frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 14 December 2011
doi: 10.3389/fnhum.2011.00158

Auditory scene analysis: the sweet music of ambiguity

Daniel Pressnitzer [1,2]*, ***Clara Suied*** [1,2,3] ***and Shihab A. Shamma*** [1,2,4]

[1] *Centre National de la Recherche Scientifique and Université Paris Descartes, UMR 8158, Paris, France*
[2] *Département D'études Cognitives, Ecole Normale Supérieure, Paris, France*
[3] *Fondation Pierre-Gilles de Gennes pour la Recherche, Paris, France*
[4] *Electrical and Computer Engineering, University of Maryland, College Park, MD, USA*

Edited by:
Robert J. Zatorre, McGill University, Canada

Reviewed by:
Joel Snyder, University of Nevada Las Vegas, USA
Pierre Divenyi, Veterans Affairs Northern California Health Care System, USA

***Correspondence:**
Daniel Pressnitzer, Département D'études Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75230 Paris Cedex 05, France.
e-mail: daniel.pressnitzer@ens.fr

In this review paper aimed at the non-specialist, we explore the use that neuroscientists and musicians have made of perceptual illusions based on ambiguity. The pivotal issue is auditory scene analysis (ASA), or what enables us to make sense of complex acoustic mixtures in order to follow, for instance, a single melody in the midst of an orchestra. In general, ASA uncovers the most likely physical causes that account for the waveform collected at the ears. However, the acoustical problem is ill-posed and it must be solved from noisy sensory input. Recently, the neural mechanisms implicated in the transformation of ambiguous sensory information into coherent auditory scenes have been investigated using so-called bistability illusions (where an unchanging ambiguous stimulus evokes a succession of distinct percepts in the mind of the listener). After reviewing some of those studies, we turn to music, which arguably provides some of the most complex acoustic scenes that a human listener will ever encounter. Interestingly, musicians will not always aim at making each physical source intelligible, but rather express one or more melodic lines with a small or large number of instruments. By means of a few musical illustrations and by using a computational model inspired by neuro-physiological principles, we suggest that this relies on a detailed (if perhaps implicit) knowledge of the rules of ASA and of its inherent ambiguity. We then put forward the opinion that some degree perceptual ambiguity may participate in our appreciation of music.

Keywords: auditory perception, perceptual organization, bistability, auditory illusions, music

INTRODUCTION

This paper aims at highlighting some cross-connections that, we argue, may exist between auditory neuroscience, perceptual illusions, and music. More precisely, we address the issue of auditory scene analysis (ASA). ASA refers to the ability of human listeners to parse complex acoustic scenes into coherent objects, such as a single talker in the middle of a noisy babble, or, in music, a single melody in the midst of a large orchestra (Bregman, 1990). It has long been and still is one of the hot topics of auditory neuroscience, with its share of important advances and ongoing controversies (e.g., Shamma and Micheyl, 2010 for a review). ASA has also been studied in a musical context, with the hypothesis that many of the established rules of polyphonic writing in the Western tradition may be underpinned by perceptual principles (Huron, 2001). The aim here is not to repeat those arguments, but rather to provide a brief review, aimed at the non-specialist and biased toward perceptual illusions: we argue that illusions seem to be both a powerful investigation tool for neuroscientists and an important expressive device for musicians.

We will first briefly discuss the potential of illusions to reveal fundamental principles of perception in general. We will then describe the problem that ASA has to solve and what we know of the neural processes involved. For our purposes, we will emphasize recent studies, both behavioral and neuro-physiological, that have made use of so-called bistability illusions based on ambiguous stimuli. Finally, through a few musical illustrations and a computational model, we will suggest that perceptual ambiguity has been part of the composer's repertoire for quite some time. We will then conclude by speculating on the potential role of ambiguity in the appreciation of music.

ILLUSIONS AS A SIGNATURE OF PERCEPTUAL INFERENCE

Illusions are a vivid way to remind us about some basic but essential facts about perception. Take for instance the change-blindness illusion from vision (e.g., O'Regan et al., 1999)[1]. In a change-blindness paradigm, major parts of an image or a film can be modified, in full "sight" of the observer, but these changes will go unnoticed if they are not attended to. This has been taken as strong confirmation that visual awareness is not the result of the passive and obligatory registration of sensory information impinging on the retina, but rather, it is by essence an active exploration of the visual scene (O'Regan and Noe, 2001).

Another useful example is the classic Ponzo illusion (e.g., Murray et al., 2006), illustrated in **Figure 1**. Here, all blue objects in the figure have the same physical length, but they are usually perceived as one being taller than the other – even when the observer is fully aware that she/he is being "tricked." But are we really being tricked? Arguably, quite the opposite: the illusion reveals that we are able to

[1] A particularly dramatic illustration of the illusion can be found here: http://youtu.be/voAntzB7EwE

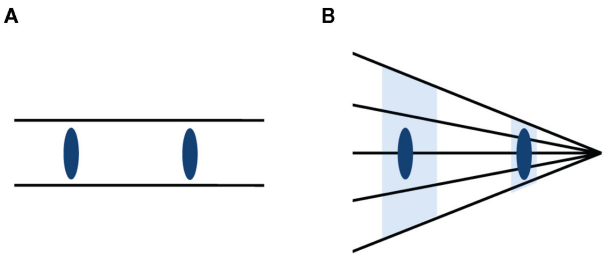

FIGURE 1 | Illustration of a visual illusion based on context. Two identical oval shapes are presented, either in a context suggesting that they are located at the same distance from the observer **(A)** or, using simple rules of perspective, in a context suggesting that one is further away than the other **(B)**. Most observers report experiencing different sizes in the **(B)** case, even though the retinal size of all oval shapes in all panels of the figure will be identical. This illusion is termed the Ponzo or corridor illusion.

make sense of complex sensory information on the basis of ecological plausibility. We are not really interested in the size of the image projected on the retina by the two objects, what is termed the proximal information. Rather, we would like to know what the size of each object is likely to be in the external world, what is termed the distal information. The perspective lines suggest that the second object is located further away than the first one. Somehow, the visual system is able to recognize that fact. As a result, the same proximal length on the retina is "seen" as two different distal sizes. The (useful) distal size is what enters awareness.

Note that, as introspection suggests, this does not seem to be a result from a laborious and voluntary reasoning about the laws of optics on the part of the observer: brain-imaging showed that the even the early stages of visual processing (primary visual cortex, V1) were modulated by the size illusion (Murray et al., 2006). Furthermore, the salience of visual illusions seems to increase with development: the mature visual system is more susceptible to it, perhaps because it "knows" more about the laws of optics thanks to experience (Kovacs, 2000; Doherty et al., 2010). More susceptibility to illusion means more accuracy in interpreting the visual scene.

How is this possible? Illusions have been studied since the beginning of experimental psychology, so any definitive answer would prove incomplete and controversial. The only point we wish to make here is that illusions seem to highlight the ongoing interaction between sensory input, which is noisy and inconclusive by nature, and some knowledge about the world that is embodied in perceptual systems (Barlow, 1997; Gregory, 1997, 2005). The precise way this is achieved is still a matter of debate. In vision, the Bayesian framework, which explicitly takes into account prior knowledge about the structure of the world, has been shown to account for several behavioral and physiological findings (e.g., Kersten and Yuille, 2003; Kersten et al., 2004 for reviews). Interestingly, in this framework, some perceptual illusions actually appear as optimal percepts given some simple prior rules governing physical objects (Weiss et al., 2002). Note that alternative schemes exist, where the observer does not try to make inferences about the state of the world (Purves et al., 2011 for a recent review). Here, perception keeps track of previous experiences in order to disambiguate future experiences by learning for instance the statistics of natural images. In all frameworks, we would suggest that illusions are an adaptation of perceptual systems to the regularities of the external world.

