Contemporary Music Review
2000, Vol. 19, Part 2, p. 33–59
Photocopying permitted by license only

Acoustics, Psychoacoustics and Spectral Music

Daniel Pressnitzer and Stephen McAdams

Introduction

The aim of this article is to examine the points at which acoustics, psychoacoustics and what has been called *'spectral music'* meet. The motivations behind this bringing together of three more or less barbaric words are to be found in the principle of the spectral approach itself.

The tonal system is governed by a set of harmonic rules that embody a compromise between the will to modulate among keys, the system of symbolically notating music, the available set of instruments, certain laws of acoustics, as well as many other concerns. This imposing edifice was patiently constructed by an accumulation of experience and benefited from a slow, cultural maturation. But the bases of this edifice were shaken in the evolution of contemporary music by recent developments in our relation to sound: previously of a fleeting and evanescent, ungraspable nature, sound has been captured and manipulated by way of recording technology. The theory of signals, associated with the computational power of modern computers, has made it possible to analyze sound, to understand its fine structure, and to fashion it at will. The potential musical universe has thus 'exploded' in a certain sense. Sound synthesis opens truly unheard perspectives, extending the act of composition to the sound material itself. The distinctions between note, frequency, timbre, and harmony become fuzzy, or even irrelevant, and accumulated traditional experience finds itself impotent to organize the emerging sound world.

After such a shock, new means of formalizing and structuring needed to be defined. Rather than establish a series of arbitrary rules, the spectral

intuition consisted in founding compositional systems on the structure of sound, and thus in deriving fields of musical relations from sound itself. The wager of such an approach is to give to a listener reference points that are naturally understandable, while allowing the use of the new potential offered by micro-compositional work at the level of sound. In other words, the structures sought should be latently intelligible, since the elements necessary for their comprehension are contained in the materials. If the work of the forerunners of this approach was founded only on intuition and experimentation, the will to go further, to not let oneself be enclosed by a limited number of effects or gestures, requires a more rigorous conceptualization and formalization of the fundamental ideals. A perfect understanding of acoustic phenomena is thus necessary and we will address this domain by insisting on the importance of different representations of sound. Though it is necessary, this comprehension is not sufficient: what counts in the end is certainly (at least in the logic of the spectral approach) what is perceived and understood by the listener. It is at this level that psychoacoustics, which extends and validates the reflection on purely physical structures, enters the picture.

As such, it is not by a will to 'scientism' at any cost that the spectral composers were undoubtedly drawn to interest themselves in these disciplines, but simply because of a necessity that proceeds from their approach. Numerous questions thus naturally come to the fore: What in fact is a sound? What are its possible representations? What are the interpretations made by perception to extract from a sound what is relevant for the listener? Can we exploit them musically? Can we speak of 'sound objects' in our psychological representations? How do we think music? Doubts can also appear, notably with respect to a fundamental aspect of music: if tonal harmony is considered as a sort of syntax, allowing expression by changes in tension that occur as one deviates from certain rules, it relies undoubtedly on a strong, even implicit, cultural learning. Is it possible, in no longer using this solid, conventional prop, to find a basis contained in the material of sound, for the expression of tension? In an attempt to respond to these questions, and especially to incite new ones, we will present a set of facts concerning sound and its perception, starting with its birth in the acoustic world.

The acoustic world

Representations

A vibrating body creates in the surrounding air the propagation of a pressure wave, in the same manner that the agitation of an object on a

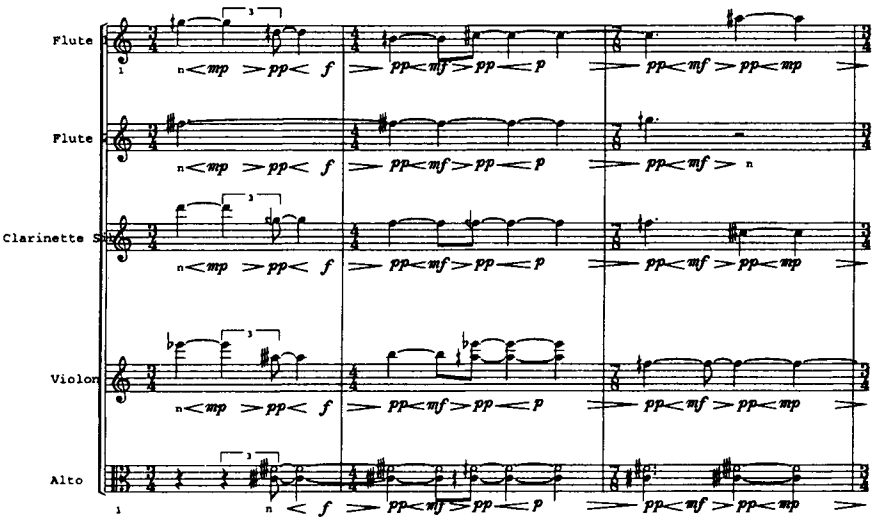

Figure 1

surface of water provokes the propagation of wavelets. This is the physical reality of sound, the variation of acoustic pressure over time, and this reality is unique. It can, nevertheless, be represented in different ways according to the information that one wishes to emphasize. Take the example of a chord from the pieces *Streamlines* by Joshua Fineberg (1995). Its classic musical representation is the score[1] (Fig. 1). Centuries of experience allow the use of this symbolic representation for compositional purposes, but in fact it constitutes more a set of instructions to the performers than a description of the sound actually produced. For example, if the instrumentation changes, the music transcribed in the score is transformed. In Baroque music, where exact instrumentation was often not specified, and in some pieces of contemporary music, where the instrumentation is specified vaguely as in the percussion piece *Ionisation* by Edgard Varèse (1933), the sound structure itself may be completely different from one rendering to the next.

To obtain a trace of a specific sound, it is possible to capture it through a recording device that will convert the pressure variations into something that can be visualized (Fig. 2). This representation of the pressure wave reflects all of the fine-grained temporal evolution (within the resolution limits of the visual representation). It is therefore particularly well-adapted to manipulations of sound such as cutting and splicing, reversal

1. Classic up to a certain point, considering the presence here of quarter-tones!

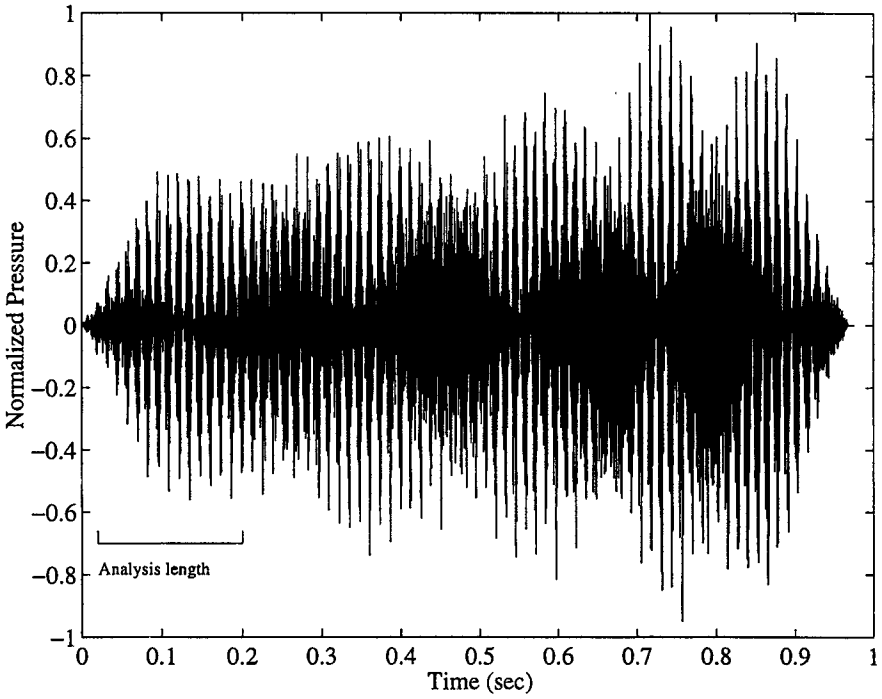

Figure 2

in time or repetition. In its early days, *musique concrète*, because of the available technology, worked with razor blades on tapes containing the exact magnetic retranscription of this temporal wave and thus used a lot of these kinds of transformations.[2] However, this temporal image does not translate in an obvious fashion the different pitches that it is possible to distinguish in listening attentively to such a chord.

