temporal, a domain in which sounds are experienced at a single instant and grasped in time, backward in memory (sound storage) and forward in predictability.

We have shown the isomorphism between the spectral features visible as temporal geometries of the musical pieces and the two-dimensional (frequency against time) *signature* produced in the early stage of perceptual processing.

Certain electroacoustic musical syntaxes are reified in listening as perceptual chains, grouped by 'old+new' heuristics and confirmed by a continuous validation process about future sound events. In particular, in this

article we have postulated that these styles of music are essentially based on pure temporal proto-objects insofar as they embody the way in which our early stages of perception process the auditory scene, opening up to a world of sound perceived in a pure temporal domain.

This outlook, able to link together processing stages of working memory, and the aesthetic appreciation of musical episodes, with our natural way to sort acoustic stimuli and the specific characteristics of certain styles of music, may favour the emergence of innovative possibilities in approaching electroacoustic music practices and contemporary art music in general.

## REFERENCES

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S. and Grady, C. L. 2001. 'What' and 'Where' in the Human Auditory System. *Proceedings of the National Academy of Sciences of the United States of America* **98**(21): 12301–6.
- Andean, J. 2016. Narrative Modes in Acousmatic Music. *Organised Sound* **21**(3): 192–203.
- Baddeley, A. 2012. Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology* **63**: 1–29.
- Bayle, F. 1993. *Musique Acousmatique: Propositions ... Positions*. Paris: INA-GRM/Buchet Chastel.
- Bergson, H. [1934] 1960. *La Pensée et Le Mouvant*, 35th edn. Paris: Presses Universitaires de France.
- Blood, A. J., Zatorre, R. J., Bermudez, P. and Evans, A. C. 1999. Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions. *Nature Neuroscience* **2**(4): 382–7.
- Bonnet, F. J. 2012. *Les Mots et Les Sons: Un Archipel Sonore*. Paris: Éditions de l'éclat.
- Bregman, A. S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.
- Casati, R., Dokic, J. and Di Bona, E. 2020. Sounds. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*. Stanford: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/sounds/>
- Chion, M. (1983) 2009. *Guide to Sound Objects. Pierre Schaeffer and Musical Research [Guide des Objets Sonores]*, trans. John Dack and Christine North. Paris: Buchet Chastel.
- Clarke, E. 2005. *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press.
- Cowan, N. 2001. The Magical Number 4 in Short-term Memory. A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences* **24**(1): 87–185.
- Denham, S. L. and Winkler, I. 2015. Auditory perceptual organization. In J. Wagemans (ed.) *Oxford Handbook of Perceptual Organization*. Oxford: Oxford University Press, 601–20.
- Dibben, N. 2001. What Do We Hear, When We Hear Music?: Music Perception and Musical Material. *Musicae Scientiae* **5**(2): 161–94.