Illusions thus serve to illustrate a very simple but important idea. Perception is an active construct, more akin to a moment-by-moment gambling process than to a rolling camera or open microphone. In more elegant terms, Helmholtz (1866/1925) famously suggested that perceptual awareness was built from a succession of "unconscious inferences," aiming at producing the hypotheses about the state of the world that are most beneficial for guiding behavior. In the following, we will entertain the view that, as ASA also has to deal with ambiguous sensory data, it can be approached as a problem of perceptual inference.

AUDITORY SCENE ANALYSIS: THE PROBLEM AND HOW THE AUDITORY SYSTEM MAY SOLVE IT

THE ACOUSTIC PROBLEM

Among the various current opinions on ASA, there seems to be at least one that reaches a consensus among investigators: realistic auditory scenes can be rather complicated. At the core, sound is a one-dimensional physical phenomenon. An acoustic pressure wave impinges on one of our eardrums and it can only do one of two things: it can push the tympanic membrane a little bit, or it can pull it a little bit. Moreover, there is no occlusion between acoustic waves originating from different sources: as waves propagate through the air, they sum linearly at each point. As a consequence, at any moment in time, the little push or pull on the eardrum may be caused by one sound source out in the world, but it may also be caused by two sound sources, or by many sound sources, the number of which is unknown (**Figure 2**). This is what is known as an ill-posed problem in mathematics. There are too many unknowns (in fact, an unknown number of unknowns) for too few observations. The problem cannot be solved without further assumptions.

Each author has, at one point or another, tried to convey the intricacy of auditory scenes by a metaphor. Helmholtz (1877) evokes the interior of a nineteenth century ball-room, complete with "a number of musical instruments in action, speaking men and women, rustling garments, gliding feet, clinking glasses, and so on." He goes on to describe the resulting soundfield as a "tumbled entanglement of the most different kinds of motion, complicated beyond conception." His choice of metaphor may not have been totally neutral. The complexity of natural acoustic scenes is clearly large in general, but that of *musical* acoustic scenes can be positively daunting. Consider for instance the picture of **Figure 3A**, a photograph taken before the première of Gustav Mahler's eighth symphony. This work has also been dubbed the "Symphony of a Thousand," an obvious reference to the size of the orchestra and choir. An illustration of the resulting acoustic waveform (Sound Example S1 in Supplementary Material) is presented in **Figure 3B**. It seems impossible to infer, from there, how many sources where present and what they were doing.

But is the auditory system really expected to make sense of each and every one of the acoustic sources? This is not the case, fortunately. Mahler has a thousand potential acoustic sources at the tip of his finger when writing his score, but he will in fact use various

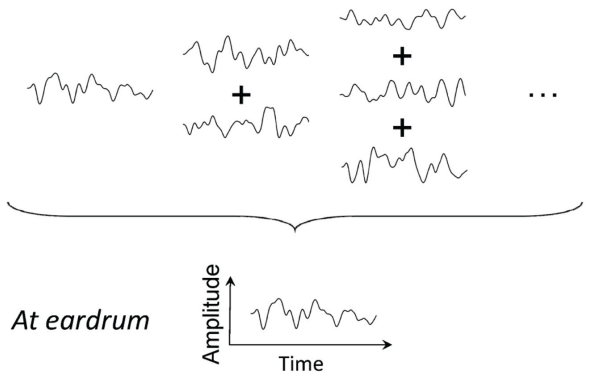

FIGURE 2 | Scene analysis as an ill-posed problem. All the pressure waves from different sound sources combine linearly in the air. As a result, any waveform observed at the ears may have been caused by one, two, or in fact an unknown number of sound sources. Determining the number and nature of sound sources from the compound waveform is an ill-posed problem.

devices to create only a tractable number of concurrent auditory objects (this tractable number may be rather low for the listener, Brochard et al., 1999). He does that by means of what could rightly be termed auditory illusions (we know there are many sources, we hear only one melody). This is a first hint of the intricate connections between ASA, illusions, and music, to which we will come back later.

THE SENSORY PROBLEM

Auditory processing begins by the transduction of the one-dimensional physical motion caused by acoustic waves into patterns of nervous activity in the auditory system. Because of the biophysics of the cochlea, the sensory receptor for hearing, this acoustic information is first broken into frequency sub-bands. The detailed mechanics of this transformation are beyond the scope of this paper, but for a review, see Pickles (2008). This so-called tonotopic organization is broadly preserved along the auditory pathway up to at least the primary auditory cortex.

When applied to a piece of music, tonotopic organization produces a representation similar to the illustration of **Figure 3C**. Tonotopy seems to help revealing patterns that were not obvious from the sound-wave. However, it also produces challenges of its own: now, the energy produced by a single sound source will be spread over several frequency channels and, consequently, will recruit distinct sets of sensory neurons. The problem that ASA has to solve can thus be rephrased as follows: given the flow of sensory information from the cochlea, which resembles a time–frequency analysis, the listener must determine the likely combination of physical sources in the world. Unfortunately, this is still an ill-posed problem. An exact solution being impossible, perceptual gambling must begin.

CUES TO ASA

A vast amount of psychophysical data has been accumulated on the topic of ASA (sometimes also referred to as the cocktail party problem, Cherry, 1953). A classic book also exists, which gave its modern name to the field (Bregman, 1990). More recent reviews are available (Carlyon, 2004; Snyder and Alain, 2007; Shamma and Micheyl, 2010). Here we will not go into any details, but rather sketch two possible accounts of ASA while emphasizing the role of inference processes in both of them.