A representation that allows this is the one based on Fourier's theory. This theory states that any complex signal can be decomposed into a sum of sinusoidal waves, over an infinite time frame, by specifying precisely their relative amplitudes and phases. It is thus possible to decompose a complex sound into a sum of sine tones, which are called the partials of the sound, the set of which form its spectrum (Fig. 3). This Fourier transform represents well the same chord, but in a different form, unveiling its frequency content. This knowledge of frequency components present in the sound, which may under certain conditions be heard as 'spectral

2. One should note in passing the influence of the tools used on the end result.

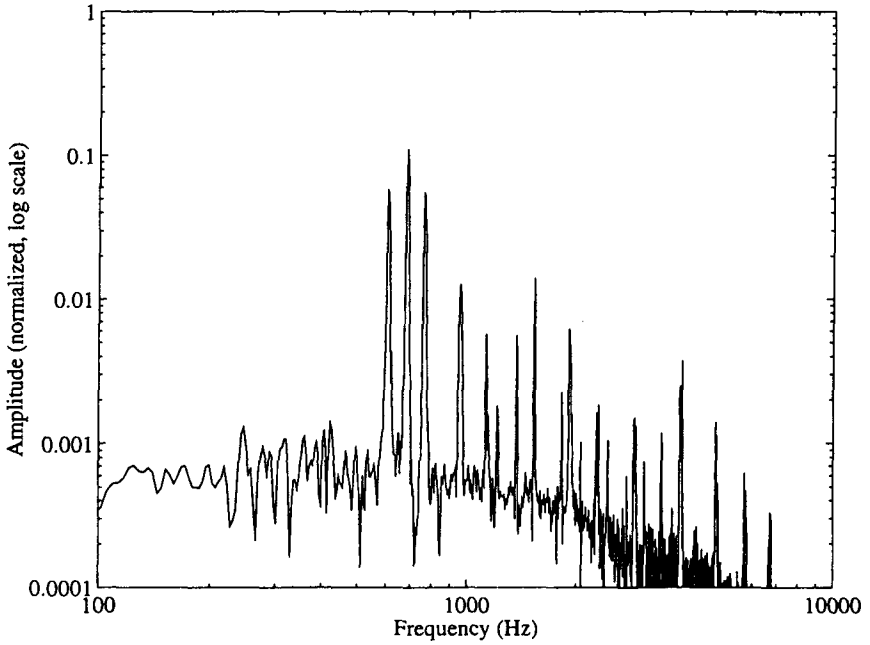

Figure 3

pitches' (Terhardt, 1974), is at the origin of the emblematic idea of the music that is, appropriately, called 'spectral'. As an example among others, Jonathan Harvey's piece *Mortuos plango, vivos voco* (1980) uses the spectrum of a bell sound and its transformations as a foundation for the harmony. If such a representation is a source of fertile inspiration, it makes the temporal information about the sound no longer explicit in the transform as the analysis is (theoretically) over an infinite duration.

The true nature of a sound phenomenon as we perceive it is double: it evolves over time, which is represented by the temporal wave, and it also has a certain frequency content, visible in the spectrum. The short-term Fourier transform reconciles these two types of informations. It is thus called a 'time-frequency representation'. The sound is sliced up, by an analysis time window, into successive instants. The Fourier transform is then computed for successive instants by sliding the window over the temporal waveform bit by bit. In this way the evolution of the frequency content of the sound is represented over time. The result of this analysis can be presented either in the form of a spectrogram (Fig. 4), having graphical similarities to a musical score, or in the form of a three-dimensional perspective plot that expresses the acoustic variations in

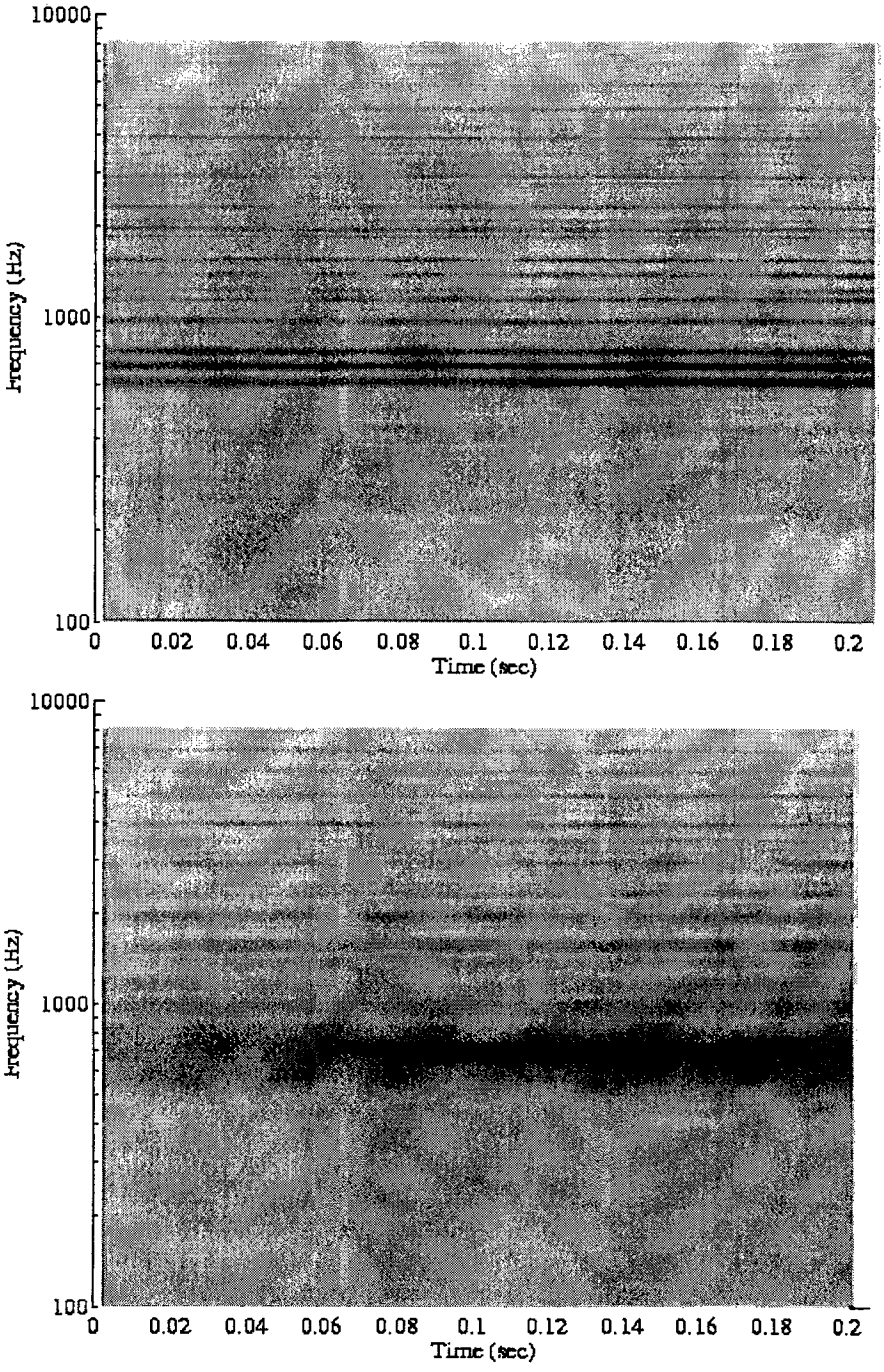

Figure 4

terms of frequency, amplitude, and time.[3] This type of analysis is essential for performing additive synthesis, or its orchestral derivations. In *Partiels* (1975), for example, Gérard Grisey explores the sound of a trombone by assigning to different instruments the production of a given partial of the trombone spectrum analyzed with its dynamic temporal evolution. The representation as a short-term Fourier transform is quite general, but it should be noted that there is an inevitable compromise between the temporal resolution (the duration analyzed) and the frequency resolution (the analysis precision). For a high frequency precision, a long analysis duration is needed, and so there is a loss of temporal precision.

