

Perception and cognition of music components

Elvira Di Bona

Abstract

In this paper, I analyze how the auditory system works when exposed to different kinds of music components, such as melodies, harmonies, the mere successions of sounds, and the ordinary simultaneities of sounds. I will claim that, while melodies and harmonies are perceivable, the mere successions of sounds and the ordinary simultaneities of sounds are processed at the cognitive level. This conclusion is based on the analysis of how the auditory system employs primitive grouping, schema-based grouping, and cognitive schema-based grouping when ordering the auditory landscape by the identification and categorization of meaningful musical streams.

Keywords: Melodies; Harmonies; Perception; Cognition; Succession of sounds; Simultaneity of sounds; Music sounds.



Quest'opera è distribuita con licenza [Creative Commons Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Sommario

In questo articolo, analizzo come funziona il sistema uditivo quando è esposto a diversi tipi di componenti musicali, come le melodie, le armonie, le semplici successioni di suoni e le simultaneità ordinarie di suoni. Sosterrò che, mentre melodie e armonie sono percepibili, le semplici successioni di suoni e le simultaneità ordinarie di suoni sono elaborate a livello cognitivo. Questa conclusione si basa sull'analisi di come il sistema uditivo fa uso del raggruppamento primitivo, del (semplice) raggruppamento basato su schemi, e del raggruppamento cognitivo basato su schemi al fine di identificare e categorizzare il paesaggio uditivo in flussi musicali significativi.

Parole chiave: Melodie; Armonie; Percezione; Cognizione; Successione di suoni; Simultaneità di suoni; Suoni musicali.

1. Introduction

When we hear a *melody*, we tend to hear a sequence of sounds as forming a unitary whole. That is, we group consecutive sounds in a unified composite that evolves over time and of which we can recognize, most of the time, a beginning, a middle, and an end. Similarly, when we hear a *harmony*, that is simultaneous sounds that are harmonic – which means that either they are clear integer multiples of a fundamental or they can be traced back to integer multiples of a fundamental –, we tend to group those sounds as forming a unitary whole, which is heard as a blended composite unity. Melodies and harmonies, together with rhythm¹, are the paradigmatic music components, which constitute the vast majority of Western music. Because of the direct and fast way in which we can identify melodies and harmonies when we listen

¹ In this work, I will not deal with rhythm as a music component taken by itself but only as a crucial feature determining, on the one hand, the occurrence of melodies and harmonies and, on the other hand, the occurrence of the mere succession of sounds and the ordinary simultaneity of sounds.

to music – that is it seems that we can easily group consecutive sounds in a composite evolving unity, and group simultaneous sounds in a blended composite unit – we have good *prima facie* reasons to think that melodies and harmonies are genuinely perceivable. That is, they are processed at the perceptual level.

Nevertheless, let us now focus on the musical compositions in which we are not able to identify clear melodies and harmonies. These are musical works – most of them are part of the Western tradition and composed after the 1950s – in which we do not hear consecutive sounds as constituting a unitary whole because they are heard as a mere *succession of sounds*. That is, despite the hearer can still perceive a sense of unity that accompanies this succession of sounds, she does not seem to be able to sequentially unify them in a way that she can detect a beginning, a middle, and an end. Analogously, there are some music compositions in which simultaneous sounds are not blended as forming a harmony because they do not stand in a harmonic relationship to each other.

So these simultaneous sounds are heard as merely taking place at the same time. That is as if the hearer recognizes them as an ordinary *simultaneity of sounds*² even though these sounds can still be perceived as accompanied by a feeble sense of unity. For the hearer to identify and, eventually, recognize that the succession of sounds and the simultaneity of sounds are music components it takes more time and effort, if compared to the time and effort it takes to identify and, eventually, recognize melodies and harmonies. Therefore, we have good *prima facie* reasons to think that the succession of sounds and the simultaneity of sounds are music components that are not genuinely perceived, but processed at the cognitive level.

² I referred to succession of sounds and simultaneity of sounds as “mere” succession of sounds and “ordinary” simultaneity of sounds to underline the fact that these sounds are not unified in melodies and harmonies, as melodies and harmonies are also succession of sounds and simultaneity of sounds. In the rest of the paper, for the sake of simplicity, I will omit “mere” and “ordinary” and talk of succession of sounds and simultaneity of sounds.

In this paper, I will go beyond the *prima facie* reasons to claim that melodies and harmonies are perceivable items, and that succession of sounds and simultaneity of sounds are processed at the cognitive level. I will show that melodies and harmonies are perceivable properties of music compositions because they are graspable via a psychological *perceptual* process. In order to do that, I will appeal to Bregman's auditory scene analysis discussion of the psychological process which is responsible for the perceptual grouping of such music components. I will then show that succession of sounds and simultaneity of sounds are, instead, graspable at the *cognitive* level.