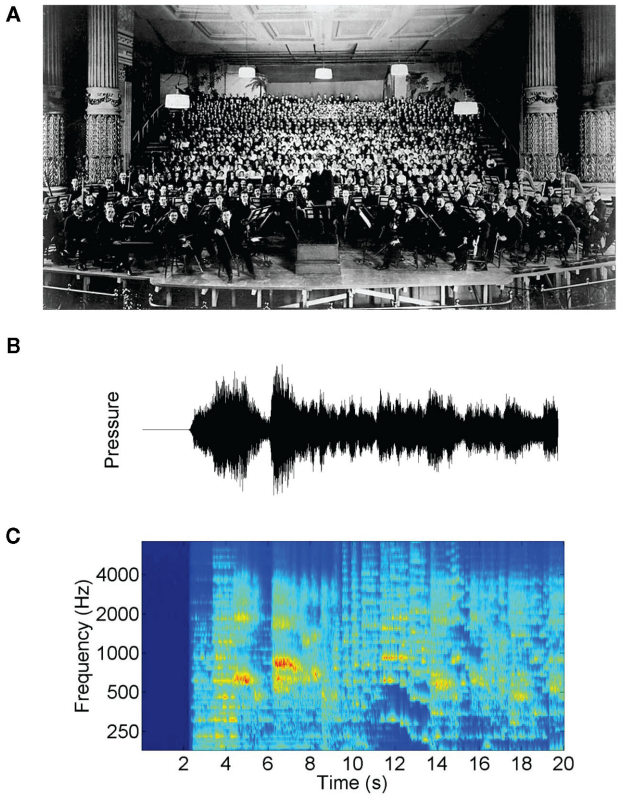

FIGURE 3 | The problem faced by auditory scene analysis. (A) Music creates acoustic scenes with a large number of potential sound sources, as illustrated by this picture taken before the American premiere of Mahler's eighth symphony – dubbed the "Symphony of a Thousand." *Image source: Wikipedia.* **(B)** The acoustic waveform of the first few seconds of Mahler's eighth symphony (see also Sound Example S1 in Supplementary Material). At any moment in time, the information available to the auditory system is the pressure value at the ear. This value may be reflecting vibrations from an unknown number of physical objects. The challenge of auditory scene analysis is to infer the most likely distal causes that account for the proximal pressure values. **(C)** Cochleogram of the waveform in **(B)**. The picture was obtained by passing the acoustic waveform through a model simulating the early stages of auditory processing (Shamma, 1985). The acoustic information is now spread over a two-dimensional time–frequency plane, as would be the case in the tonotopic channels of, e.g., the auditory nerve. The challenge is now at least twofold: to group all the activity belonging to one source and only to that source, even though it may be spread over many tonotopic channels; to bind over time the activity produced by a given source over time. In spite of the problem of being ill-posed, human and other animals are remarkably able at solving it and we are able to follow, e.g., a single speaker in a noisy environment. However, in the case of music, scene analysis usually fails: we cannot hear out each and every singer of the choir, even though they are distinct sound sources. This is precisely one of the points of the paper: how composers steer the inherent ambiguity of auditory scene analysis to achieve "illusory" musical effects.

Bregman (1990) suggested that ASA may be broken into two sub-problem. The first one is local in time and is termed vertical organization. Vertical refers to the frequency axis of **Figure 3C**: at any given moment in time, ASA needs to parse the distribution of energy over frequency channels into its plausible distal cause(s). The issue is twofold: acoustic sources are in general complex, so they produce activation over several auditory channels. It is thus important to be able to recognize these channels as belonging to one source. Also, pressure waves originating from different acoustic sources add up with each other, so it may be useful to be able to parcel out the contribution of each source to any given patch of activity in the time–frequency plane.

The general principle of ASA for Bregman seems to be one of perceptual inference based a heuristic ensemble of cues, each expressing a little knowledge about the way the acoustic world usually behaves. For instance, a cue to vertical organization is harmonicity. Harmonicity refers to the fact that any periodic sound can be represented by a stack of pure tones, and that these tones will exhibit a harmonic relationship: their frequencies will all be integer multiples of a fundamental frequency, f_0, corresponding to the inverse of the repetition period. Harmonicity is a strong cue: it would be particularly unfortunate that several independent distal sources satisfied the harmonic relation just by chance. On the contrary, a harmonic relation is the obligatory correlate of any periodic sound. Accordingly, when we hear harmonic series, perceptual awareness is overwhelmingly that of a single source and not of a disparate collection of pure tones. However, there are many instances of natural sources which do not produce fully periodic sounds and hence exact harmonic series (piano strings for instance). So, the harmonicity cue must include some tolerance (Moore et al., 1986). There are many other cues to vertical organization, each having a strong or weak effect on the perceptual outcome: location (Darwin, 2008), onset synchrony (Hukin and Darwin, 1995), spectral regularity (Roberts and Bailey, 1996), to cite a few. Importantly, just as is the case for harmonicity, none of the cue is infaillible and all are potentially corrupted by noise.

The other sub-problem of ASA is termed, predictably, horizontal organization. It refers to the horizontal time axis of **Figure 3C**. Acoustic sources tend to extend over time, and sound sources do not necessarily produce energy in a continuous fashion. It seems nevertheless advantageous to consider a series of footsteps as a single source, and not to reset the perceptual organization of the scene for each step. For horizontal organization, a putative distal source is also called a "stream." Musical melodies are a prime example of streams.

The cues to horizontal organization, again, seem to follow the plausibility principle. Because of the physics of sound production, an acoustic source will tend to be slowly varying over time. It is unlikely that two consecutive sounds produced by the same source, such as a single talker, will display in rapid succession wide differences in pitch, timbre, or location. Streams will thus favor the grouping together of sounds that are perceptually similar, and segregate sounds which are perceptually dissimilar. Any similarity cue seems to be able to subserve streaming (Moore and Gockel, 2002). Just like for vertical organization, the precise degree of dissimilarity that can be tolerated within a single stream seem to be highly variable, but more on that in Section "Visual Bistability."

Recently, what seems to be a radically different account of ASA has been proposed (Elhilali et al., 2009; Shamma et al., 2011). It suggests that there is one simple and general principle that could govern the formation of auditory streams. The general idea is that sound is analyzed through a multitude of parallel neural channels, each expressing various attributes of sound (periodicity, spatial location, temporal and spectral modulations, etc.). The problem of ASA is then to bind a sub-set of those channels together, with the aim that all channels dominated by a given acoustic source will be bound together and, if possible, not bound with channels dominated by other sources. The suggested principle is *temporal coherence* between channels (as measured by correlation over relatively long time windows). Coherent channels are grouped as a single stream, whereas low coherence indicates more than one stream.

In spite of many differences between these two frameworks, we would argue that there is a core connection between them, when one considers the need for perceptual inference for ASA. This is explicit in Bregman's case, as organization cues are based on the usual behavior of sound sources (even though the neural implementation of each heuristic is not always specified). In contrast, the coherence model does not seem to be easily construed as an inference model: it does not explicitly store knowledge about the external world, not does it include a "decision" stage. However, coherence is a direct and simple way to embody neurally a plausibility principle. Indeed, a single source will tend to induce coherent changes in all channels, irrespective of the nature of the channel. Moreover, these changes will be incoherent with those of other sources. Thus, more often than not, binding coherent channels will lead to isolating single acoustic sources. Note that coherence will never be a perfect trick, either: as soon as there is noise or more than one source in a given channel, choices need to be made among the likely candidates for binding.

This brief account of current issues in ASA is obviously oversimplified. In particular, we have not mentioned the crucial importance of learning and familiarity on the ability to extract information from a scene (e.g., Bregman, 1990; Bey and McAdams, 2002; Agus et al., 2010; McDermott et al., 2011), the role of attention (Thompson et al., 2011), or the strong multi-modal influences on the formation of perceptual objects (e.g., Suied et al., 2009). There are also many open issues that remain to be solved. However, we would argue that the general picture that emerges is that ASA truly behaves as if it were an inference process relying on a variety of sensory cues. These cues are evaluated from the proximal acoustic wave and concomitant neural activity, but they are also weighted with respect to their physical plausibility by means of a form of embodied knowledge (not necessarily explicit and not necessarily operating in a top-down manner) of some of the laws of the acoustics of sound sources.

BISTABLE ILLUSIONS AS TOOL TO PROBE THE PHENOMENOLOGY OF SCENE ANALYSIS

AMBIGUITY AND ASA

We have mentioned that a wide variety of cues can influence ASA. These cues must somehow be pooled to produce a single perceptual outcome that is able to guide behavior. Often, most cues point toward a highly plausible hypothesis in terms of the

number and nature of the distal sources. However, in many cases, the cues can also provide incomplete or even conflicting evidence. For instance, approaching footsteps can be obfuscated by background noise, but a single decision must be reached as to act or ignore. In fact, because of the very nature of ASA as an ill-posed problem, it can be argued that, fundamentally, the system cannot be fully certain of the distal information so there is *always* some degree of ambiguity to be resolved.

VISUAL BISTABILITY

This is where perceptual illusions based on ambiguity enter the picture. The examples of **Figure 4** illustrate what is termed bistable perception in vision: an unchanging stimulus presented for a certain amount of time evokes spontaneous perceptual alternations in the mind of the observer. As the physical description of the stimulus does not match its subjective experience, bistable perception can rightly be termed an illusion. A variety of bistable illusions have been described by visual scientists. Reversible figures such as **Figure 4** are bistable (Long and Toppino, 2004). Binocular rivalry, where incompatible images are presented to the each of the two eyes, also produce alternations between images (Helmholtz, 1866/1925; Alais and Blake, 2005). Finally, there are bistable motion stimuli such as moving plaids (Hupé and Rubin, 2003)[2].