There are other representations that allow an optimization of this compromise (Loughlin, Atlas, and Pitton, 1993), some, such as the wavelet transform, adapting it to the frequencies analyzed (Combes, Grossman, and Thamitchian, 1989). This produces for a given sound various equivalent representations.

The idea of continuum

In the physical world, frequency, time, and intensity are considered as continuous dimensions. Music, on the other hand, has been built on discrete scales of pitch and duration made necessary, among many other reasons, by the desire to notate events and by instrumental playing constraints. The different representations we just presented, associated with sound synthesis, give us access to the physical continua. This allows us, as Varèse noted, to catch a glimpse of an alternative to the 'fil à couper l'octave'.[4] From this point onward, between each degree of the scale there can be a continuous thus infinite world to be discovered and organized.

Another continuum is revealed by these mathematical representations. What difference is there between the spectrum of a note associated with a timbre and the spectrum of a chord considered as an element of harmony? The answer is to be found on the computer screen: at first sight, there isn't any! A simple note is a collection of spectral components, thus a chord; and a chord is a collection of partials, thus a timbre. Sound synthesis allows the organization of the note itself, introducing harmony into timbre, and reciprocally sound analysis can introduce timbre as a generator of harmony. This ambiguity is strikingly demonstrated in Jean-Claude Risset's piece *Mutations* (1969), where the same

3. It should be noted that it is possible to obtain spectrograms in an analogous fashion with a bank of filters. From a theoretical point of view, the two descriptions of the short-term transform (running window and filter bank) are equivalent (Allen and Rabiner, 1977).

4. Literally a 'slicing up of the octave', i.e. the equal-tempered scale.

material is treated alternately as a harmonic chord or a gong-like timbre. If harmony and timbre are so intimately linked, the clear-cut traditional classification of chords between perfect consonances, imperfect consonances and dissonances may become irrelevant. Timbre manipulation opens up the possibility to look for a continuous scale that could repro-

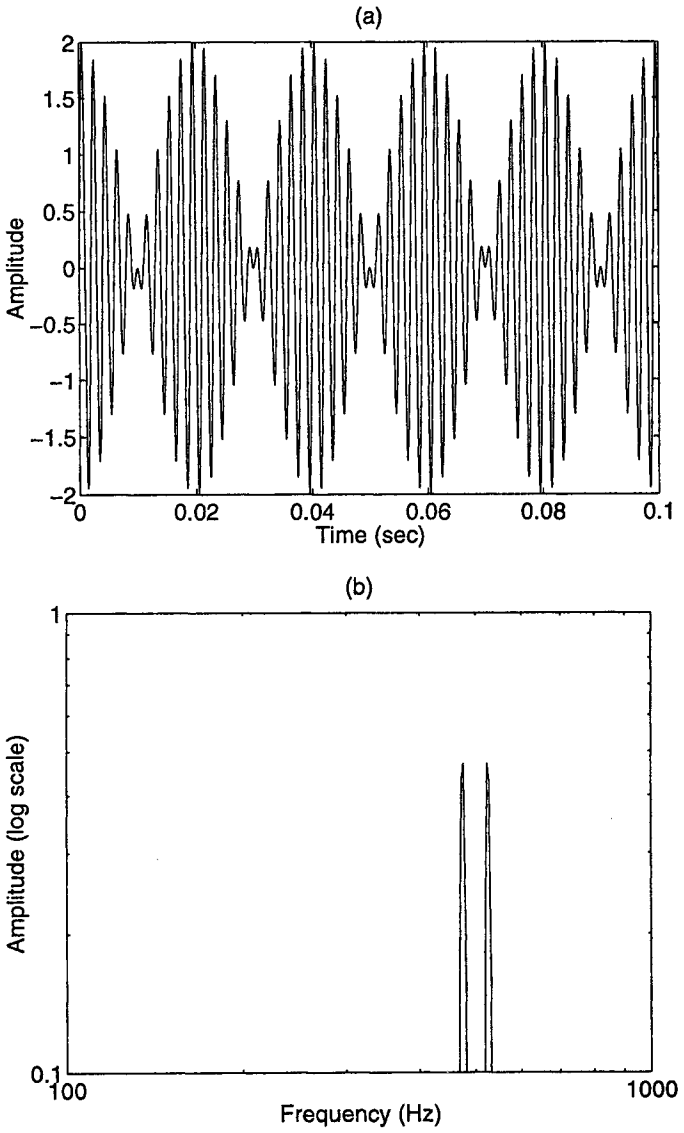

Figure 5

duce, in some respects, the expressive means associated with the tonal notions of consonance and dissonance.[5]

The exploration of such a dimension has been undertaken by many composers, essentially in an intuitive fashion. Tristan Murail, for example, has ordered timbres and aggregates with a measure of inharmonicity (*Désintégrations*, 1982). Kaija Saariaho has defined a sound/noise axis intended to reproduce the harmonic capacity to create tension and relaxation (*Verblendungen*, 1982–84). Joshua Fineberg has adopted a hierarchy founded on the pitch of virtual fundamentals (*Streamlines*, 1995). Is it possible that a single phenomenon is hidden behind these different intuitive criteria? Hermann von Helmholtz proposed an axis of reflection in drawing our attention to the attribute of sound that he called 'roughness' (von Helmholtz, 1877). Two pure tones produced simultaneously and having closely related frequencies create amplitude fluctuations in the waveform that are called 'beats' (Fig. 5). These fluctuations, according to their rate of beating, can give rise to a grainy quality in the sound. An example of this rough quality can be heard, emerging from silence, at the beginning of *Jour, Contre-jour* (Grisey, 1980). Helmholtz thought he saw in roughness the acoustic basis for the dissonance of musical intervals. Western music employs principally instruments with harmonic spectra. The partials of their complex spectra are superimposed when an interval is played, resulting in beats if their frequencies do not coincide perfectly. Intervals with simple frequency ratios, such as the octave or the fifth, do have a significant degree of harmonic coincidence and thus less beating. However, intervals such as the tritone create a situation where harmonics of one note beat with those of the other note (Fig. 6). Coming back to the previous examples, an inharmonic sound can be a source of roughness when superimposed on harmonic sounds; sounds described as noisy can often be rough; a sound with a very low fundamental frequency has partials that are quite close to one another which creates beating. As such, it may be that the same acoustic feature guided the composers mentioned above in the elaboration of their 'harmonic' criteria. If this was to be the case, such a feature could be used to define a new continuum related to the vast and complex notion of musical dissonance, as some kind of an 'acoustic nucleus' for it (Mathews and Pierce, 1980).

One might be overwhelmed before the immense field of possibilities that is thus opened. The mass of data available to the composer that are derived from progressively more precise and sophisticated acoustic

5. In the Western early polyphonic period, the scale of consonance and dissonance contained up to six different degrees. The later simplification of this scale can be paralleled with the progressive affirmation of syntactic tonal rules (Tenney, 1988).

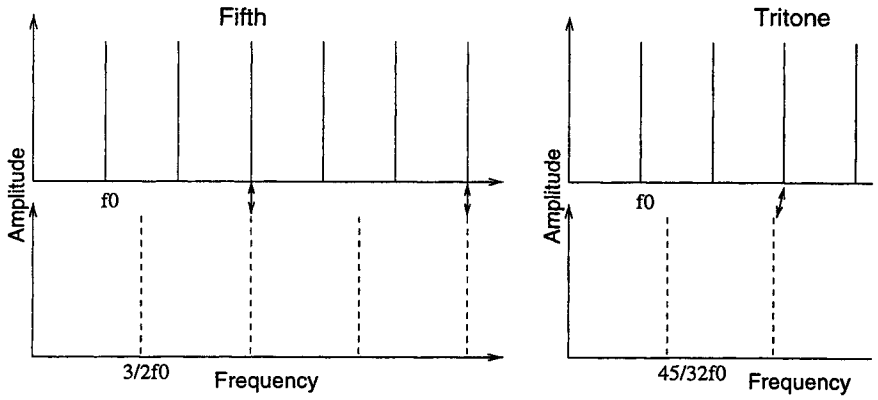

Figure 6

analyses, the abundance of, at times redundant or perceptually irrelevant, masses of numerical data coming out of analysis programs, can mask the salient features of musical sound. For example, when the interval between two pure tones is large enough, they are heard separately and without beats, resulting in no roughness whatever, even though the beats still exist in the acoustic world as can be seen on an oscilloscope. So what happened to them?

