

JENS HJORTKJÆR

Sound objects – Auditory objects – Musical objects

Introduction

Objects are fundamental to experience but how do we experience an object in sound perception? Pierre Schaeffer suggested the concept of a 'sound object' in his comprehensive *Traité des Objets Musicaux* aiming to describe the sonorous anatomy of musical sounds.¹ This ambitious turn to perceptual music research was nourished by the emergence of electronic sound processing technologies in the 20th Century. New tools for electronic synthesis had allowed composers to explore musical timbre as a source of musical invention and organization, challenging the conception of pitch structures as the fundamental constituent of music. At the same time, electronic sound had detached sounds from their physical sources and suggested a new understanding of the 'sounds themselves' as purely perceptual objects. The program for description of sound objects launched by Schaeffer was, however, not restricted to electronic sounds but aims at identifying sonorous features of musical events that would enable the composer to work with 'sound' instead of (or in addition to) working with traditional musical parameters, such as harmonies and melodies.

Auditory neuroscience today is still struggling to understand sound objects and how auditory processing in our brains gives rise to 'auditory objects'.² As in Schaeffer's program, much previous auditory research has focused on the representation of 'basic' sound features corresponding to traditional musical parameters (pitch, loudness, timbre, duration, etc.). However, recent research has also suggested that object features that are behaviorally relevant have a privileged role in sensory processing. Rather than being exclusively occupied with constructing faithful representations the sound coming into our ears, the brain is always involved in abstracting and selecting meaningful information about our auditory environment and about objects that are relevant to the perceiving organism. This also suggests that sound objects must be viewed in the behavioral context of perceiver and in the biological context in which sound perception evolves.

1 P. Schaeffer: *Traité des Objets Musicaux* (Paris: Éditions du Seuil, 1966).

2 T.D. Griffiths & J.D. Warren, "What is an auditory object?", *Nature Reviews Neuroscience* 5 (2004); J.K. Bizley & Y.E. Cohen, "The what, where and how of auditory-object perception", *Nature Reviews Neuroscience* 14 (2013).

A phenomenology of sound objects

The common conception of an object is that of a physical thing that we experience. We may see a ball but we also experience it in other senses. If we bounce the ball then we automatically relate the sounds of the different impacts in the bouncing sequence to our conscious perception of the same object. In everyday sound environments, sounds from multiple different sound sources reach our ears at the same time and yet we experience distinct 'objects', e.g. running water, a person talking, a car passing by, music on the radio. Separating and integrating these events over time is a formidable complex task accomplished by the auditory system ('auditory scene analysis'³) and yet experienced at ease.

This conception of auditory objects, however, is different from Pierre Schaeffer's phenomenological notion of a sound object. To Schaeffer, the sound object is the result of a particular mode of listening. In fact, Schaeffer defines a sound object negatively in relation to its physical source: it is the perceptual gestalt that results from reducing away any reference to the particular source that gives rise to the sound. This relies on the idea of different modes of listening where 'listening' (*écouter*) is naturally oriented towards the cause of a sound event in contrast with perceiving or sensing (*ouïr*) the raw sound as it is given in passive experience.⁴ A sound object is established by suspending our habitual listening for a particular source and turning this *intention* of the sound as a sign of something back on 'the sound itself'. We may hear the singing of a plumbing system in a hotel⁵ not as sounds caused by the pipes (the sound source) of the plumbing system (the meaning), but as sounds that have particular sonorous features. The sound event would create a particular *type* of sound object (a *grosse note*) characterized, for instance, by a medium sustained duration with a complex eccentric variation in pitch content.⁶

Schaeffer underlines that the sound object that we experience when turning to the 'sound itself' is not the physical sound signal but rather how the sound is qualitatively perceived.⁷ Schaeffer's research program aims at understanding the complex 'correlations' between the physical sound signal and the perceived sound.⁸ The experience of pitch, for instance, is not identical to the frequency content of the sound but relates to it in complex nonlinear ways.⁹ While Schaeffer is also critical of psychoacoustics studying simple relations between the sound signal and perception,¹⁰ many of the

3 A. Bregman, *Auditory Scene Analysis* (Cambridge: MIT Press, 1990). Bregman prefers the term 'stream' to the term 'object'.

4 Contrasted with 'abstract' modes of listening: hearing (*entendre*) with an intention to listen and comprehending (*comprendre*) the meaning of what we hear. Schaeffer, *Traité*, p. 116.

5 *Ibid.* p. 441.

6 *Ibid.* p. 457.