The paper is organized in the following way: Bregman's view will be preliminarily introduced in Section 2. In Section 3, I will first analyze the case of melodies and, to illustrate how the perceptual psychological process takes place when hearing them, I will analyze Johannes Sebastian Bach's Third Brandenburg Concerto and will refer also to some music by Francisco Tárrega's and Augustín Barrios'. Then, I will show how, to detect, instead, successions of sounds the psychological process which is involved is cognitive rather than perceptual. I will analyze some music compositions by Elmir Mirzoev, Salvatore Sciarrino, and György Ligeti which embed successions of sounds instead of melodies. In Section 4, I will first analyze harmonies and then the case of simultaneities of sounds to show how the former are perceivable while the latter are cognitively accessible. My analysis will be based on Bregman's account and McAdam and Bigand's conception of the cognitive psychology of audition. Finally, in Section 5, I will conclude with some general remarks on the different ways in which we grasp music components.

2. Albert Bregman's *The Auditory Scene Analysis*

The cognitive psychologist Albert Bregman in his monumental work *The Auditory Scene Analysis. The Perceptual Organization of Sound* (1990) describes how the auditory system works when immersed in the messy

auditory landscape. Bregman tries to understand how the auditory system can make sense of the chaotic auditory landscape which surrounds us by organizing the auditory stimuli (that is by segregating and grouping them) in a way that we will attribute a specific auditory stream³ to the auditory source which has produced it. This ability of the auditory system is very important for the detection and the recognition of the objects which are the causes of the sounds we hear. The auditory system works thanks to a two-fold perceptual process: *primitive grouping* and *schema-based grouping*. The primitive grouping is responsible, first, for the detection of auditory streams' components such as frequencies, amplitudes, and spatio-temporal properties⁴. Then, it is in charge, on the one hand, of grouping stream's components by attributing them to a specific auditory stream and, on the other hand, segregating the stream's components that do not belong to the same specific auditory stream. The primitive grouping takes place at two levels: *sequentially* and *simultaneously*. That is, it groups and segregates the sounds that evolve over time and those that take place simultaneously. The primitive grouping operates by employing Gestalt principles which are similar to the ones suggested by the psychologists of the Gestalt in 1920s, such as *proximity*, *similarity*, *common fate*, *exclusive allocation*, *context*, *closure*, and *organization* (Bregman 1990, 196–202, 248–293).

Once the auditory system, thanks to the primitive grouping, has segregated and grouped auditory stimuli attributing them to a specific auditory stream,

³ In this paper, when talking in general about how the auditory system groups and segregates the auditory landscape I will use “auditory stream” and “sound” interchangeably. Nevertheless, in some cases, talking of auditory streams captures more exhaustively the multifaced auditory landscape. For example, when a sound is made of many specific elements (such as footsteps, rain sound, and so on), it is more appropriate to talk about auditory stream than sound.

⁴ Frequencies, amplitudes, and spatio-temporal properties are sound waves' characteristics. They are physical and, thus, objective characteristics. I take for granted that they correspond to the sound audible subjective properties of pitch and loudness. I am well aware that it is not always the case, as in some circumstances a frequency, for example, does not correspond to a specific pitch – and the same applies to amplitude vs loudness. These differences do not matter in this paper, so I will use waves' characteristics and sound subjective properties interchangeably.

the schema-based grouping comes into play. At the schema-based grouping level, which is a higher-order process than the primitive grouping but still perceptual, then, the auditory system applies schemas about recurring sound patterns to properly identify and categorize the different sounds or streams of sounds. These schemas are then applied to the material already grouped and segregated via the primitive grouping. According to Bregman, all the different categories of sounds that surround us are perceptually organized by using the two-fold perceptual process just described. That is, usually sounds are organized into three different categories: environmental (or everyday) sounds – that are the sounds that surround us in everyday life such as the slamming of doors, the jiggling of keys, the baby crying, and so on – speech sounds – that are the sounds that people utter and which have a semantic meaning – and musical sounds – that are the sounds which constitute music. All three categories of sounds are grouped, segregated, identified, and categorized by virtue of the primitive grouping and the schema-based grouping. Let me make an example to show the application of the two-fold process within the auditory context of environmental sounds. Imagine being in the street and your auditory landscape is populated by different auditory streams, such as the sounds of cars that pass close to you, of someone talking to you, of a vehicle horn, and of an ambulance that is approaching you. In this situation, the primitive grouping will be responsible for segregating the different auditory stream's components and grouping the components that belong to a specific stream, such as, say, a siren's wail, to that stream. Then, when using the schema-based grouping, the auditory system will be able to recover the right schema stored in memory to properly identify and then recognize the auditory stream as a siren's wail.