Perception

The ear: from without to within

The notion of perception is implicitly contained in the word 'sound'. A sound is not just any kind of variation in acoustic pressure, but a pressure variation that can generally be heard: our ears must be able to code its features. This coding is conditioned by the physiology of the peripheral auditory system. Thus, the very first given of psychoacoustics is the definition of the realm of validity of the word 'sound', in other words the audible field (Fig. 7). An audiogram traces the average hearing threshold: that is, the intensity necessary to just detect a pure tone of a particular frequency. This curve simply translates our capacity to detect a sound signal and is a poor way to characterize the auditory system. The relation between the pressure wave and what we hear of it can only be understood by studying certain physiological mechanisms of perception.

The air vibrations of a sound wave are transmitted and amplified by the external and middle ears: the pinna, the ear canal, the eardrum, the middle ear ossicles, up to the cochlea. These vibrations are communi-

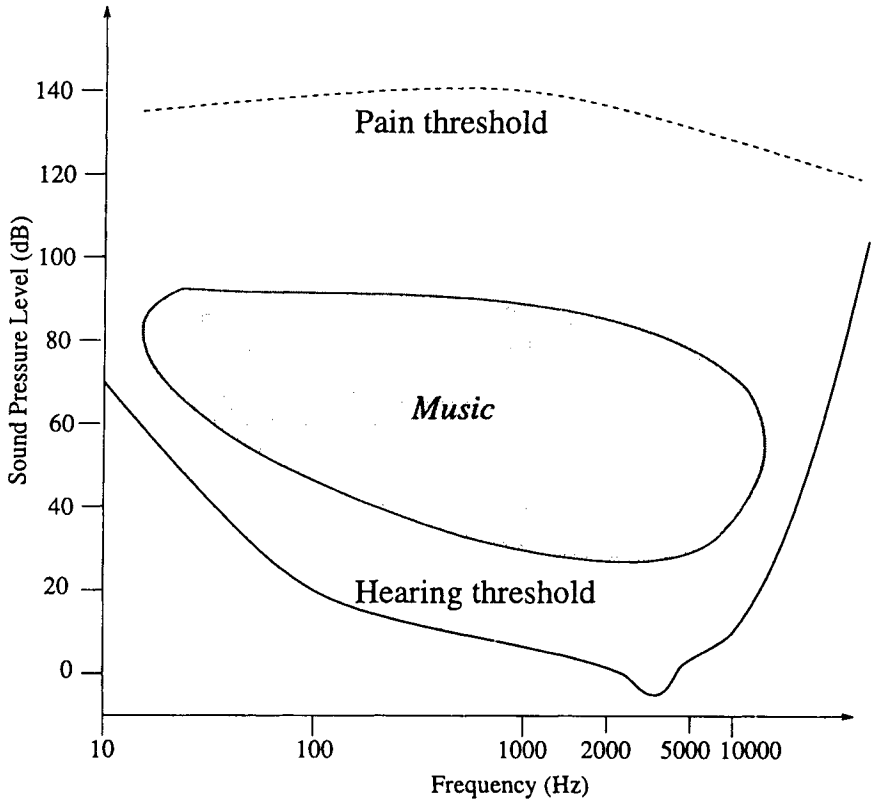

Figure 7

cated in the inner ear to the basilar membrane, and the waves propagate and are damped along this membrane. The stiffness of the membrane varies along its length. Due to this property the higher frequencies create a maximal displacement at its base (near the ossicles), while lower frequencies maximally stimulate the other (apical) end. These deformations then result in electrochemical changes in the hair cells that are arranged along the length of the membrane. These cells in turn stimulate the fibers of the auditory nerve, along which electrochemical impulses are sent toward the brain.

The essential point to understand here is that the first thing that happens to an acoustic signal in the inner ear is some kind of a spectral analysis. In fact, starting from the temporal wave, the basilar membrane spatially decomposes the signal into frequency bands. The second important point is that the hair cells, in addition to coding the frequency position of the signal components, also preserve to a certain degree their

temporal information by producing neural firings at precise moments of the stimulating wave they are responding to. This phenomenon, called phase-locking, decreases with increasing frequency and eventually disappears between about 2000 to 4000 Hz near the upper end of the range of musical pitch. The auditory system thus performs a double coding of the sound, both spectral and temporal, in such a way that all the cues present in both kinds of representation may be available simultaneously in the sensory representation sent to the brain.

More than just a transmission

The coding mechanisms introduce certain phenomena that generate paradoxes and ambiguities. Sound components called difference tones or combination tones are a first example of the necessity to become interested in perception in addition to physical representations. Two pure tones presented simultaneously to the auditory system stimulate the basilar membrane at positions that are associated with their respective frequencies, but also at positions corresponding to frequencies that are the completion, toward lower frequencies, of the harmonic series. The causes and behavior of all these distortion products are not fully understood. However, it is easy to hear the difference tone that corresponds to the simple difference between the frequencies physically presented to the ear. This tone is all the more audible at higher levels. It has even been used compositionally, for example, in György Ligeti's *Zehn Stücke für Bläserquintett* (1968). The phenomenon can be heard at the end of the first piece, for instance. The difference tone belongs to the world of physiology and perception: even if it is not present in the stimulating waveform, it is created physically in the inner ear. It can also create auditory beats in the same way as a 'real' sound since it is mechanically present on the basilar membrane.

The frequency decomposition realized by the basilar membrane is mechanical: the displacements of the membrane are not limited to specific points but are spread out over a portion of it. If two components of a complex signal are close in frequency, these displacement will overlap. There is thus a minimal resolution, called the critical band (Greenwood, 1961), inside of which the ear cannot separate two simultaneous frequencies. The width of this band varies as a function of the center frequency considered (Fig. 8). The mechanisms underlying the critical band play an essential role: two components will not at all have the same perceptual effect if they are resolved, thus heard separately, or if they fall within a single critical band. Let us just make clear that the width of the critical band does not at all represent a frequency band within which all sounds are perceived to have the same pitch. Due to the

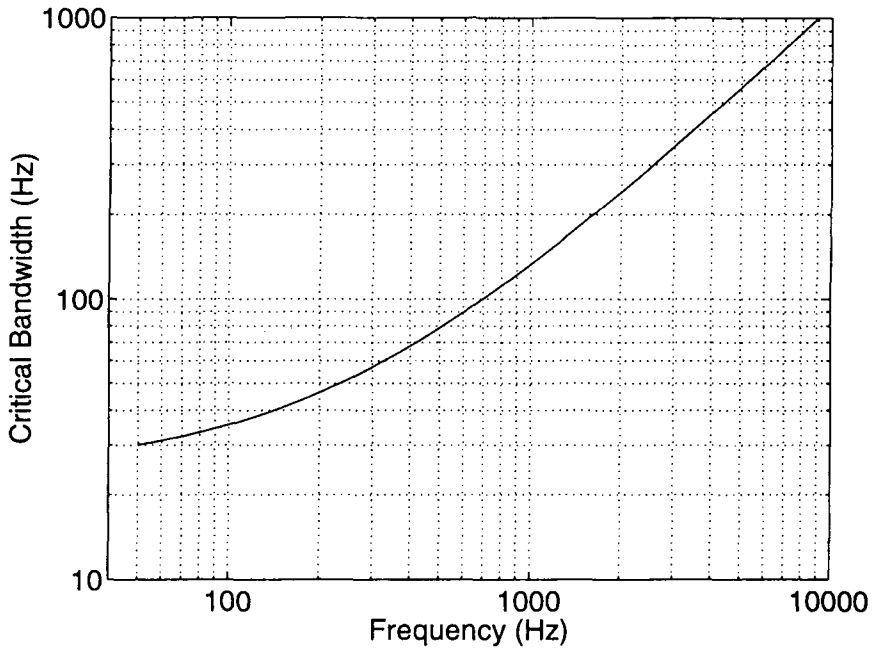

Figure 8

temporal coding, the ear is sensitive to differences in frequency that are less than 1%. The critical band is not a limit in precision, but a limit in selectivity.