7 *Ibid.*, p. 269.

8 Schaeffer uses the term 'anamorphosis' for the ways in which the physical signal becomes distorted in perception. *Ibid.*, p. 216f.

9 *Ibid.*, p. 188.

10 *Ibid.*, p. 170f.

formal parameters of his typo-morphology express intuitions about psychoacoustic concepts of basic perceptual parameters of sound (pitch, loudness, duration, roughness, harmonicity, etc.).

Although the sound object as a perceptual gestalt is not a physical object in the world, as Schaeffer argues, properties of physical objects may however still influence sound perception. As we will discuss in the following, research on auditory processing suggest that the process of extracting object features sounds in our environment is integral to our auditory system. In a biological context, the perceiving organism is always involved in extracting relevant information about the environment in order to interact with it, and not only to construct representations of ‘the sound itself’. Although we may have intuitive notions about what the ‘basic’ features of sound are, it is not necessarily clear why particular sound parameters become perceptual constituents of sound experience in the first place.

The standard model of auditory object processing and its challenges

The abstraction of object properties is accomplished by a complex processing sequence in the auditory system. A visual object is grouped into a coherent gestalt by a process involving e.g. extraction of edges. But what are the ‘edges’ of auditory objects? Sounds are represented by their frequency content over time from the level of the inner ear, but simple spectrotemporal modulations do not necessarily give rise to distinct gestalts. For instance, the different partials in a violin tone are spectro-temporal ‘edges’ but they are integrated into the perception of a tone as a single gestalt. Instead, neurons in the auditory brainstem have been proposed to detect the degree of periodic regularity over time and transform the pitched sound into a stable ‘auditory image’ (the neural correlate of the perceived gestalt).¹¹ The pitch of the tone is then represented in pitch maps on the surface of the auditory cortex.¹²

The abstraction of more complex object properties is thought to involve cortical mechanisms beyond the primary auditory cortex. In a functional neuroimaging study, Zatorre et al. found that brain activity along the anterior part of the superior temporal cortex co-varied with the saliency of auditory object features.¹³ This supports the notion of an anterior functional stream emerging from the auditory cortex involved in identifying sound categories (what an object is) (Fig 1 right). Similarly as in the visual system, the auditory ‘what’ stream is proposed to work in parallel with a postero-dorsal ‘where’ stream involved in extracting the spatial location of a sound (where an object is).¹⁴ The ventral stream has also been implicated in abstracting sound features that allow us to

11 R.D. Patterson et al., “Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform”, *Journal of the Acoustical Society of America* 98 (1995).

12 C. Pantev et al., “Tonotopic organization of the auditory cortex: Pitch versus frequency representation”, *Science*, 246 (1989).

13 R. Zatorre, M. Bouffard, P. Belin, “Sensitivity to auditory object features in human temporal neocortex”, *Journal of Neuroscience* 24 (2004).

14 J.P. Rauschecker & B. Tian, “Mechanisms and streams for processing ‘what’ and ‘where’ in auditory cortex”, *Proceedings of the National Academy of Sciences of the United States of America* 97 (2000).

identify, for instance, a violin timbre of a tone regardless of variations in pitch, loudness, duration, reverberation, etc.¹⁵

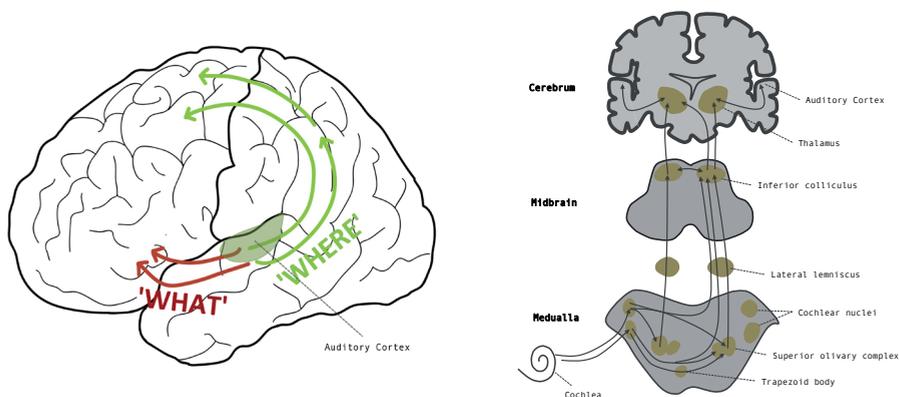


Fig 1. Left: The ascending auditory pathway. Right: Auditory 'what' and 'where' streams.