These illusions are seemingly very diverse, but they all have two important things in common. First, they present the visual system with a profoundly ambiguous situation. The information that reaches the retina for **Figure 4** may well have been caused by two faces looking at each other, or, alternately by a vase. Second, it seems that confronted with such an insoluble dilemma, the perceptual system's response is to explore in turn the different possible interpretations (and not to consider an "average" interpretation, as two faces and a vase which contours match exactly is a highly unlikely situation). This is not an obvious outcome: it may well have been possible to imagine that the two alternative interpretations would have been simultaneously available to awareness, but apparently this is not the case. A moment-by-moment decision seems unavoidable.

Recent investigations of visual scene analysis have made extensive use of bistability illusions (for reviews, Leopold and Logothetis, 1999; Long and Toppino, 2004; Sterzer et al., 2009). This enduring interest is perhaps because bistability highlights fundamental processes involved in perceptual organization. As we argued for ASA, sensory scenes contain by necessity some degree of ambiguity. The problem of "inverse optics," just as "inverse acoustics," is ill-posed. Our perceptual systems constantly operate in this inference regime, but we are generally not aware of it because, fortunately, one highly plausible interpretation usually trumps all the others. That this interpretation mostly corresponds to reality is an impressive sign of the sophistication of perception, and not of the simplicity of the problem (as attempts at artificial vision and audition remind us). With this in mind, bistability can be seen as a clever trick by neuroscientists to study the general inference processes that are always at work in perceptual organization.

[2]Demonstrations for the plaid stimulus are available at: http://audition.ens.fr/sup/.

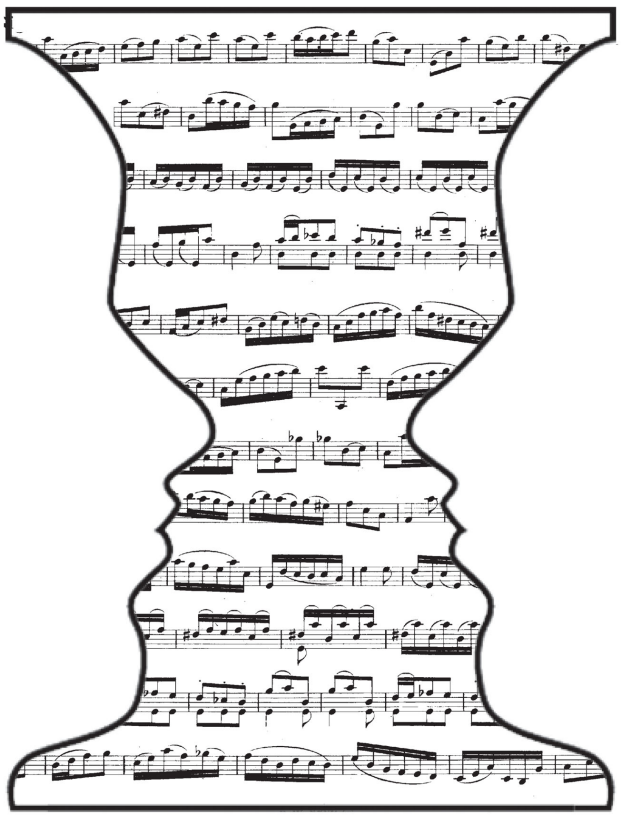

FIGURE 4 | Illustration of visual bistability. This ambiguous picture may be seen as either two faces looking at each other, or as a single vase (*original illustration: J. M. Hupé*). When looking at the picture for a prolonged period of time, the two interpretations may alternate in the mind of the observer. Another example of such bistable alternations can be obtained with ambiguous motion stimuli. For a demonstration, see for instance: http://audition.ens.fr/sup/

As an aside, it is interesting to consider the kind of neural models that have been proposed for visual bistability. Whereas some studies implicate higher brain regions such as pre-frontal cortex, which are most naturally associated with decision and inference (Sterzer and Kleinschmidt, 2007), there are also formalisms based on low-level competition between incompatible percepts (Lankheet, 2006) or non-linear neural dynamics (Kelso, in press) that achieve the same phenomenology. This highlights the fact that the "decision processes" we refer to here may take many different forms when implemented with neurons, some of them bottom-up and largely automatic.

AUDITORY BISTABILITY

The history of auditory bistability is much more modest than that of visual bistability, but recent studies suggest that it may also provide a useful experimental probe for ASA. A surprisingly simple paradigm already reveals the existence of auditory bistability: in its various forms, the "streaming paradigm" uses only two pure tones of different frequencies, arranged in repeating patterns (**Figure 5**). Depending on the frequency and time difference between the tones, listeners report either grouping all

tones together in a single melody (a single stream) or splitting the sequence in two concurrent melodies (two streams). Early on, it was noticed that perception of one or two streams could change across stimulus presentations for a same listener and even within a single presentation (van Noorden, 1975; Bregman, 1978). It had thus been remarked that streaming presents a "striking parallel" with apparent motion, a visual stimulus that is bistable (Bregman, 1990, p. 21). However, the changes in percept were usually assumed to be under the volitional control of the listener (van Noorden, 1975) and were not until recently subjected to the range of experimental and theoretical tools applied to visual bistability.

In fact, auditory streaming is a perfectly fine instance of bistability, as shown by a formal comparison between the perception of ambiguous stimuli in audition and vision (Pressnitzer and Hupé, 2006). In this study, the auditory stimulus was a streaming sequence (van Noorden, 1975; **Figure 5**), and the visual one consisted of bistable moving plaids (Hupé and Rubin, 2003; **Figure 4**). A common point between the two is that they can be perceptually grouped as a single object (a stream or a plaid), or split in two different objects (two streams or two sets of lines). Also, the "correct" organization is ambiguous. The dynamics of percept alternations in auditory streaming were found to display all of the characteristics that define visual bistability (e.g., Leopold and Logothetis, 1999). Percepts were mutually exclusive, that is, subjects reported successively one or two streams but rarely an intermediate percept between the two. The percept durations were random and followed a log-normal statistical distribution. Finally, the effect of volition was limited and highly similar between modalities: when instructed to try and maintain one perceptual interpretation in mind, observers were unable to lengthen the average duration of the target interpretation; rather, they were only able to shorten the duration of the unwanted interpretation. Other authors have independently strengthened the case for auditory streaming as a bistable phenomenon (Denham and Winkler, 2006; Kashino et al., 2007). Interestingly, in those studies, bistability for streaming seemed to be the rule rather than the exception as it could be observed over a surprisingly broad range of stimulus parameters.

An apparently unrelated example of auditory bistability can be found in a paradigm termed verbal transformations (Warren and Gregory, 1958). Listeners were presented with a rapid sequence of repeated verbal material, typically syllables or words (e.g., "life life life"). After prolonged exposure, most listeners reported subjective alternations between the original material and some transformed speech forms (e.g., switches between "life life life" and "fly fly fly"). Warren and Gregory (1958) argued that verbal transformations were the auditory analog of reversible figures in vision. Recently, Sato et al. (2007) and Kondo and Kashino (2007) confirmed that the dynamics of switches between speech forms were similar to other examples of bistable stimuli.

These examples suggest that, despite large acoustic differences, many of the stimuli used to study ASA may share a common point. When in the right regime, the decision processes involved in ASA are revealed by bistable perceptual switches, which display strikingly similar characteristics across all stimuli.