Even though the schema-based grouping is a perceptual process, there are cases in which it is not sufficient to identify and categorize an auditory stream. In these cases, the auditory system still employs a form of schema-based grouping but not of the perceptual kind we already discussed rather of a higher-level kind, that is the 'cognitive' schema-based grouping (Di Bona

2023, 5). When the sensory information is particularly complex and vague and the perceptual schema-based grouping does not suffice to provide a coherent representation of the auditory landscape, the cognitive schema-based grouping comes into play by using cognitive resources. The cognitive mechanisms which are used by the cognitive schema-based grouping are ‘higher-level processes that bring into play mental representations, decision making, inference, and interpretations by the perceptual system’ and these systems ‘are necessary to elaborate a coherent representation of the sound world (McAdams and Bigand 1993, 1-2)’. Imagine that you have never heard a siren wail in your entire life. The first time you hear it, most probably you will not have the right schema stored in your memory – as you have never heard such sound before – so that you can identify and recognize the sound you hear as the sounds of a siren’s wail. Therefore, in order to make sense of the new sound you hear, the primitive grouping and the schema-based grouping are not enough. McAdams and Bigand (*ibidem*), to make clear the functioning of the cognitive mechanism at the basis of the cognitive schema-based grouping, suggest the following example:

Imagine for an instant that we are being guided through an Amazonian rain forest: we would hear exactly the same noise as the native of the region that accompanies us but we would be incapable, because of our lack of knowledge of the environment, to extract from the sound background events corresponding to the cries of iguanas and Macaques, the songs of Wistiti monkeys, or the rustling of the leaves of tropical trees, nor would we be able to assign meanings to entire sound structure that might in the long run be important for survival.

The idea is that natives listening to the Amazonia rain forest have the right schemas to directly grasp the different auditory happenings; whereas the non-natives, in the same auditory circumstance, would not have the same psychological resources as the natives so they will need to use the cognitive

process of interpretation to find the correct schemas which allow them to make the auditory soundscape they are immersed in intelligible.

The moral is that in some auditory circumstances, the auditory system will need to employ not only the primitive grouping and the schema-based grouping, but also the cognitive schema-based grouping in order to provide a meaningful representation of the new auditory stream it is in front of. Bregman maintains that there are complex auditory stimuli that go beyond his analysis so he does not mention the cognitive schema-based grouping. Nevertheless, McAdams and Bigand (1993) focus solely on the high-level processes which are necessary to “interpret” the auditory stimuli that are particularly vague to be processed only via the primitive grouping and the (perceptual) schema-based grouping. As stated in the quotation of their work above, the cognitive high-level processes they refer to involve decision-making and inferences.

As for the primitive grouping and the schema-based grouping, also the cognitive schema-based grouping is used to order the auditory landscapes made of all three categories of sounds. Nevertheless, in what follows, I will discuss only musical sounds, leaving aside speech sounds and environmental sounds⁵. I will focus on music components and show that while melodies and harmonies can exhaustively be identified and recognized by the exclusive employment of the primitive grouping and the schema-based grouping, the music components of the successions of sounds and simultaneities of sounds require the cognitive schema-based grouping to be identified and recognized, as the primitive grouping and the schema-based grouping do not suffice.

3. Melodies and successions of sounds

Chapter V of *The Auditory Scene Analysis* is dedicated to the perceptual organization of music. As I have already mentioned in Section 2, the auditory

⁵ For a discussion of the perceivability of speech sounds and environmental sounds see Di Bona (2020, 2022b).

scene analysis takes place sequentially and simultaneously. Generally speaking, these two dimensions correspond to the two dimensions of the staff in which notes are written. Indeed, in the staff, you have a horizontal dimension and a vertical dimension. Notes evolving over time are represented in the horizontal dimension, while simultaneous notes are represented in the vertical dimension. The horizontal dimension corresponds to sequential grouping and it is also the spatial dimension in which typically melodies are represented, whereas the vertical dimension represents simultaneous notes and it is also the spatial dimension in which harmonies are typically represented. In this paragraph, we focus on the sequential grouping therefore on the grouping (and segregation) of melodies.

According to Bregman, the primitive grouping works by virtue of the application of Gestalt principles. These principles determine the grouping and segregating of frequencies, amplitudes, and spatio-temporal properties that constitute the multiplicity of the musical landscape. They contribute, indeed, to the formation of melodies. Let us see now how each Gestalt principle operates practically.