This 'standard model' of auditory processing¹⁶ suggests that the perceived sound object is the result of a hierarchical process of extracting higher-level object features based on earlier representations of sound features in the ascending auditory pathway (Fig. 1 left). But results about the representation of 'basic features of sound' in the auditory cortex are far from being conclusive after many years of research. The *visual* cortex is sensitive to basic sensory features of visual stimuli that emerge in continuous map-like representations of edge orientation, position, contrast, etc. But the nature of representations of 'basic' continuous properties of sound in the auditory cortex, such as loudness, spatial location, duration, or even pitch, is still debated. Many sound features are represented with high resolution at the level of the auditory brainstem, whereas cortical neurons are known to respond more 'sluggishly'.¹⁷ Although they may respond selectively to a given pitch, they may respond differently to different sorts of pitched sounds and may respond even stronger to other parts of a complex sound (like background noise¹⁸). On the other hand, a number of neurophysiological studies have reported that neurons in primary auditory cortex respond selectively to specific classes of sounds that are behaviorally relevant to the perceiving animal. For instance, electrophysiological studies have reported selective responses to con-specific vocal calls in natural auditory environments but not to isolated sounds with similar low-level acoustic features.¹⁹

15 J.D. Warren, A.R. Jennings & T.D. Griffiths, "Analysis of the spectral envelope of sounds by the human brain", *Neuroimage* 24 (2005).

16 I. Nelken, "Processing of complex stimuli and natural scenes in the auditory cortex", *Current Opinion in Neurobiology* 14 (2004).

17 I. Nelken & O. Bar-Yosef, "Neurons and objects: the case of auditory cortex", *Frontiers in Neuroscience* 2 (2008).

18 O. Bar-Yosef, Y. Rotman & I. Nelken, "Responses of neurons in cat primary auditory cortex to bird chirps", *Journal of Neuroscience* 22 (2002).

19 Ibid.

This has led to the suggestion that the auditory cortex is already sensitive to ‘objects’ in the sense of behaviorally relevant sound categories rather than to continuous sound features *per se*.²⁰ In this view, auditory cortex neurons are already involved in finding features of particular object classes that are invariant across physical variation in the sound. For instance, speakers of a particular language are typically less sensitive to acoustic variations within phonological categories (e.g., different spoken instances of the /da/) but highly sensitive to small variations between categories (e.g. between acoustically similar instances of /ba/ and /da/).²¹ Encoding of the high-level auditory object (collapsing acoustically different instances of /ba/) enables us to identify and use these sounds efficiently in a given context, but at the expense of lower-level information about the sound.

But if the auditory cortex is processing high-level objects, how is it then still possible for humans to discriminate fine-grained physical features of sound?²² So-called Reverse Hierarchy Theory (RHT) propose that while only high-level objects are immediately accessible to perception, access to lower-level sensory information can be accomplished in situations that allow reverse processing along the processing hierarchy.²³ Sound discrimination between different instances of the same sound object is possible in particular situations that, for instance, eliminate the need to understand the object-level meaning of the sound and focus instead on its sensory details. In a behavioral study supporting RHT, Nahum et al. showed that listeners make different use of lower level sound information (here, different phase information between the two ears) in different listening situations.²⁴ Listeners did not make use of available low-level acoustic cues to discriminate between phonologically similar words during a semantic association task, but only during explicit identification that allowed them to focus on acoustical details. This, perhaps contrary to the common intuition, indicates that we do not always access the full range of sound information coming into the ears, but only relevant object categories.

Sensitivity to auditory objects at the level of the auditory cortex ensures fast and flexible integration of relevant auditory information into ongoing behavior. But sensitivity to regularities of a particular sound environment is also found beyond real-time perception in how experience shapes the *descending* auditory system at longer time scales. Strait et al. compared the brainstem response to musical tones in pianist and non-pianist musicians.²⁵ They found that the temporal neural response in the brainstem of pianists followed the particular amplitude envelope of piano tones with a higher level of detail compared to both non-pianist or to non-piano timbres. This sug-

20 Nelken & Bar-Yosef, “Neurons and objects”.

21 A.M. Liberman et al., “The discrimination of speech sounds within and across phoneme boundaries”, *Journal of Experimental Psychology* 54 (1957).

22 For instance, listeners identify frequency differences down to around 0.2%, which is well below the half-tone difference in equal temperament of around 0.6%. Nelken & Bar-Yosef, “Neurons and objects”.