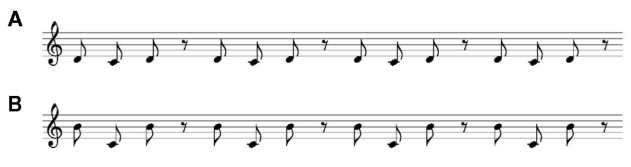

FIGURE 5 | The auditory streaming paradigm. (A) Two different notes (two tones at different frequencies) are presented in a repeated high–low–high (HLH) pattern. Here, the frequency difference between the note is relatively small (two semitones), so when played at a reasonable tempo, the perception is usually of a single melody with a galloping rhythm (HLH–HLH–HLH...). **(B)** Same as A, but with a larger frequency difference (11 semitones). In this case, many listeners experience the perception of two concurrent melodies (H–H–H–... and –L–L–L...) that can be attended individually, but not simultaneously. The sound demonstrations are also available at http://audition.ens.fr/sup/

BISTABLE ILLUSIONS AS TOOL TO PROBE THE NEURAL BASES OF ASA

A major interest of the bistability paradigm for neurophysiologists is that it dissociates the subjective report of the observer from the external stimulus. If a neural correlate of the changes in perceptual reports can be found, then it cannot be confounded by some passive propagation of the stimulus statistics (as those are unchanging). Rather, the correlate must be related to a brain network involved in creating the percept that reached awareness (Tong et al., 2006).

Auditory bistability is being used to investigate the neural correlates of ASA, and in particular the neural correlates of streaming. Overall, results indicate that neural correlates of bistable percepts during streaming may be found at many levels of the auditory pathways. Gutschalk et al. (2005) for instance played a long-duration streaming sequence and asked his listeners to report continuously on their perception of one or two streams. Magneto-encephalography (MEG) revealed that the event-related fields evoked by the tones in the sequence differed if the subjective perception was that of one or two streams, for the same physical stimulus. Localization of the source of the fields suggested that the effect originated from secondary auditory cortex. Cusack (2005) used the same auditory bistability paradigm with an fMRI technique. He observed differential activity for one or two streams in the intra-parietal sulcus, a locus beyond the main auditory pathways associated among other things with cross-modal processing. Using a similar paradigm but focusing on the moment of the perceptual switches, Kondo and Kashino (2009) found switch-related activations in primary auditory cortex, but also in an earlier processing stage, the auditory thalamus. Schadwinkel and Gutschalk (2011) investigated streaming based on subjective location differences and, together with cortical activation, found a correlate of bistable switches in the auditory midbrain (inferior colliculus). Single-unit recordings for streaming based on frequency, in the bistable regime, suggest correlates in primary auditory cortex (Micheyl et al., 2005) but also as early as the cochlear nucleus, the first synapse after the auditory nerve (Pressnitzer et al., 2008). Because of the nature of the technique, these last two studies fall short of co-registration of bistable percept with neural activity, but they do show that the average temporal dynamics of changes in percepts due to bistability is found at the earliest stages of the auditory hierarchy.

When auditory bistability is based on verbal transformation, yet other types of correlates have been found, this time involving fronto-parietal networks implicated in speech (Kondo and Kashino, 2007; Basirat et al., 2008). A recent study has directly compared the two types of auditory bistability in the same subjects (Kashino and Kondo, in press). It confirmed that, even though the auditory bistability networks overlap to some extent, notably for thalamic and primary cortical areas, they also differ to reflect the nature of the competition (speech forms versus tone frequency and rhythm).

This brief overview shows that a bewildering array of neural correlates has been claimed for auditory bistability, encompassing many levels of the auditory pathways. This absence of a single locus is reminiscent of the current view of visual bistability (Tong et al., 2006; Sterzer et al., 2009). It could be that technical details account for the differences between studies. However, it could also be that the bistability processes for ASA are indeed applicable to many levels of processing and hence well-suited to a distributed neural implementation: ASA is such an essential function for hearing that its basic mechanisms seem to be pervasive throughout the auditory system. In any case, it seems that bistable illusions are now firmly part of the experimental assay for the investigation of ASA.

SOME MUSICAL ILLUSTRATIONS

After this brief review of ASA and the use of ambiguity illusions by neuroscientists, it is now time to turn back to music. Before delving into specific examples, a few general points should be made. We have already suggested that musical auditory scenes have the potential to be the most complicated acoustic mixtures encountered by human listeners, because of the sheer number of different acoustical sources involved. We have then mentioned a few of the cues that are considered reliable for ASA, based on physical plausibility. For vertical organization for instance, tones that are synchronous and that satisfy a harmonic relation are highly unlikely to come from different sources, as the likelihood of such a chance combination is really small. But not so in music: in fact, if the composer so decides, and provided the performers are skilled enough, it is well possible to have a collection of different sources playing in synchrony and following harmonic ratios (a choir, for instance). For horizontal organization, it is highly unlikely that a single source widely and rapidly changes its pitch and timbre[3]. But not so in music: musical instruments covering a broad pitch range (e.g., the piano) or even the voice (think human beat-box) can be used to such effect.

What happens then to the subjective experience of the listener? The reasonable assumption is that the general rules of ASA described above still apply, but that the outcome of perceptual inference may or may not reflect the physical description of the scene. The musician can attempt to facilitate the emergence of one or more distinct melodic lines from an otherwise complicated acoustic mixture, or on the contrary to promote the illusory fusion of many sources into one perceptual object. This could be described as steering the inherent ambiguity of ASA toward one of several possible perceptual interpretations. Interestingly, when enjoying music, the listener may be especially willing to entertain different solutions to the ASA problem as there is no obviously harmful potential consequence to making a mistake (as opposed to failing to detect footsteps in the savannah).

Let us now survey what may feel like a haphazard collection of musical examples, borrowed from different genres and historical periods. The eclecticism is intentional, and it is in fact only limited by space constraints and by the music collection of the different authors. It is our hypothesis that similar examples would be found in many musical traditions, including of course non-Western ones.

THE ART OF VOICE-LEADING

A lot of music around the world, starting from what is arguably the most valuable kind of all, lullabies, involve a single acoustic source. However, perhaps because of the social value of music (McDermott and Hauser, 2005; Fitch, 2006), there are also many examples across cultures of musical ensembles involving more than one source. Musicians may then wish to avoid the acoustic muddle that would result from a random superposition of sound sources and try to simultaneously express several distinct melodic lines. This is what is termed polyphony. In Western music, it has been conceptualized through numerous treatises, providing various kinds of advice on how to best achieve "voice-leading."

A particularly fascinating example of voice-leading is to be found in the Musical Offering from J. S. Bach (**Figure 6**). The circumstances of the composing of this piece are worth repeating. The title refers to a single melodic line that the emperor Friedrick II of Prussia presented to Bach, perhaps as a challenge to his composing skills. The theme can be seen and heard at the beginning of the example of **Figure 6** and Sound Example S2 in Supplementary Material. From this royal "offering," Bach was reportedly able to improvise on the spot a polyphonic canon involving several voices. Later on, he returned a score containing several variations on the theme, including the *tour de force* that is the "Ricercar, a 6." In parts of this later canon, six different melodic lines are present.

To help the listener distinguish between the voices, the writing takes advantage of many of the rules of ASA (Huron, 2001). For instance, synchronous harmonic intervals are carefully avoided to avoid fusion between voices, while the pitch steps within a voice are relatively small to promote stream formation. These are two of the most potent cues to vertical and horizontal grouping, as we have seen above. Many, more subtle rules of ASA also seem to be enforced in the music of Bach, as discussed in Huron (2001).