The proximity principle states that we tend to group tones that are close in frequency and segregate, instead, those which are far in frequency. To exemplify this principle, Bregman considers the case of a sequence of high-pitched tones (As) and low-pitched tones (Bs) that alternate following a certain speed. Instead of hearing an auditory stream made of an alternation of As and Bs, we tend to group As all together as forming an auditory stream, and all Bs as forming another auditory stream. That is because As are similar in frequency – given that they are all high-pitched tones – but different from Bs which are low-pitched tones. Likewise, we group all the Bs together because they are similar in frequency – given that they are all low-pitched tones. The necessary condition in order to group tones together is not that the tones should be the same, but that they should be sufficiently close in the scale of the pitches. Moreover, also the tone rate at which the tones are performed is very important. If the tones are played quite slowly, then As and Bs will

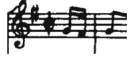
not be heard as forming two separated auditory streams but as constituting a single (and quite slowly performed) auditory stream. This principle suggests that when we hear melodies, it is because we have grouped together as composite unities tones that are close in frequency.

The proximity principle has been largely used to compose many melodies, especially in Baroque music. Baroque composers seem to be well aware of the fact that in order to create a good melody – that is a series of tones that is heard as a composite unity which then could also be easily performed and well-remembered by the perceiver – they had to make sure that the rate of the tones ensure segregation and that the tones were sufficiently close in frequency. The requirement of the tones to be close in frequency in order to group tones as forming the same stream was stated already in the early 1920s by Ortmann (1926), who found out that melodies written by Franz Schubert, Robert Schumann, Johannes Brahms, and Richard Strauss were constituted, in the majority of the cases, by the smallest intervals between tones. Concerning the tone rate of the tones, instead, Dowling (1973) found out that it is extremely difficult to have a clear segregation when the median tone rate is 6.3 tones per second. To reach this conclusion, Dowling examined twenty recordings of Baroque music. Let us now analyze the functioning of the Gestalt principles of context and exclusive allocation, and show also another application of the proximity principle. In this regard, Bregman and Rudnický (1975) suggested a specific experiment. A listener had to hear two tones, A and B, and distinguish between the case in which the two tones were heard as forming an isolated element – in which case the listener was able to order the tones saying which one was higher – and the case in which, instead, the tones were somehow heard as a specific element but not as clearly as to be able to tell which tone is higher and which one is lower. When A and B occurred one after the other, the listener was able to say whether the order was high-low or low-high. When, instead, the two tones were embedded in a longer sequence anticipated and followed by two tones “F” (“F” stands for flankers) of equal frequency, the listener was not able to order A and B anymore. The idea is

that when A and B are embedded in a longer perceptual sequence, the listener perceives the whole thing differently compared to when the tones are heard as an isolated element. In the former case, the listener does not have the ability to individuate the specific properties of A and B (as, for example, their pitch). Bregman and Rudnický also examined what happens when the Fs and AB are embedded into an even broader auditory stream which, together with the Fs, included also the further tones, Cs.

The Cs tones preceded and followed the FABF sequence, originating the larger perceptual sequence CCCFABFCCC. They found out that when the Cs tones were much lower in frequency than Fs, Fs were grouped together with AB again, and the AB order was lost once more. Instead, when Cs were close in frequency to Fs, AB became audible again, and the listener was able to order them again. In this second case, the listener tended to group the tones as forming the perceptual stream CCFFCC. As we have already seen, tones tend to group with those that are closest in frequency, thus proximity explains why Cs were grouped with Fs instead of AB. The perception of the order of frequency of tones A and B was based on the allocation of Fs, depending on whether the AB element was inserted in a longer sequence “surrounded” by Fs or was isolated. The allocation of Fs either in the sequence CCFFCC or in the sequence CCCFABFCCC is explained by the exclusive allocation principle. That is to say that Fs cannot be allocated to both sequences, in this sense the allocation is ‘exclusive’. Finally, when deciding whether Fs should be allocated to either one sequence or the other, the auditory system employs the principle of context. Namely, it evaluates the possible perceptual contexts in which the Fs are embedded, before deciding to which sequence to attribute them. To summarize, focusing on how we hear AB helps us understand the functioning of the proximity principle; focusing on how we hear Fs helps us understand the functioning of exclusive allocation; and focusing on the audibility of AB and Fs inserted in broader contexts exemplifies the principle of context. The principles of context, exclusive allocation, and proximity are used to group and segregate many melodies. I will discuss a specific example

of melody that comes out of the interaction of different musical instruments, each producing a specific melodic element (Di Bona, 2022, pp.256-257). I will focus on the first movement of Bach’s 3rd Brandenburg concerto.