- Di Bona, E. and Santarcangelo, V. 2018. *Il suono. L'esperienza uditiva e i suoi oggetti*. Milano: Raffaello Cortina Editore.
- Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagné, J.-P., Leclerc, C. and Lepore, F. 2005. Blind Subjects Process Auditory Spectral Cues More Efficiently Than Sighted Individuals. *Experimental Brain Research* **160**(2): 194–202.
- Egermann, H. and McAdams, S. 2013. Empathy and Emotional Contagion as a Link Between Recognized and Felt Emotions in Music Listening. *Music Perception: An Interdisciplinary Journal* **31**(2): 139–56.
- Friston, K. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1456): 815–36.
- Furukawa, S., Washizawa, S., Ochi, A. and Kashino, M. 2013. How Independent Are the Pitch and Interaural-time-difference Mechanisms That Rely on Temporal Fine Structure Information? In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon and H. E. Gockel (eds.) *Basic Aspects of Hearing: Physiology and Perception*. New York: Springer New York, 91–9.
- Grebosz-Haring, K. and Weichbold, M. 2020. Contemporary Art Music and Its Audiences: Age, Gender, and Social Class Profile. *Musicae Scientiae* **24**(1): 60–77.
- Griffiths, T. D. and Warren, J. D. 2004. What Is an Auditory Object? *Nature Reviews Neuroscience* **5**(11): 887–92.
- Grisey, G. 1987. Tempus ex Machina: A Composer's Reflections on Musical Time. *Contemporary Music Review* **2**(1): 239–75.
- Grisey, G. [1982] 2008. La Musique: Le Devenir Des Sons. In G. Lelong and A.-M. Réby (eds.) *Écrits: Ou l'invention de La Musique Spectrale*. Paris: MF Éditions, 45–56.
- Hervé, J.-L. 2004. Formes et Temporalité dans les dernières Oeuvres de Gérard Grisey. In D. Cohen-Levinas (ed.) *Le Temps de l'Écoute. Gérard Grisey ou la Beauté des Ombres Sonores*. Paris: L'Harmattan, 16–21.
- Hyde, K. L., Peretz, I. and Zatorre, R. J. 2008. Evidence for the Role of the Right Auditory Cortex in Fine Pitch Resolution. *Neuropsychologia* **46**: 632–9.
- Imberty, M. 2005. *La Musique Creuse Le Temps. De Wagner à Boulez: Musique, Psychologie, Psychanalyse*. Paris: L'Harmattan.
- Jeong, E. and Hokyung, R. 2016. Nonverbal auditory working memory: Can music indicate the capacity? *Brain and Cognition* **105**: 9–21.
- Juslin, P. N. 2013. From Everyday Emotions to Aesthetic Emotions: Towards a Unified Theory of Musical Emotions. *Physics of Life Reviews* **10**(3): 235–66.
- Juslin, P. N., Liljeström, S., Västfjäll, D. and Lundqvist, L.-O. 2010. How Does Music Evoke Emotions? Exploring the Underlying Mechanisms. In P. N. Juslin and J. A. Sloboda (eds.) *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press, 605–42.
- Kendall, G. S. 2014. The Feeling Blend: Feeling and Emotion in Electroacoustic Art. *Organised Sound* **19**: 192–202.
- Koelsch, S. and Siebel, W. A. 2005. Towards a Neural Basis of Music Perception. *TRENDS in Cognitive Sciences* **9**(12): 578–84.
- Konečni, V. 2008. Does Music Induce Emotion? A Theoretical and Methodological Analysis. *Psychology of Aesthetics, Creativity, and the Arts* **2**(2): 115–29.
- Kramer, J. D. 1988. *The Time of Music: New Meanings, New Temporalities, New Listening Strategies*. München: Schirmer/Mosel Verlag GmbH.
- Kubovy, M. and van Valkenburg, D. 2001. Auditory and Visual Objects. *Cognition* **80**: 97–126.
- McAdams, S. 1996. Audition: Cognitive Psychology of Music. In R. R. Llinás and P. S. Churchland (eds.) *The Mind-Brain Continuum: Sensory Processes*. Cambridge, MA: MIT Press, 251–79.
- McGinn, C. 1999. *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.
- Menesson, C. 2008. Scelsi, le Temps ou la Respiration du Son. In P. A. Castanet (ed.) *Giacinto Scelsi Aujourd'hui. Actes des journées européennes d'études musicales consacrées à Giacinto Scelsi (1905–1988)*. Paris: Publication Cdmc, 69–74.
- Norman-Haignere, S., Kanwisher, N. G. and McDermott, J. H. 2015. Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron* **88**(6): 1281–96.
- Peretz, I., Vuvan, D., Lagrois, M.-É. and Armony, J. L. 2015. Neural Overlap in Processing Music and Speech. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**(1664). [www.ncbi.nlm.nih.gov/pmc/articles/PMC4321131/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4321131/)
- Santarcangelo, V. 2015. Auditory Objects. A New Way to Define Old Things. A Philosophical Account. Unpublished doctoral dissertation, Università degli Studi di Torino, Turin.
- Santarcangelo, V. and Terrone, E. 2015. Sounds and Other Denizens of Time. *The Monist* **98**, 2: 168–80.
- Schönwiesner, M. and Zatorre, R. J. 2009. Spectro-Temporal Modulation Transfer Function of Single Voxels in the Human Auditory Cortex Measured with High-Resolution fMRI. *Proceedings of the National Academy of Sciences* **106**(34): 14611–16.
- Schulze, K. and Koelsch, S. 2012. Working Memory for Speech and Music. *Annals of the New York Academy of Sciences* **1252**: 229–36.
- Scruton, R. 1997. *The Aesthetics of Music*. Oxford: Clarendon Press.
- Scruton, R. 2009. Sounds as Secondary Objects and Pure Events. In M. Nudds and C. O'Callaghan (eds.) *Sounds and Perception: New Philosophical Essays*. Oxford: Oxford University Press, 50–68.
- Shamma, S. 2001. On the Role of Space and Time in Auditory Processing. *TRENDS in Cognitive Sciences* **5**(8): 340–8.
- Strawson, P. F. 1959. *Individuals An Essay in Descriptive Metaphysics*. London: Methuen.
- Strawson, P. F. 1980. Reply to Evans. In Z. Van Straaten (ed.) *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Clarendon Press, 273–82.

- Summerfield, C. and Egner, T. 2009. Expectation (and Attention) in Visual Cognition. *Trends in Cognitive Sciences*, **13**(9): 403–9.
- Susini, P., Lemaitre, G. and McAdams, S. 2011. Psychological Measurement for Sound Description and Evaluation. In B. Berglund, G. B. Rossi, J. T. Townsend and L. R. Pendrill (eds.) *Measurement Persons: Theory, Methods, and Implementation Areas*. New York: Psychology Press, 227–53.
- Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J. and Hillyard, S. A. 2005. Effects of Spatial Congruity on Audio-Visual Multimodal Integration. *Journal of Cognitive Neuroscience* **17**(9): 1396–409.
- Tenney, J. and Polansky, L. 1980. Temporal Gestalt Perception in Music. *Journal of Music Theory* **24**(2): 205–41.
- Van Noorden, L. P. A. S. 1975. Temporal Coherence in the Perception of Tone Sequences. Unpublished doctoral dissertation, University of Technology, Eindhoven.
- Wanke, R. 2019. A Unified Approach to Perceiving Sound in Current Experimental Music Genres. *Filigrane. Musique, Esthétique, Sciences, Société*, 23. <http://revues.mshparisnord.org/filigrane/index.php?id=879>
- Winkler, I., Böhm, T. M., Mill, R. W., Bendixen, A. and Denham, S. L. 2012. Multistability in Auditory Stream Segregation: A Predictive Coding View. *Philosophical Transactions of the Royal Society B* **367**(1591): 1001–12.

#### DISCOGRAPHY

- Pan Sonic (M. Vainio and I. Väisänen). 2004. *Lautturi on Kesto 234.48:4*. UK: Blast First, 180BX-CD.
- Verrando, G. 2013. *Dulle Griet (2009), Triptych#2 (2008) on Mdi Ensemble, RepertorioZero – Dulle Griet*. France: AEON, 1328-CD.