In addition, and perhaps revealingly, some of the canons of the Musical Offering are known as "ambiguous canons," bearing the religious inscription "Quaerendo invenietis" (Seek and you shall find). We may speculate on another meaning of this inscription. Indeed, as we have seen, ASA is per nature ambiguous, and especially so for such complex architectures as those imagined by Bach. The listener is thus left to his own devices to solve the perceptual riddles contained in the music. In the twentieth century, Anton Webern paid tribute to this masterpiece of controlled ambiguity by orchestrating it (Sound Example S3 in Supplementary Material). By assigning different timbres to parts of the canon, he suggests to the listener his own reading of the intricate polyphony, as there

[3]Timbre is notoriously difficult to define. Here, it is meant as "the timbre of a given sound source," including the co-variations in spectral shape and temporal envelope that accompany changes in pitch for most musical instruments.

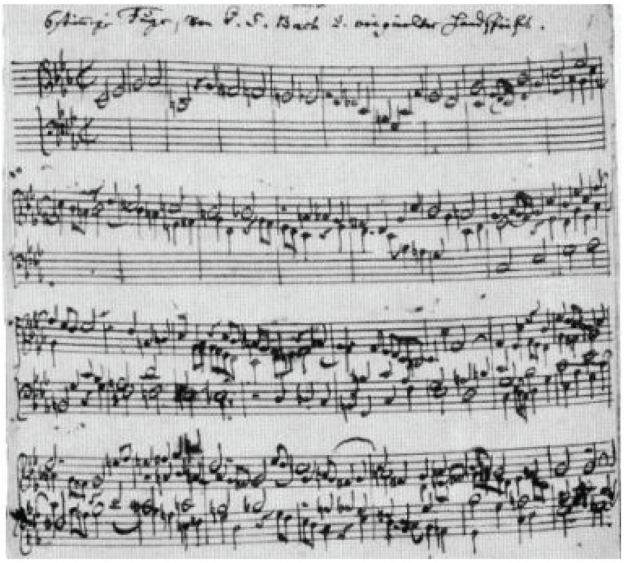

FIGURE 6 | The musical offering, J. S. Bach (ca 1747). Ricercar, a 6. In this score in Bach's handwriting, the Royal Theme that was "offered" to Bach by Frederick II of Prussia can be seen at the top of the page. It is a monodic melody, with a single voice. After a few bars, however, additional melodic lines can be seen to appear. Voice-leading becomes increasingly intricate as the music progresses (see also Sound Examples S3 and S4 in Supplementary Material).

were no indications of instrumentation on the original score. In his own words: "My orchestration aims at uncovering the relations between motifs. [. . .] Is it not about awakening what is still asleep, hidden, in this abstract presentation that Bach gave and which, because of that, did not exist yet for many people, or at least was completely unintelligible?" (Letter to Hermann Scherchen, own translation).

THE ART OF FUSION

With more than one instrument, it is also possible to aim at the opposite effect: blending all the different sources into a coherent ensemble where they eventually become indistinguishable. Early church music (plain-chant) for instance aimed at fusing all singers into a single melodic line. Later on, fusion between different instruments became the realm of orchestration. Any work written for a symphonic orchestra provides examples of complex sonorities achieved by the perceptual fusion of a mixture of acoustic instruments. String quartets are subtler examples: a talented quartet may seemingly oscillate between perfect osmosis between the parts and clearly distinct melodic lines. The illustration we chose in Sound Example S4 in Supplementary Material is taken from the work of Gil Evans, who took to a particularly elegant level the art of "arranging" the instruments of a jazz big-band into a rich orchestral palette. In this example, the compound timbre of the orchestra converses with the soloist in a natural fashion. However, it would be perfectly impossible for the listener to describe the exact combination of instruments that is making up the orchestral "chimera" (Bregman, 1990).

ILLUSIONS AS MUSICAL DEVICES

Using the rules of ASA to promote fusion across instruments or, on the contrary, to create distinct voices may be described as implicitly relying on auditory illusions (not all instruments may be heard, and, conversely, not all melodies are produced by a single physical instrument). There are also composers who have made explicit use of illusions as a structural principle for their writing (Risset, 1996; Féron, 2006). Composers known as proponents of "spectral music" built a whole method from the ASA paradox of breaking down the spectral content of natural sounds, which are usually perceived as single sources, to then write complex chords heard as orchestral timbres, thus fusing instruments that are usually heard individually (Grisey and Fineberg, 2000; Pressnitzer and McAdams, 2000).

But the work of Gyorgy Ligeti in particular bears the mark of perceptual illusions as musical devices in their own right. Take for instance the two pieces illustrated in Sound Example S5 and S6 in Supplementary Material. In the case of "Lontano," many instruments are fused into a slowly evolving texture and it is extremely difficult to isolate any one of them. In Ligeti's own word, "Polyphony is written but one hears harmony". The same orchestral configuration is used for the "San Francisco Polyphony," but here, the various instruments are individually heard, with indeed a feeling of a rich polyphony. Through these two pieces, most of the rules of ASA are used to create dramatically different perceptual outcomes with a same orchestra. This use of auditory illusions was a fully planned and deliberate musical esthetics, as stated by Ligeti himself (Sabbe, 1979): "Yes, it is true, I often work with acoustical illusions, very analogous to optical illusions, false perspectives, etc. We are not very familiar with acoustical illusions. But they are very analogous and one can make very interesting things in this domain."

BISTABILITY IN MUSIC

As it turns out, the bistability illusions that neuroscientists have only recently started to use appear almost *verbatim* in some musical pieces. The bistability of the streaming paradigm is the basis of what has been termed pseudo-polyphony or implied polyphony. Implied polyphony consists at leading several concurrent voices with a single instrument. Among the numerous possible examples, here again it is easiest to refer to the work of Bach (Davis, 2006). In his "Suites for solo Cello," the music played by the largely monodic instrument incorporates interleaved musical lines. The segregation of successive notes into two or more concurrent melodies is achieved by relying on the usual cues to streaming, and most notably as there is a single timbre, on the frequency and time intervals between notes (**Figure 6**, Sound Example S7 in Supplementary Material). Here it is interesting to note that the example of **Figure 7** incorporates a broad range of frequency intervals, starting from one that should promote grouping and ending on one that should promote streaming. In the middle is the ambiguous range where the listener may explore various organizations.

Finally, the verbal transformation paradigm bears strong resemblance with the work of minimalist composers such as Steve Reich or Philipp Glass, where a few musical elements are often recycled and re-used until the perception of the listener subtly changes. The link with verbal transformation is particularly obvious for the

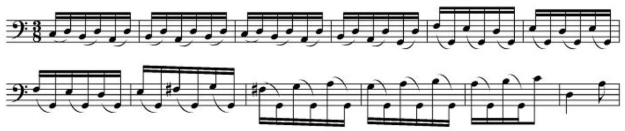

FIGURE 7 | J. S. Bach: Gigue from Cello Suite No. 3 in C Major, mm. 21–32. This part of the Cello suite bears an uncanny resemblance to the stimuli used by neuroscientists to study auditory streaming. In a musical context, the aim seems to create the illusion of polyphony with a largely monodic instrument. Reprinted from Davis (2006).