The movement starts with this melodic fragment  which is repeated many times and played by different instruments (such as violins, violas, cellos, double basses, and harpsichord) to form a variety of melodies. It is a fragment of three notes, the first and the third having the same frequency and the middle one a lower frequency. This fragment is a fundamental element to be heard for the melodic lines to be detectable. In order for this element – which is similar to the AB element analyzed by Bregman – to be perceived clearly – that is to say that the listener can correctly order the three sounds in terms of their frequency –, J.S. Bach wrote it in a way that it never appears in a larger sequence which is anticipated and followed by notes of the same frequency of the AB element (as in the FABF case). Also, the tones that usually follow and precede FABF, which following Bregman’s analysis we can be labeled as Cs, are written in a way as to be closer in frequency to Fs instead of AB, so that AB is always somehow discriminable (Di Bona *ibid*). Figure 1 is J.S. Bach’s score of the incipit of the 1st movement of the 3rd Brandenburg Concert. AB, FABF, and CCFABFCC are marked in the score in order to make clear how the principles of proximity, context, and exclusive allocation are taken into account by J.S. Bach to compose a melody.⁶

⁶ For a more detailed analysis of the score see Di Bona (2022). Di Bona (2022) focuses on melodies created by different musical instruments (melodic chimera) but obviously everything which is said about multi-instrumental melodies work also (and *a fortiori*) for melodies played by a single instrument, which is the focus in this paper.



Figure 1. J.S. Bach's score of the incipit of the 1st movement of the 3rd Brandenburg Concert. AB, FABF and CCFABFCC are marked in order to make clear how the principles of proximity, context and exclusive allocation are taken into account in order to compose a melody.

Another Gestalt principle to be discussed as it is crucial for sequential integration, so for the formation of melodies, is the principle of closure. The principle of closure explains the phenomenon of the illusion of continuity. The illusion of continuity can have different forms as it can occur in different auditory contexts.⁷ The effect takes place when there are tones that alternate with noise bursts. When the alternation is quite rapid, the listener hears a sequence of tones instead of a sequence of tones and noise bursts (which is what actually constitutes the sequence). Bregman claims that, despite distracting elements, the auditory system tends to unify some tones into “strong” perceptual forms (such as a steady sound in audition or a circle in

⁷ See Di Bona (2023) for an analysis of the illusion of continuity (or amodal completion) in the auditory context of environmental sounds and speech sounds.

vision), by virtue of the application of the Gestalt principle of closure (Ibid: 25.28). In the musical context, the effect has been studied extensively by D. Deutsch (1999). She reported that in many compositions in twentieth-century guitar literature, the illusion of continuity is employed in order to create the effect of hearing a steady sound which is a melodic line. Indeed, steady sounds produced by the illusion of continuity (which contributes to form melodic lines) can be found in the works of Francisco Tárrega's *Recuerdos de la Alhambra* and Augustín Barrios' *Una Limosna por el Amor de Dios*. In both cases, the listener hears notes which are not there, that is because when tones and pauses are quickly alternating, and the pauses are sometimes substituted by a different tone, the listener generates the missing tones.

The Gestalt principles of proximity, closure, context, exclusive allocation, and organization are the tools used by the primitive grouping to group and segregate auditory stimuli to form unified wholes such as melodies. This process is a sequential process, that is the auditory system groups (and segregates) consecutive auditory stimuli to form melodies. After the primitive grouping has completed its work allowing us to distinguish a melody from another melody or form a different kind of sound (such as an environmental sound or a speech sound), the schema-based grouping can use the already organized material by the primitive grouping to identify and categorize auditory streams. So, the schema-based grouping will allow us to tell whether we are hearing an auditory stream (such as melody1) from another auditory stream (such as melody2). Moreover, it will also allow us to categorize an auditory stream as a specific one by virtue of the application of schemas already stored in our memory.

That is, if we have already heard joyful melodies many times in the past (maybe because they are composed in major keys) we have already stored in memory the schema "joyful melody", therefore we are most probably able to tell whether the melody we are hearing is joyful. The same applies to sad melodies, and so on. It could also happen that if we are an expert in, say, J.S. Bach's melodies, we have stored in our memory schemas corresponding to

J.S. Bach's melodies therefore when listening to the incipit of the third Brandenburg Concert, we would be able to categorize the melody not only as such or as joyful but also as a Bachian melody. To conclude this section, it is important to recall that both the primitive and the schema-based grouping contribute to the *perceptual* organization of the auditory scene analysis. They suffice, at the sequential level, to group, segregate, identify and (in most cases) categorize a succession of sounds as melodies, therefore melodies are perceivable items. Note that the analysis I am pursuing applies only to the melodies in the Western music tradition composed within the tonal system⁸.