This selectivity limit is obvious in the masking phenomenon. Simultaneous masking is related to the overlap of excitation patterns on the basilar membrane and to the amount of activity present in the auditory nerve that represents each sound component present. Simultaneous masking can be conceived as a kind of 'swamping' of the neural activity due to one sound by that of another (usually more intense) sound. For example, high-level components (noise, partial) create a level of activity that overwhelms that created by lower-level components, which are subsequently not perceived at all or are perceived as being of lower level than they would if presented alone. The different components of a complex sound can also interact, mutually masking one another, some having a sensation level that is lower than their actual physical level would lead one to expect. Masking relations are determined largely by the excitation pattern on the basilar membrane. This pattern is actually asymmetrical, extending more to the high-frequency side than to the low-frequency side, and all the more so as the level of the sound

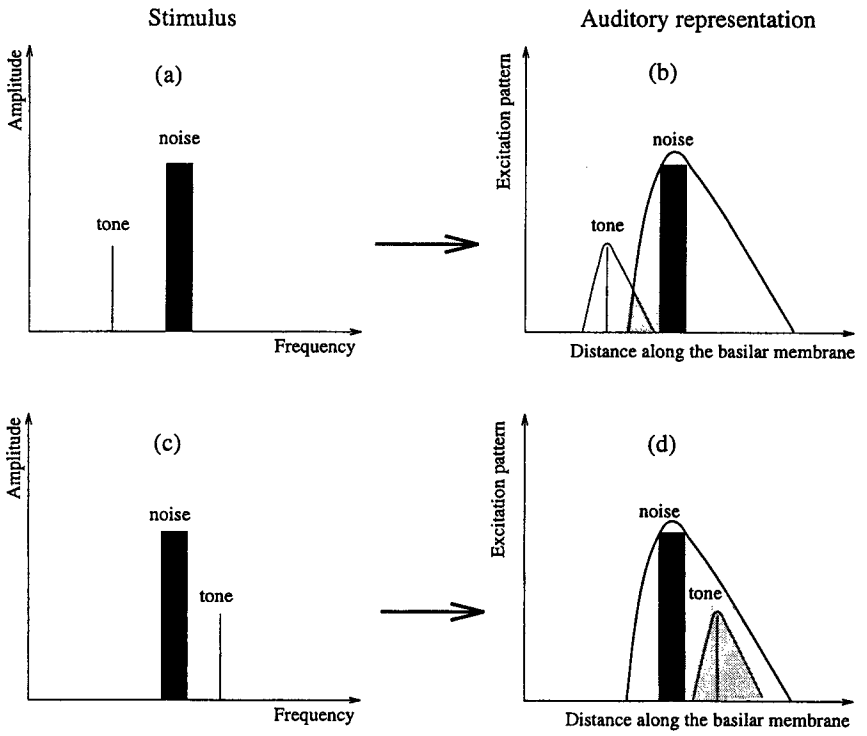

Figure 9

increases. Therefore, the frequency and amplitude relations between sounds will affect their masking relations in a non-trivial way (Fig. 9). The knowledge of these relations is nevertheless essential to understand which part of a musical message will actually be perceived.

The critical band also influences the perception of beats between two tones. Acoustically, the rate of the beats increases with their frequency difference. As such, as the two pure tones are mistuned from unison, we should hear beats that result from their interaction becoming progressively more rapid. This is in fact what happens at the beginning of the separation. But very soon, (after approximately 10 beats per second or a 10-Hz frequency difference) our perception changes from a slow fluctuation in amplitude toward an experience of more and more rapid fluctuations, that produce roughness. Finally, if the separation becomes large with respect to the critical band, the strength of the sensation of beating diminishes, leaving us with the perception of two resolved pure tones. Three very different perceptual regions can therefore arise from the same acoustical stimulus. Let us come back for a minute to our on-going

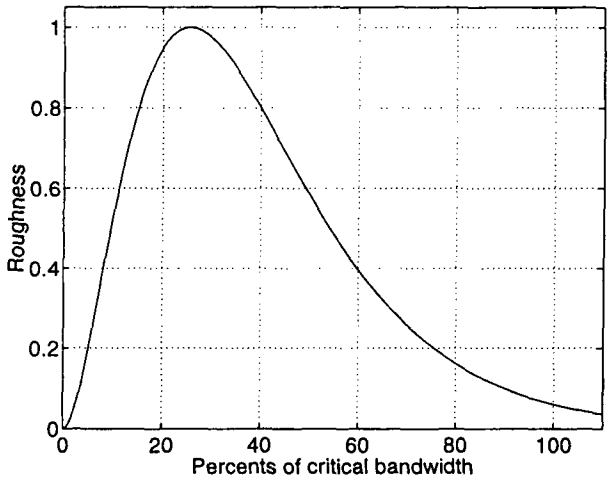

Figure 10

roughness example. The roughness of beating tone pairs (measured thanks to experiments involving judgments by human listeners) has been found to depend not on the absolute frequency difference, but rather on the frequency difference related to the width of the critical band for a given center frequency (Fig. 10). Roughness should not therefore be thought of as an acoustic feature of sound, it definitely belongs to the world of perception. This has several consequences. As the width of the band varies (Fig. 8), a given pitch interval won't have the same roughness in different registers. Thirds, for example, are free of roughness in the upper register but can be quite rough in the lower one. To be able to predict that, one needs some kind of model that could extract the relevant features from the acoustical signal and combine them.

Modeling

We have seen that all we can hear in a sound is not obvious in any of its physical representations. In the case of roughness, these representations can even be seriously misleading. Shouldn't it be possible to propose models that allow one to predict, on the basis of data obtained from psychoacoustics, which percepts would be induced by a physical stimulus? It is necessary to take certain precautions: a classic psychoacoustic study precisely characterizes a particular phenomenon by creating artificial stimuli and by analyzing the judgements of listeners within a controled laboratory situation. Quantitative data are thus carefully obtained for each of the phenomena mentioned above and for many others as well.

The relations obtained can serve as the basis for models, but in general each model describes a particular mechanism within the constraints we just mentioned. In using these models for musical purposes, it is necessary to take into account a large number of phenomena, which of course interact in a complex way. It is extremely complicated, not to say impossible, to establish a coherent ensemble from all of these sundry parts.

However, some of these phenomena are beginning to be understood in physiological terms, from which derives the idea of modeling the causes rather than reproducing the effects. In modeling the behavior of the human ear, all of the interactions and distortions are taken into account in an implicit manner (to the extent that the model captures well the properties of the auditory system). In such physiological models (Patterson, Allerhand, and Giguére, 1995; Seneff, 1988), a new representation of the sound signal is in fact proposed. This representation provides an image of what is effectively heard, within the limits of our ability to learn to read it. The previous representations were entirely oriented toward the mathematical description of the signal. Physiological models are to the contrary adapted to our perception. The sound synthesis process could be oriented by such representations, recentering the work