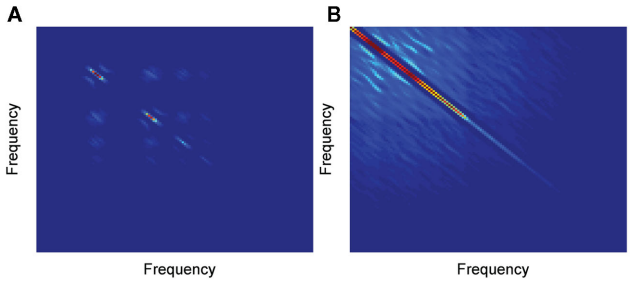

FIGURE 8 | An illustration of the coherence model of ASA applied to music. The coherence matrices (see text, A Coherence Model of ASA Applied to Musical Analysis) are displayed for excerpts of two pieces by György Ligeti, Lontano [**(A)**, Sound Example S5 in Supplementary Material] and the San Francisco Polyphony [**(B)**, Sound Example S6 in Supplementary Material]. Coherence has been averaged for the first 20s of each piece, and the displays are organized by frequency channels. As the qualitative difference between panels suggest, the perceptual organization experienced in the two cases is radically different (even though the composition of the orchestra stays basically the same).

piece "It's gonna rain" by Steve Reich, which consists largely of a tape loop of this single utterance. The perceptual effect is much richer than suggested by this factual description, because of the multistable alternations between speech forms that emerge during listening. Verbal transformations have also found their way into popular music, as illustrated by Sound Example S8 in Supplementary Material where Carl Craig uses the device to create a tense and shifting atmosphere preparing the appearance of an unambiguous beat.

A COHERENCE MODEL OF ASA APPLIED TO MUSICAL ANALYSIS

To close the loop between the neuroscience of ASA and music, we now apply a recent computational model of ASA to the analysis of some of the sound examples discussed above. The model is that of coherence (see Subsection "Cues to ASA"). Its implementation details are available elsewhere (Elhilali et al., 2009). Briefly, the computational architecture is inspired by the known neurophysiology of the auditory system. It postulates that the auditory system first decomposes the sound into different frequency bands, as occurs in the cochlea. Subsequently, these bands are used to construct parallel channels estimating elementary spectro-temporal attributes, which are combinations of temporal and spectral modulations (Chi et al., 2005). Finally, a pair-wise correlation of all attributes is performed to obtain what is termed the coherence matrix.

The coherence matrices for Sound Examples S5 and S6 in Supplementary Material, Ligeti's pieces of Section "Illusions as Musical Devices," are illustrated in **Figure 8** and in the Movies S1 and S2 in Supplementary Material. Each cell in the matrix represents a pair-wise correlation, with warm colors indicating non-zero coefficients. The diagonal is the correlation of a channel with itself, thus it indicates the overall energy of the sound. The off-diagonal elements are what matters for ASA: as explained earlier, a single source tends to have all its attributes temporally modulated in unison. Consequently, when these attributes channels are pair-wise correlated, the resulting patterns in the coherence matrix appear highly regular and sparse. When many incoherent attributes are driving the channels, as would be the case for many independent sources, the coherence matrix has weak and diffuse off-diagonal activation.

The coherence matrices predicted by the model are strikingly different for Lontano (**Figure 8A**) and the San Francisco Polyphony (**Figure 8B**). For Lontano, an exceptional degree of coherence is observed and the matrix is sparse and ordered. This is because numerous instruments play the same temporal rhythm and slow melodic progression, although with different pitches and timbres. For the aptly named San Francisco Polyphony, it is the exact opposite, as many instruments play independent melodies with independent rhythms. This is to put in parallel with the different impressions conveyed by these two pieces, with the first sounding like a single rich harmony while the latter like many voices that never coalesce.

The potential usefulness of these displays stems from their ability to illustrate how complex sound scenes would tend to be perceived by listeners (Shamma et al., 2011). Specifically, when the coherence pattern are highly ordered, it implies that attending to any one attribute points to all others and binds them together (mathematically analogous to the feature reduction attained by principal component analysis of the matrix). By contrast, when many channels are uncorrelated, attending to one attribute does not link it to any other, and hence all voices remain independent. Thus, the range of possible perceptual organizations is somewhat constrained by the form of the coherence matrix.

It is probably too early to suggest a quantitative use of such models in musicological analysis. In particular, it is not easy in the general case to reduce the matrix to a single measure that estimates the number of perceived sound sources. This is perhaps related to the fact that, because of ASA ambiguity and bistability, there is usually no single perceptual answer. However, the qualitative analysis presented here seems sufficient to show that Ligeti used to great effect one very potent principle of ASA, coherence, in order to achieve the perceptual "illusions" he was so keenly interested in.

CONCLUDING REMARKS: THE ESTHETIC VALUE OF AMBIGUITY?

From our brief tour of ASA, it seems clear why neuroscientists would be interested in perceptual illusions based on ambiguity: they seem to highlight the ongoing inference processes at work during perceptual organization and thus may serve as useful

probes to investigate the architecture of the system. But why so many musicians, from different styles, have apparently chosen to play with ambiguity as an integral part of their composing devices?

The short answer is of course that we do not know for sure. It is our hypothesis here that a lot of the music available gravitates around a sweet spot including some degree perceptual ambiguity (with counter-examples of blindingly obvious organizations, of course, as is always the case with artistic endeavors, so the hypothesis is to be taken in the statistical sense). Future research, perhaps using computational models such as the one we have outlined, will have to substantiate the claim. In the meantime, a few anecdotal observations are consistent with a role of ambiguity in the appreciation of music. First, perhaps more than any other art forms, it seems that we are incredibly keen to listen over and over again to the exact same piece of music[4]. Why is this so? Repeated listening comes with an enhanced ability to uncover musical streams that may have been missed the first time around. Memories (or schemas, in Bregman's terms) are sure to form through repeated exposure (Agus et al., 2010). They can then help to pull-out streams from complex scenes, so repeating the same piece over and over allows one to explore its inherent ambiguity. Second, music is also one industry where customers are happy to spend good money for purchasing the same work several times, but from different performers. Concert-goers do the same, too. In French, performers are called "interprètes." It reminds us that, from the score, several readings are possible. It is likely that musical interpretation often plays with ASA: highlighting the clarity of the voices, or, on the contrary, seeking fusion between parts.

It could finally be further speculated that there is something deep in the fact that we seem to look for ambiguous auditory scenes when creating and listening to music. Zeki (2004), discussing ambiguity and visual art, pointed out that there is fundamentally no ambiguity without perception. Physical information is just there, ambiguity only arises when a perceiver is trying to "make sense" of this information. Music is an especially challenging stimulus to make sense, as most of it is abstract without any clear reference to an external object. By embedding several latent perceptual organizations into complex acoustical scenes, music may well be able to challenge the listener with a rich set of possibilities that can be freely entertained, with no other potential consequence than being surprised, rejoiced, or moved.

[4]We would like to thank I. Nelken for pointing out this fact quite vividly, by suggesting to ask ourselves how many times we have seen our favorite movie compared to how many times we have listened to our favorite piece of music.

ACKNOWLEDGMENTS

We would like to thank Robert Zatorre for editing the manuscript, as well as the two reviewers for numerous insightful comments. This work was supported by the Agence Nationale de la Recherche (ANR-BLAN-08 Multistap, Daniel Pressnitzer); the Fondation Pierre Gilles de Gennes pour la Recherche (Clara Suied); the Chaire d'Excellence Blaise Pascal and the program Recherche à Paris (Shihab A. Shamma).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http:// www.frontiersin.org/human_neuroscience/10.3389/fnhum.2011.00158/abstract

All sound examples are at http://audition.ens.fr/dp/Frontiers_Sound_Examples/

Sound Example S1 | Gustav Mahler, Symphony no 8 in E flat major, 1907. *Veni, creator spiritus*. Chicago Symphony Orchestra, direction: Sir Georg Solti, 1972. Decca Legends.

Sound Example S2 | Johann Sebastian Bach, The Musical Offering, 1747. *Ricercar, a6*. Musica Antiqua Köln, direction: Reihnard Goebel, 1979. Archiv Production.