Let us now take into account music compositions that cannot be categorized as tonal music compositions in which, therefore, we cannot individuate a clear tonality on which the compositions are based on. Many compositions written starting from the 1950s are of this kind. When listening to such compositions, the question is whether we can unify successive sounds as forming melodies in the same way in which we unify successive sounds as forming melodies when listening to tonal music. As we already discussed, according to Bregman, in order to unify successive sounds into composite wholes, there should be specific auditory stimuli that are grouped (and segregated) by virtue of Gestalt principles. Let me analyze some music compositions in which Gestalt principles are not followed, so it seems that we cannot use primitive grouping to group auditory streams. I will start by analyzing Elmir Mirzoev's *Allegoria Sacra* for two performers with violin, prepared piano, percussions, Tibetan singing bowls with microphones, tape, and multimedia. The piece is based on an interaction between two performers who most of the time play violin and piano but also play percussions, and Tibetan singing bowls and read out poems. The first sentence played by the

⁸ Obviously, I am not claiming that all tonal Western music compositions have melodies. There might be compositions within this field – such as some of Vivaldi's second movements of various concertos for violin solo or two, three, or four violins that are exclusively made of chords) in which even though the tonality is very clear, the composer does not intend to compose recognizable melodies. In these cases, she will probably not be using Gestalt principles usually at play when forming melodies.

violin, after a kicking incipit by the piano, is made of tones (such as E flat, A, E, and B flat, and then, after a pause, D, F, B, B flat an octave below, E), played in *tremolato*, which are not close in frequency. This explains why the auditory system does not group these tones. Moreover, it seems that the principles of context, closure, and organization are not employed in a clear way to allow the grouping of tones into melodies. That is, we would not find clear and recognizable segment made of notes we can order in terms of frequency and not inserted in a larger perceivable sequence which would make it invisible. Not even the principle of proximity applied in the form of the illusion of continuity has been employed to group and segregate the auditory streams composed by Mirzoev. Nevertheless, in this composition, there are still some passages that can be experienced as somehow unified, because of the rhythmic repetition of the tones which helps group them. What is most important, though, is that in order to complete the processes of grouping, segregating, and then identifying and categorizing successive sounds in unitary wholes (which are not melodies but mere unified percepts), we should definitively employ the cognitive schema-based grouping. Indeed, the primitive grouping starts a quite partial and insufficient grouping and segregation process. Then, most of the time, we do not have already stored in memory the schemas that allow us to identify and recognize the auditory streams as composite melodic elements. In order to complete the identification and recognition process and provide a coherent representation of the sound world, the auditory system has to interpret the auditory stimuli by using the cognitive mechanism of decision-making and inferences and decide on which element directs the attention. Then, it can compare the tones it is hearing with the tones stored in the memory that were heard in the past to come up with plausible schemas that allow it to complete the categorization process. All of it can take place thanks to the cognitive schema-based grouping. This process is not immediate, it takes time. Indeed, the listener should listen to *Allegoria Sacra* many times before she can automatically identify and recognize the auditory musical streams constituting the piece.

Before the listener is well acquainted with *Allegoria Sacra*, she will interpret what she is currently listening to, by comparing what she is listening to what she has listened to in the past and which resembles the music components of *Allegoria Sacra*. She will compare past musical experiences with the current one and *decide* whether there are similarities which will eventually lead to the identification and recognition of musical passages.

I will briefly mention also Salvatore Sciarrino's *Il Silenzio degli Oracoli* (1989) for wind quintet and Gyorgy Ligeti's *Artikulation* (1958). These are two examples of musical pieces in which, like in Mirzoev's composition, melodies are absent but successions of notes are somehow groupable because they are composed by tones that have a recurrent rhythm that contributes to the grouping. More specifically, in Sciarrino's piece, all the wind instruments play musical patterns that vary in rhythm and notes. For example, there are some passages in which the horn player plays notes by blowing into the instrument without producing sound, and others in which, when blowing into the instrument, she has to pronounce phonemes indicated in the score. After the listener has listened to these passages several times, she will most probably be able to group the tones because she will use the rhythmicity of the patterns as a hint to do that. Moreover, and most importantly, in order to properly group those tones, identify them and recognize the specific succession of notes as auditory streams, the listener has to use the cognitive-schema grouping. That is, a cognitive mechanism of interpretations of tones is at play to help provide a coherent representation of the musical landscape. The same with Ligeti's work, which is a piece composed entirely of electronic sounds. In this case, also a minimal sense of a unitary succession of sounds is very hard to detect because, not only there are no tones that are close in frequency, but Ligeti does not write rhythmic recurrent patterns. Therefore, even the smallest sense of primitive grouping is very difficult to obtain. What matters in this specific composition, then, is the cognitive schema-based grouping which is crucial for the listener to make sense of what she is listening to. This means that, as

already stated, the listener will have to compare what she is listening to with similar electronic sounds she has listened to in the past. Then, she will have to infer and somehow guess which are the successive sounds that can be grouped together, their individuation and categorization.