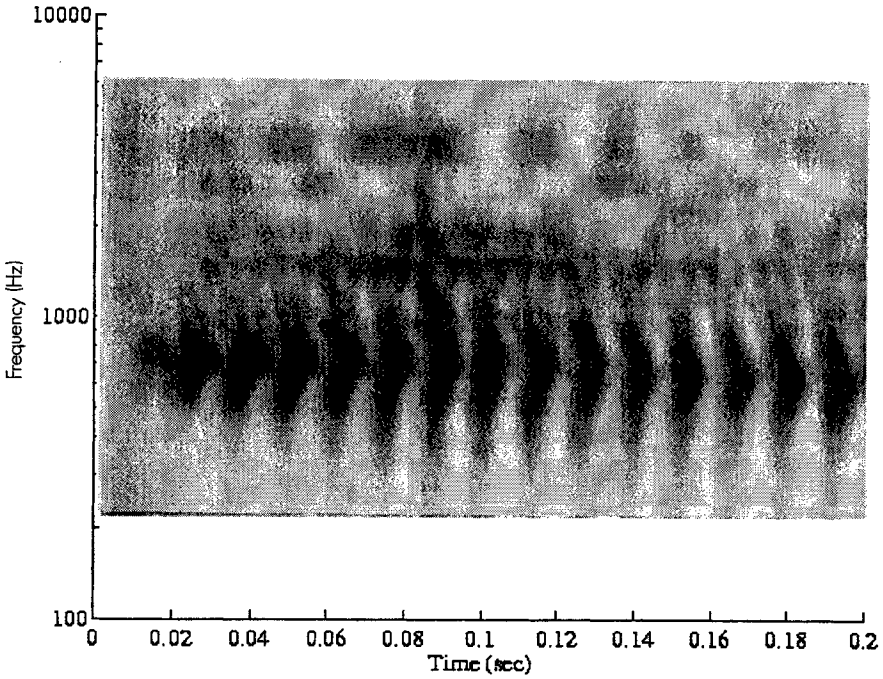

Figure 11

on the relevant perceptual parameters (Cosi, De Poli, and Lauzzana, 1994). The analysis can benefit as well: for example, roughness, hidden in the other representations since it belongs intrinsically to the world of sensation, is here revealed by the fluctuations within each critical band (Fig. 11). These fluctuations are a reflection of the perceived grainy quality. The images produced by such models thus potentially characterize the evoked sensations. The sensations are however only very basic bricks upon which the mental representation of the acoustic world that surrounds us is organized.

Auditory scene organization

Auditory representations

When listening to a noisy environment, or to a piece of music, our auditory experience is usually quite different from the collection of interleaved sinusoids, with frequencies and amplitudes varying over time, that nonetheless constitute the only available information that reaches the ears. Quite to the contrary, we structure the acoustic world in terms of coherent entities that we can generally detect, separate, localize, and identify. For example, in a concert hall we hear separately the melody played by a flute soloists, the cello ensemble, a sudden percussion entry, and our neighbour sighing — in certain concerts. This capacity is quite impressive. The chaotic form of the time-frequency representations of the superposition of all these vibrating sources, that resemble the peripheral analysis we just described, is totally unintelligible to the human eye and even to the most powerful computers. Auditory organization is nonetheless of vital importance for the survival of the species, if only to be able to distinguish the flute solo from a fire alarm! This importance allows us to state that a listener will always try, whatever the situation or listening strategy, to structure the acoustic world that confronts his or her ears. The creation of a structured representation is what allows music to be more than a simple succession of percepts.

The metaphor of the auditory image intuitively incorporates the mode of structuring that is used. An auditory image can be defined as a psychological representation of a sound entity that reveals a certain coherence in its acoustic behavior (McAdams, 1984). This definition is broad enough to allow it to be employed at several levels: a single percussion sound is one auditory image, the collection of events composing the rapid melody played by the flute is another, all those emitted by the cello section playing in harmony a third. From research attempting to make sense of all the possible sound cues that the brain uses to organize the

sound world appropriately, it seems that there are two principal modes involved in auditory image formation: perceptual fusion of simultaneously present acoustic components and the grouping of successive events into streams.

Vertical organization: perceptual fusion

One of the first immediate effects of vertical organization is the grouping together in a same image of the multiple partials of a complex sound spectrum, as analyzed by the ear, into coherent parts. This kind of organization allows us to hear a note played by a violin as a *single* note rather than a collection of harmonic partials. The main object of vertical organization is therefore at each and every instant to group what is likely to come from the same acoustic source, and to separate it from what is coming from different acoustic sources. One of the characteristics of music, as we shall see, is to constantly try to break down this simple rule. However, the cues used by the auditory system to form vertical images remain the same in musical and non-musical contexts. Understanding them is of course the key to being able to go beyond the simple equivalence between a vertical image and an acoustic source.

The position of a source in space introduces differences between the waves received by each ear (time delays, intensity differences) that allow a listener to localize it. A first cue of grouping is thus made available, one image for each spatial source. Localization can play a musical role, as in the religious antiphonal music of Gabrieli as early as 1600. Compositional writing for the classical orchestra, with its codified disposition of the instruments, integrates more or less consciously the cues of localization in its form. It is thus undoubtedly to favor the formation of 'vertical' auditory images that some composers using tape or live electronics music do not hesitate to spatialize their scores. Nevertheless, when one listens to sound emitted by a single loudspeaker, and thus originating from a single physical source, it is possible to have a more or less clear representation of different auditory images. Imagine for a moment listening to a monophonic recording of a wind quintet played over a loudspeaker: in general you should have no trouble perceptually segregating the five sources. Other strategies are thus available to the ear, based on regularities in the environment that would have been learned through an evolutionary process (Bregman, 1990). It is highly unlikely, for example, that different partials start and stop at exactly the same time if they do not come from the same source. The ear tends to group together partials that start together. If several partials evolve over time in similar fashion (this is called the common fate regularity), they will have a high probability of coming from the same source: as a matter of fact, a

modulation of the amplitude or of the fundamental frequency of a natural sound affects all of its partials in a similar way. And finally, a harmonic series will probably come from the same sound source due to the physics of sound production in forced vibration systems such as bowed strings and blown air columns. All these cues can be used together to form vertical images.

Horizontal organization: streams

The second level of the auditory image metaphor enters the realm of temporal evolution and studies the formation of auditory streams. A stream is a sequence of events that can be considered to come from a same source. A stream constitutes a single auditory image that is distributed over time. For example, the voice of someone speaking or a melody played on a musical instrument possesses a certain perceptual unity and thus forms a coherent image.

The general law governing the formation of streams appears to be based on spectral continuity (McAdams and Bregman, 1979). This law also reflects some kind of regularity in the environment, as a sound source tends to change its parameters progressively over time. Continuity is evaluated according to several cues. The most studied ones were frequency and time proximity (van Noorden, 1975). Events that are close according to these dimensions will tend to form a separate stream, just as melodies having a small range and rapid tempi in different registers from one another will be segregated into different voices. There is a trade-off between time and frequency proximity for stream formation. Actually, most combinations of these to parameters give rise to an ambiguous perception where streams can be voluntarily built or modified. Timbral similarity is another grouping cue brought into play by orchestration.

The formation of auditory streams has dramatic effects on the perception of the acoustic events. Judging the timing between two successive events is usually a trivial task, however, if the events are part of two different streams the judgments become impossible (Bregman and Campbell, 1971). Melody recognition is also affected by the grouping of all the correct notes in a same stream (Dowling and Harwood, 1986).

Emergent attributes

Vertical and horizontal grouping mechanisms can interact in a complex manner. It is more than likely that at some moment of an evolving auditory scene, energy in a certain frequency region could be attributed to several different vertical images, or to an ongoing stream. In this case, choices are made and once this energy has been attributed to an auditory

image, it is taken away from the others. This 'stealing' of energy between images can lead to the consequence that attributes such as pitch and brightness (Bregman and Pinker, 1978) or loudness (McAdams, Botte and Drake, in press) or roughness (Wright and Bregman, 1987) can be altered by auditory organization. The expression 'sound object', if used regardless of its original historical context of *musique concrète*, can therefore be misleading. A sound object implies a sound event possessing a basic unity, defined by some characterizable features. It is in fact the case that all sound properties are dependent on dynamic relations with the context, within which attributes as important as pitch, loudness, roughness, and other dimensions of timbre can find themselves significantly changed as streams and vertical images are formed and reorganized. There exists no indivisible, stable sound object, but rather auditory images that possess more or less coherence in a dynamic relation.

Each vertical auditory image, once it is formed and only then, possesses emergent attributes. These emergent attributes are born of the fusion of its components. An emergent attribute is different from the sum of elementary attributes of the components contributing to the image. The addition to spectral components higher in frequency to a pure tone can lead to the perception of a pitch lower than the original pure tone one, as in the case of the missing fundamental (Schouten, 1940; Terhardt, 1974). The recognition of a certain sound source emerges from the fusion of many components into a global timbre, each of them taken alone often being unable to unveil the sound origin (McAdams and Bigand, 1993). Correct auditory grouping is therefore essential to build these emergent attributes and grasp a representation of a natural environment.