Sound Example S3 | Johan Sebastian Bach/Anton Webern, The Musical Offering, 1935. Orchestration of the fugue no 2 from the Musical Offering. London Symphony Orchestra, direction: Pierre Boulez, 1991. Sony Classical.

Sound Example S4 | Miles Davis/Gil Evans, George Gershwin's Porgy and Bess. 1958. *Summertime*. Columbia

Sound Example S5 | György Ligeti, Lontano für großes Orchester, 1967 Sinfonie-Orchester des Südwestfunks, Baden-Baden, direction: Ernest Bour, 1969. Wergo.

Sound Example S6 | György Ligeti, San Francisco Polyphony für Orchester, 1973/74. Sinfonie-Orchester des Schwedischen Rundfunks, direction: Elgar Howarth, 1977. Wergo.

Sound Example S7 | Johan Sebastian Bach, Suites for solo Cello in C major, ca. 1720. *Suite no 3 in G major*. Janos Starker, 1965. Mercury Living Presence.

Sound Example S8 | Carl Craig, More songs about food and revolutionary art, 1997. *Dominas*. Planete SSR.

REFERENCES

Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (2010). Rapid formation of robust auditory memories: insights from noise. *Neuron* 66, 610–618.

Alais, D., and Blake, R. (eds). (2005). *Binocular Rivalry*. Cambridge, MA: MIT Press.

Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352, 1141–1147.

Basirat, A., Sato, M., Schwartz, J. L., Kahane, P., and Lachaux, J. P. (2008). Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *Neuroimage* 42, 404–413.

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Percept. Psychophys.* 64, 844–854.

Bregman, A. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Bregman, A. S. (1978). Auditory streaming is cumulative. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 380–387.

Brochard, R., Drake, C., Botte, M. C., and McAdams, S. (1999). Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1742–1759.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn. Sci. (Regul. Ed.)* 8, 465–471.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979.

Chi, T., Ru, P., and Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* 118, 887–906.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651.

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1011–1021.

Davis, S. (2006). Implied polyphony in the solo string works of j. S. Bach: a case for the perceptual relevance of structural expression. *Music Percept.* 23, 423–446.

Denham, S. L., and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* 100, 154–170.

Doherty, M. J., Campbell, N. M., Tsuji, H., and Phillips, W. A. (2010). The ebbinghaus illusion deceives adults but not young children. *Dev. Sci.* 13, 714–721.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317–329.

Féron, F. X. (2006). *Des illusions auditives aux singularités du son et de la perception: L'impact de la psychoacoustique et des nouvelles technologies sur la création musicale au xxe siècle.* Unpublished PhD thesis, Université Paris IV - Sorbonne, Paris.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352, 1121–1127.

Gregory, R. L. (2005). The Medawar lecture 2001 knowledge for vision: vision for knowledge. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1231–1251.

Grisey, G., and Fineberg, J. (2000). Did you say spectral? *Contemp. Music Rev.* 19, 1–3.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25, 5382–5388.

Helmholtz, H. (1877). *On the Sensations of Tone.* New York: Dover.

Helmholtz, H. (1866/1925). *Treatise on Physiological Optics Southall.* New York: Dover.

Hukin, R. W., and Darwin, C. J. (1995). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Percept. Psychophys.* 57, 191–196.

Hupé, J. M., and Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res.* 43, 531–548.

Huron, D. (2001). Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Percept.* 19, 1–64.

Kashino, M., and Kondo, H. (in press). "Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations," in *Multistability in Perception: Binding Sensory Modalities,* eds J. L. Schwartz, N. Grimault, J. M. Hupé, B. C. J. Moore, and D. Pressnitzer.

Kashino, M., Okada, M., Mizutani, S., Davis, P., and Kondo, H. M. (2007). "The dynamics of auditory streaming: Psychophysics, neuroimaging, and modeling," in *Hearing – From Sensory Processing to Perception,* eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Berlin: Springer), 275–283.

Kelso, J. A. S. (in press). "Multistability and metastability: understanding dynamic coordination in the brain," in *Multistability in Perception: Binding Sensory Modalities,* eds J. L. Schwartz, N. Grimault, J. M. Hupé, B. C. J. Moore, and D. Pressnitzer.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.* 55, 271–304.

Kersten, D., and Yuille, A. (2003). Bayesian models of object perception. *Curr. Opin. Neurobiol.* 13, 150–158.

Kondo, H. M., and Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage* 36, 123–130.

Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* 29, 12695–12701.

Kovacs, I. (2000). Human development of perceptual organization. *Vision Res.* 40, 1301–1310.

Lankheet, M. J. (2006). Unraveling adaptation and mutual inhibition in perceptual rivalry. *J. Vis.* 6, 304–310.

Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci. (Regul. Ed.)* 3, 254–264.

Long, G. M., and Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychol. Bull.* 130, 748–768.

McDermott, J., and Hauser, M. D. (2005). Probing the evolutionary origins of music perception. *Ann. N. Y. Acad. Sci.* 1060, 6–16.

McDermott, J. H., Wrobleski, D., and Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1188–1193.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80, 479–483.

Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica* 88, 320–332.

Murray, S. O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.* 9, 429–434.

O'Regan, J. K., and Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–1011.

O'Regan, J. K., Rensink, R. A., and Clark, J. J. (1999). Change-blindness as a result of "mudsplashes." *Nature* 398, 34–34.

Pickles, J. O. (2008). *An Introduction to the Physiology of Hearing,* 3rd Edn. New York, NY: Academic Press.

Pressnitzer, D., and Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16, 1351–1357.

Pressnitzer, D., and McAdams, S. (2000). Acoustics, psychoacoustics, and spectral music. *Contemp. Music Rev.* 19, 33–59.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.

Purves, D., Wojtach, W. T., and Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 3), 15588–15595.

Risset, J. C. (1996). Real-world sounds and simulacra in my computer music. *Contemp. Music Rev.* 15, 29–47.

Roberts, B., and Bailey, P. J. (1996). Regularity of spectral pattern and its effects on the perceptual fusion of harmonics. *Percept. Psychophys.* 58, 289–299.

Sabbe, H. (1979). Gyorgy ligeti – illusions et allusions. *J. New Music Res.* 8, 11–34.

Sato, M., Basirat, A., and Schwartz, J. L. (2007). Visual contribution to the multistable perception of speech. *Percept. Psychophys.* 69, 1360–1372.

Schadwinkel, S., and Gutschalk, A. (2011). Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J. Neurophysiol.* 105, 1977–1983.

Shamma, S. A. (1985). Speech processing in the auditory system. II: lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* 78, 1622–1632.

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123.

Shamma, S. A., and Micheyl, C. (2010). Behind the scenes of auditory perception. *Curr. Opin. Neurobiol.* 20, 361–366.

Snyder, J. S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* 133, 780–799.

Sterzer, P., and Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 323–328.

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 310–318.

Suied, C., Bonneel, N., and Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Exp. Brain Res.* 194, 91–102.

Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1253–1262.

Tong, F., Meng, M., and Blake, R. (2006). Neural bases of binocular rivalry. *Trends Cogn. Sci. (Regul. Ed.)* 10, 502–511.

van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences.* Ph.D. thesis, University of Technology, Eindhoven.

Warren, R. M., and Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *Am. J. Psychol.* 71, 612–613.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604.

Zeki, S. (2004). The neurology of ambiguity. *Conscious. Cogn.* 13, 173–196.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 June 2011; accepted: 16 November 2011; published online: 14 December 2011.

Citation: Pressnitzer D, Suied C and Shamma SA (2011) Auditory scene analysis: the sweet music of ambiguity. Front. Hum. Neurosci. ***5****:158. doi: 10.3389/fnhum.2011.00158*