In this paragraph, I have explained the reasons why melodies are perceivable music components while successions of sounds are processed at a cognitive level. That is because they are grasped via two different psychological processes: the latter thanks a perceptual process while the latter via a perceptual process with the aid of the cognitive mechanism of the schema-based grouping. In the next paragraph, I will discuss harmonies and simultaneities of sounds to reach a similar conclusion to the one on melodies and successions of sounds already mentioned. Namely, harmonies are perceivable while simultaneities of sounds are processed at a cognitive level.

4. Harmonies and simultaneities of sound

The primitive grouping and the schema-based grouping organize the auditory scene not only to group and segregate sounds that are successive but also in order to group and segregate sounds taking place simultaneously. The grouping of simultaneous sounds gives rise to harmonies, that is to unitary auditory compounds which are heard as an assemblage of harmonic sounds. In this paragraph, I will describe how harmonies are heard as perceptual items by showing how the primitive grouping and schema-based grouping work in the case of harmonies in a similar way to the way in which they work in the case of melodies.

Bregman (1990: 221) states that the basic question at the core of the functioning of the simultaneous integration process is: “How do we know which acoustic components have arisen simultaneously from the same physical event?”. Similar principles to the ones influencing sequential grouping are at work in simultaneous integration. In particular, the principles of proximity and common fate apply to amplitude, frequency, and spatial properties when auditory stimuli have to be ordered in auditory streams. Let

us start by discussing proximity for the frequency of simultaneous sounds. As we already said, the principle of proximity applied to successive sounds claims that we tend to group together the sounds that are close in frequency. The same applies in the case of simultaneous sounds. More specifically, thanks to the *harmonicity principle*, we tend to group simultaneous sounds that resemble in frequency especially when they are multiple integers of a specific sound which is the so-called “fundamental”. The idea is that sounds, the partials, which have a relationship with a sound that plays the role of an “attractor”, the fundamental, tend to be fused into a single percept. We usually unify those sounds close in frequency to form chords. Indeed, most musical compositions are based on simultaneous sounds perceived as chords, namely as unified harmonic composites. We could analyze any tonal music composition of the Western tradition which is then based on a clear tonality to have instances of the application of the harmonicity principle.

For example, the first movement of W.A. Mozart's Piano Sonata n°16 in C Major K.V. 545 evolves by virtue of tonalities such as C Major, G Major, G minor, D minor, A minor, F Major, and C Major. At the end of the section based on each tonality, the piano ends with the chords which are built on the tonic of the correspondent scale. Let us take the last session, and conclusion of the first movement, as an example, which is based on C Major. The last three chords of the movement are just three C major chords. Each chord is perceived as a unified auditory amalgam because the notes that compose the chord are integer multiple of a fundamental. That is, the notes which compose each chord are multiple integers (of course with approximation) of the C at 65 Hz (see Figure 2).



Figure 2. This is the last bar of the first movement of W.A. Mozart's Piano Sonata n°16 in C Major K.V. 545. Because the notes composing the chords are multiple integers of a fundamental which is C at 65 Hz, the three simultaneous groups of sounds will be perceived as fused in three well-unified chords. The principle of harmonicity is responsible for this perception.

The common fate principle is another Gestalt principle used for sequential grouping. This principle claims that when the characteristics of an auditory stream change in loudness or frequency if these changes happen simultaneously, we still hear this auditory stream as we originally heard it. As if no changes had occurred. Also, when tones start and end at the same time, which means that they are close in time (proximity principle), that is they are synchronic, they tend to be grouped in a unified auditory stream. Both common fate applied to loudness and frequency, and proximity in time are massively employed in sequential integration of simultaneous sounds originating unified percepts that are chords. It is important to note that the harmonicity principle is more powerful than proximity in time and common fate for sequential integration that is when applied it gives a strong sense of unity to simultaneous sounds. Instead, when this cannot be applied because the sounds are not integer multiple of a fundamental and the auditory system turns to proximity in time or common fate to unify sounds, still there is a sort of unitary sense of simultaneous sounds but looser than the one obtained thanks to the harmonicity principle. The vast majority of simultaneous sounds of *Allegoria Sacra* does not follow the harmonicity principle so they are not fused in a unitary composite. For example, at the beginning of the second movement of Mirzoev's piece, *Notturmo slancio*, the piano plays B Flat of 58.3 Hz (it is a very short sound that resonates for a few seconds) and none of the notes played by the violin while the piano plays B Flat are integer

multiple of B Flat (see Figure 3). Therefore, we cannot fuse the simultaneous sounds into a unified whole. Analogously, most of the simultaneous sounds written in Sciarrino's *Il Silenzio degli Oracoli* do not constitute chords that are made of harmonics, so they are not in a multiple integer relationship with a fundamental.