Auditory images and musical structures

The usual result of auditory organization is to build stable auditory images, each corresponding to a sound source, to be able to recognize them. In a musical context, the emergent attributes often come from 'chimeric' sound sources. If each instrument of the symphonic orchestra was to be heard as a single source, the perception of the musical structures imagined by the composer would certainly be quite difficult — just as difficult as if all the instruments were fused into a single auditory image. Auditory organization has therefore been explored by composers for a long time.

Deceiving processes of horizontal organization allowed the writing of virtual or implied polyphony, where a monophonic instrument expresses more than one voice at the same time. Tricking processes of vertical organization is the key to orchestration, where unheard of and augmented

timbres can be created by fusing different instruments. It can also have radical structural consequences. In his piece *Lontano* (1967), Ligeti creates dense structures within which, because of vertical fusion cues, the instruments cannot be distinguished. He then remarks that 'polyphony is written but one hears harmony'. A few years later, organization cues that induce the exact opposite consequences were to be used by the same composer to write the *San Francisco Polyphony* (1973–74), this time clearly heard as polyphony.

The organization of the auditory scene can even be an argument for musical structure. A first example of this is *Mutations* by Risset where the transformation of a chord into timbre that we mentioned earlier is done with the help of auditory organization cues. The inharmonic structure of the chord requires the convergence of other cues such as synchrony to be heard as a fused timbre. Another even more extreme example is the piece *Désintégrations* by Tristan Murail (1982). In this work, a set of intervals is first heard fused in a section with a rapid tempo, comprising a complex melodic line, though it will be heard in a 'disintegrated' version in a nearly static section. This structure is expressed by using different organizational cues (see Figs. 12, 13).

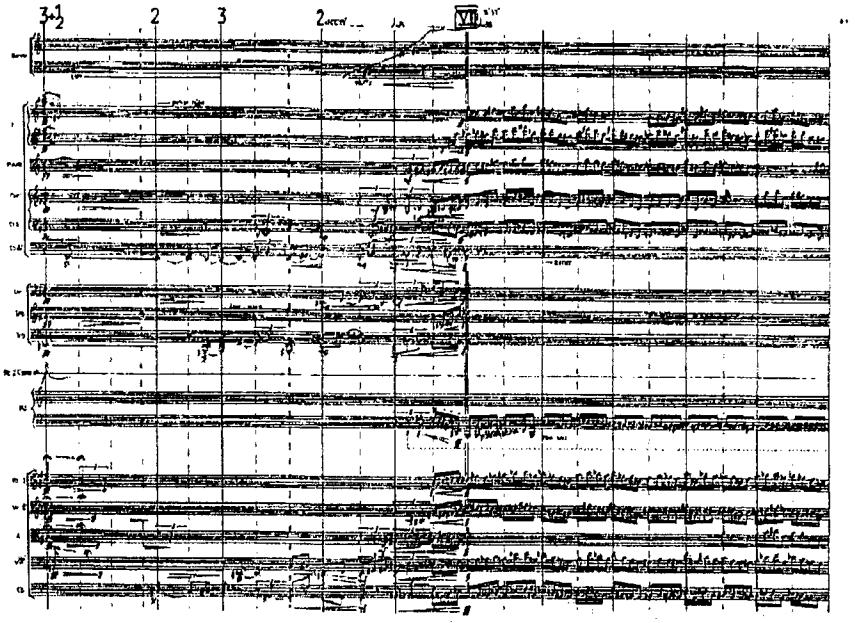

Figure 12

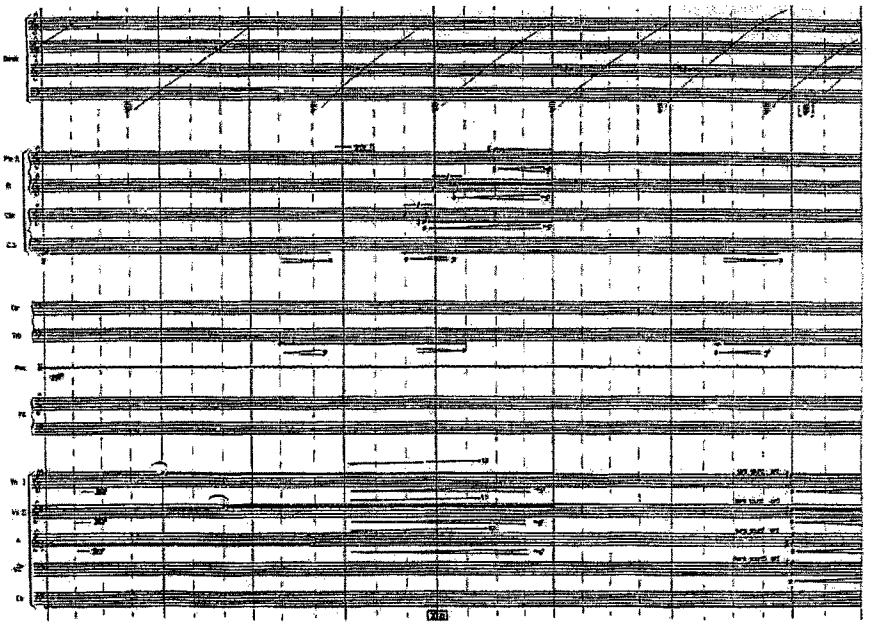

Figure 13

Listening and cognition

Memory

Of course, the bottom-up processes that we have described so far, that took us from acoustic vibrations to auditory images, are not the only ones to intervene in musical listening, nor even in the formation of auditory images. A complex set of high-level cognitive processes also come into play. Without pretending to describe exhaustively and precisely these processes (insofar as this is possible), we would like in concluding to address certain cognitive aspects of listening, stressing the ones that are associated more notably with memory.

Various notions are implicated in music cognition: we might cite attention, cultural knowledge, and temporal organization in perception. A common characteristic seems to tie together these kinds of processes as they play a role in listening: memory. Memory is linked to attending in the sense that attention seems to be predisposed to focus on events that are expected on the basis of cultural knowledge abstracted from past experience (Jones and Yee, 1993). Memory allows listeners to implicitly

learn the basic rules of the musical culture to which they belong (Krumhansl, 1990; Bigand, 1993). Finally the very notion of time, so essential to music, derives from our ability to mentally establish event sequences through the use of memory (Damasio et al., 1985).

The auditory mode of recall is remarkably powerful. Crowder and Morton (1969) have shown, in a task requiring listeners to recall a list presented either visually or auditorially, that the auditory modality has a net advantage over the visual modality for the later elements in the list. The hypothesis advanced to explain this superiority is the existence of a sensory storage: a sort of 'echoic' memory, specific to hearing, that conserves the stimulus trace for a brief period of time. This hypothesis has since been refined and there are most likely several different retention intervals (Cowan, 1984). One of these intervals would be on the order of several hundreds of milliseconds and another one on the order of several seconds. The first storage would be related to sensation, constituting a sort of 'perceptual present', while the second one would serve as a basis for what is called working memory. These fairly short durations raise all kinds of questions concerning the possibility of apprehending structures extended through time and carried by sound. The stimulus trace vanishes within a few milliseconds of the echoic memory, and the working memory cannot hold more than a few items.

Still, we can experience a perception of form and meaning over time with spoken language, for instance, which is also carried by sound. So is there a long term storage of the facsimile of the acoustic stimuli we receive? In fact, there are a great deal of both behavioral and neurophysiological data that lead us to believe that there is not a memory center where this kind of storage takes place in the brain (Rosenfeld, 1988). Memory would more likely be distributed, as a by-product of cognitive processing, in the form of *potential representations*. In other words, in the presence of a stimulus, the brain activity results in the extraction of the relevant features, making generalizations, and forming categories. This theory has some physiological basis. Perception activates the primary sensory cortices in the brain, which through their activity in time detect and encode different features of sound. These activity patterns are then projected onto association cortices. Each neuron in the association cortices receives a large number of connections from various areas of the brain, which allows for generalization: stimuli sharing similar features will activate the same groups of neurons, reinforcing the connections between the concerned sensory cortices and these neurons. When a new stimulus is perceived, if it is similar enough to a potential representation already memorized, it will be categorized as a member of the same family. In recall or in imagination or dreaming, or also in forming categories, it

would be the convergence zones that then activate the sensory cortices in the reverse direction (Damasio, 1994).