The image shows a musical score for Violin and Piano. The Violin part is in the upper staff, marked 'Vivo' with a tempo of 100. It features a 'pizz.' (pizzicato) section with a 7:6 ratio indicated above it. The Piano part is in the lower staff, also marked 'Vivo' with a tempo of 100. It features a very short B flat note marked 'sfz' (sforzando) in the bass clef.

Figure 3. This is the incipit of the second movement, *Notturmo Slancio*, of Elmir Mirzoev's *Allegoria Sacra*. The simultaneous tones are the notes played by the violin while the piano plays a very short B flat of 58.3Hz which resonates for a few seconds. The notes played by the violin are not integer multiples of the B flat of 58.3 Hz. This seems to justify why we do not fuse violin notes and the piano note into a unified series of chords that form harmonies.

Nevertheless, the application of the proximity principle to time and the application of the common fate principle to loudness and frequency could attribute a sense of unity to percepts of simultaneous sounds in some passages in both Mirzoev's and Sciarrino's works. However, this sense of unity does not lead to the audibility of harmonies, which are instead genuinely audible. That is because the schema-based grouping cannot use material that has been already exhaustively grouped (and segregated). Indeed, in order then to make sense of the simultaneous sounds and to properly identify and categorize them into meaningful auditory streams, the auditory system will employ the cognitive schema-based grouping mechanism. It will actively compare the current percept with a previous one in order to find the right schema to be applied. Decision-making and inferences will be at work, then, to make sense of the simultaneities of sounds.

5. Conclusion

In this paper, I first mentioned *prima facie* reasons according to which we can tell whether we are listening to melodies and harmonies because we seem to be able to identify and recognize them in a very fast and direct way. Then, I also mentioned *prima facie* reasons according to which, instead, when we listen to successions of sounds and simultaneous sounds, it takes time for us to identify and recognize them as constituting specific auditory streams. Moving from these *prima facie* reasons, I analyze how the auditory system works in front of different kinds of music components. I concluded that while melodies and harmonies are perceivable, successions of sounds and simultaneities of sounds are processed at the cognitive level. This conclusion is based on the analysis of how the auditory system employs the primitive grouping, the schema-based grouping, and the cognitive schema-based grouping when ordering the auditory landscape by virtue of the identification of meaningful musical streams. Surely, there is an impact on the aesthetic appreciability of music which is based on the different ways in which we

grasp music components. This impact has to be properly evaluated and I leave it for further research.

References

- BREGMAN, A. S., *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge (MA), The MIT Press, 1990.
- BREGMAN, A. S., RUDNICKY, A., “Auditory Segregation: Stream or Streams?”, *Journal of Experimental Psychology: Human Perception and Performance*, 1, 1975, pp. 263-267.
- DEUTSCH, D., “Grouping Mechanisms in Music”, in DEUTSCH, D. (ed.), *The Psychology of Music*, San Diego, Academic Press, 1999, pp. 299-348.
- DI BONA, E., “A Disjointed Account of the Illusion of Auditory Continuity: In Favor of Hearing Everyday Sounds but Against Hearing Semantic Properties”, *Inquiry*, 2023.
- DI BONA, E., “Hearing Chimeras”, *Synthese*, 200 (3), 2022a, pp. 1-20.
- DI BONA, E., “A Skeptical Approach to the Audibility of Semantic Properties”, *Argumenta*, 14 (7, 2), 2022b, pp. 543-554.
- DI BONA, E., “On Hearing Meanings: Reflections on the Method of Contrast, Adaptational Effects, and Semantic Satiation”, *Rivista di Filosofia*, 111 (2), 2020, pp. 215-235.
- DOWLING, W. J., “Rhythmic Groups and Subjective Chunks in Memory for Melodies”, *Perception & Psychophysics*, 14, 1973, pp. 37-40.
- MCADAMS, S., BIGAND, E. (eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition*, Oxford, Clarendon Press/Oxford University Press, 1993.
- ORTMANN, O., “On the Melodic Relativity of Tones”, *Psychological Monographs*, 35 (162), 1926, pp. 1-47.