The transformation of stimuli into potential representations, which are not reproductions but rather abstractions of features of a stimulus, can help to interpret a study on the memory for melodies by Crowder (1993). The recall of melodies seems first of all to be based on the pitch contour, if recall follows shortly after learning has taken place. However, if recall is delayed in time, the influence of contour decreases to the benefit of pitch relations: an abstraction has been performed from absolute pitch, an early level of perception, to relative pitch intervals and relations within a tonal framework. It is this kind of abstraction that seems to be stored in long-term memory.

This leads us back to what is implied in the possible use of different continua, for only discrete scales allow a classification of perceptual values that is apt to be coded in the brain in terms of abstract relations. In fact, the existence of discrete degrees dividing the octave is one of the rare constants that can be found throughout nearly all cultures (Dowling and Harwood, 1986). Further, numerous studies on musical cognition seem to converge toward the importance of metric and rhythmic hierarchies and thus of the discretization of time (Lerdahl and Jackendoff, 1983; Clarke, 1987). The question is also raised for the use of timbre, seen as a potential carrier of structure, and its various, apparently continuous dimensions (McAdams, 1989). These considerations on memory thus have a direct influence on musical structures, if one wants them to be intelligible.

Arousal

After all these considerations about acoustics, psychoacoustics and cognition, a text about music would seem to politely avoid an uneasy point if it did not address, even superficially, the question of arousal (or emotion) in music. This issue, which may be considered at first glance to be beyond the field of scientific investigation, or with one foot in the aesthetics domain, is in fact the subject of numerous psychological studies that could find an application in the musical domain. So, what is arousal for the cognitive psychologist?

An interesting answer is proposed by Mandler (1984). Human beings have a certain number of schemas, some innate and directly linked to survival, others acquired and eventually modifiable that are linked to past experience. Perceptions are thus evaluated in terms of expectancies. In the case of a perception that conforms to our expectancies, one might expect that little cognitive activity is necessary. On the other hand, a perception that goes against expectancy triggers both an emotional reaction

and cognitive processing, the latter in order to adapt our representation of the external world to what has been perceived. Damasio goes even further, in showing by way of numerous examples from neuropathology that preceding arousal is not only sufficient, but also absolutely necessary for the correct operation of many cognitive processes such as socially relevant decision-making and personal planning for the future (Damasio, 1994). If arousal occupies such an important place in our cognitive processing (usually associated with 'pure reason'), the question of the relation between emotion and music takes on a new dimension.

Let us try to make a parallel between what we learned from cognitive psychology and what happens in music listening. Bregman (1990) clearly states that schemas influence the way we organize the auditory scene, by providing us with a mechanism to extract certain things that we are seeking within that scene. Another well-established schema is our knowledge of tonality. These schemas, and especially the latter will give rise to expectancies, which, being violated or not, will evoke arousal. In the domain of musical tension, the work of Bigand (1993) has demonstrated the influence that implicit knowledge of tonality has on music perception in both professional musicians and non-musician listeners. Asked the judge the degree of tension of an interrupted melody, the listeners clearly expressed expectancies linked to tonality. Bharucha (1989) has simulated with neural nets the expectancies of Western and Indian listeners presented with their respective musics, and their erroneous comprehension faced with the music of the other culture. The arousal (the feeling of 'tension') can here be seen as the result of an enforcement of cultural rules that give rise to expectancies. These rules, assumed to be shared by listeners, could be called external referents. Sound charged with an extra-musical meaning can also be considered as playing with external referents. On the other hand, there are also reactions to music that originate in the musical material, based on its immediate qualities, and which generate expectancies over time. The auditory equivalent of Gestalt good-continuation laws (Koffka, 1935) have been proposed to generate expectancies that can give rise to arousal if they are violated (Meyer, 1956). Another immediate sound quality, roughness, has been shown to play a part in tension perception in the tonal context along to these other factors. The self-generated expectancies acquire even greater importance if external referents are kept in the background (Pressnitzer, McAdams, Winsberg, and Fineberg, 1996). In this case, the embedded structure is revealed by cognitive mechanisms and thus depends to a great extent on the organization into auditory images, and thus on perception, which in turn depends on the physical phenomena.

To finish with our roughness example, it is clear that in tonal music, the influence of roughness is altered by the deep implicit or explicit

acculturation that we have of basic harmonic rules.[6] However, when the composer leaves the well-trodden path of convention, the mastery of a cue allowing the expression of a simple and immediate tension can become primordial again. For example, the piece *La tempesta d'après Giorgione* by Hugues Dufourt (1977) is constructed as a movement of latent tension going (nearly) all the way to its own paroxysm. This piece employs wind instruments in a very low register. In the low register the spacing between the partials of the instrument sounds are very small compared to the critical band. Different partials thus fall into the same critical band: this induces a perception of roughness. The same material transposed to a medium or high register would lose its meaning to a large extent. As such, acoustics (the beats) and psychoacoustics (perception of these beats) contribute to the structure of the piece! Gérard Grisey can therefore exclaim, 'we are musicians and our model is sound, not literature, sound, not mathematics, sound, not theater, plastic arts, quantum theory, geology, astrology, or acupuncture' (Grisey, 1984, p. 22, as quoted by Wilson, 1989). Spectral music, in its search for expression through the material itself, without hidden or conventional reference, makes possible the recourse to certain data from acoustics and psychoacoustics not only to justify certain choices a posteriori, but also as a means of formalizing musical processes.

Conclusion

Throughout this article, we have attempted to evoke a set of facts derived from acoustics, psychoacoustics, and cognitive psychology, that have a certain resonance with the spectral approach. Through the example of roughness, we showed how a part of an essential musical feature such as dissonance could be embedded in the material itself, involving the knowledge of its acoustic ground, the study of its transformation through perception, and how it is put into perspective through the influence of cognitive processes involved in listening. The mutual curiosity of scientists and musicians is not recent, but it has often had different motivations. Concerning our example, the question of why certain intervals are consonant had sparked interest throughout history. When Pythagoras was interested in intervals, it took the form of a new manifestation of the omnipresence of numbers. Later, with Kircher, divine intervention was sought in harmonic ratios. Kepler, Galileo,

6. We might, however, ask ourselves if these rules, that actually modulate the perceived roughness, do not in fact have an unconscious foundation based on roughness (at least early in musical life).

Leibniz, Euler, Helmholtz, among others, all made attempts to propose an explanation of their own (Assayag and Cholleton, 1995).

The goal here is much more modest. To bring this goal to light we would like to cite a musical custom in New Guinea that is practiced by the Kaluli of Papua. A ceremony takes place in a hut where the members of a clan are gathered. Passing visitors enter the hut at dusk, singing the names of places they crossed within the territory of their hosts. This song produces particular reactions among the listeners. Schieffelin (1979), quoted in Dowling and Harwood (1986), tells the story as follows:

'After a while, the audience (the hosts) becomes very deeply moved. Some of them burst into tears. Then, in reaction to the sorrow they have been made to feel, they jump angrily and burn the dancers on the shoulders with the torches used to light the ceremony. The dancers continue their performance without showing any sign of pain. The dancing and singing with the concomitant weeping and burning continue all night with brief rest periods between songs.' (p. 128).

Evoking the name of a place to which a tragic memory is perhaps attached (the death of family or friends, for example) has triggered such reactions in the listeners. Obviously, the singer is not interested in acoustics, nor even in psychoacoustics, and why would he be so since he is leaning on external referents to trigger arousals. The point here is that he knows very little about the referents he is using, being a stranger to the territory of his hosts. If only the small amount of data concerning music perception that have been presented here could be used to avoid such embarrassing situations...

Author note

The authors would like to thank Joshua Fineberg for providing musical material as well as useful comments on earlier versions of the manuscript. Correspondence should be addressed to either Daniel Pressnitzer or Stephen McAdams, IRCAM, 1 place Igor Stravinsky, F-75004 Paris, France (email: daniel.pressnitzer@ircam.fr; smc@ircam.fr).