23 M. Ahissar et al., “Reverse hierarchies and sensory learning”, *Philosophical Transactions of the Royal Society, B: Biological Sciences* 364 (2008).

24 M. Nahum, I. Nelken & M. Ahissar, “Low-level information and high-level perception”, *PLoS Biology* 6 (2008).

25 D.L. Strait et al. “Specialization among the specialized: Auditory brainstem function is tuned in to timbre”, *Cortex* 48, (2012).

gests that extensive experience with a particular sound category leads to plastic changes of the auditory system at the subcortical level. Even the evolution of inner ear nerve fibres has been suggested to result from an adaptation to object classes in the natural sound environment. Lewicki suggested that tuning properties of the auditory nerve are optimal for processing information about categories of vocal and non-vocal environmental sounds in our natural ecology.²⁶

Ecological acoustics and musical instruments

The sensitivity to object features at all levels of the auditory system underlines the importance of relating perception to the environment in which perception takes place. Perceptual access to lower-level sound information, as proposed by RHT, involves 'reverse' processing in the auditory hierarchy and only occurs in particular listening situations that suspends our usual orientation towards objects in real-time behavior. This is seemingly in line with Schaeffer's phenomenological notion of sound objects as a result of reduced listening that suspends natural listening for sound sources. However, the privileged role of objects in our auditory system also questions Schaeffer's idea of turning to the 'sound itself' and that musical listening is oriented toward basic parameters of sound 'before' we attribute object properties. Instead, object features are likely to influence what sound features become perceptually relevant in the first place, also when the sound source is not the conscious focus of attention.

As mentioned above, a fundamental function of the auditory system is to extract sound properties that are invariant for objects and will allow us to identify them through physical variation. Recent perceptual research on 'ecological acoustics'²⁷ has focused on describing invariances that allow us to pick up information about e.g. the length, shape or material of an object or about sound producing actions.²⁸ This reflects an alternative to the traditional focus in psychoacoustic research on perceptual properties of the 'proximal' sound stimulus arriving at our ears. Rather than viewing object perception as a process of inference from sound features to object representations, the ecological approach argues that the physical object itself (the 'distal' stimulus) contains information that we can pick up in perception.²⁹

26 M. Lewicki, "Efficient coding of natural sounds", *Nature* 5 (2002).

27 N.J. VanderVeer, *Ecological Acoustics: Human Perception of Environmental Sounds*, (PhD Dissertation, Cornell University 1979).

28 E.g. C. Carello, K.L. Anderson & A.J. Kunkler-Peck, "Perception of object length by sound", *Psychological Science* 9 (1998); A.J. Kunkler-Peck & M.T. Turvey, "Hearing shape", *Journal of Experimental Psychology* 26 (2000); S. McAdams, A. Chaigne, V. Roussarie, "The psychomechanics of simulated sound sources: Material properties of impacted bars", *Journal of the Acoustical Society of America* 115 (2006); W.H.W Warren & R.R. Verbrugge, "Auditory perception of breaking and bouncing events: A case study in ecological acoustics", *Journal of Experimental Psychology* 10 (1984). See also D. Rocchesso and F. Fontana (Eds.), *The Sounding Object* (GNU Free Documentation License, 2003).

29 W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception", *Ecological Psychology* 5 (1993). Gaver argues that traditional psychoacoustics has, perhaps paradoxically, been occupied with *musical listening* to the 'sound itself' rather than with *everyday listening*. This distinction is thus similar to Schaeffer's distinction between natural and reduced listening modes.

For instance, an impact on a solid bar creates vibration modes in the solid material depending on the particular physical properties of the object. The ratio of harmonic frequencies propagated through the surrounding medium depends particularly on the boundary conditions of the object (e.g., whether the bar is clamped or freely moving), but less on other object properties such as length, elasticity, or mass.³⁰ This means that ratio between frequency partials in the emitted sound (the harmonicity) is a potential structural invariant that allows a perceptual system to pick up information about this particular object property. The rate of vibration, on the other hand, covaries with the mass density of the object and may carry information about its material. Hearing a pitched sound as a single gestalt (and not as unrelated partials) is a way of picking up information about an object that is a distinct and constant physical entity in the environment.

Perceptual research on musical timbre has confirmed the relevance of properties of the instrument source, even when listeners are not specifically attending to them. Timbre research has traditionally focused on describing the acoustic correlates of the particular multidimensional perceptual character of timbre. Different studies have examined the similarity between instrument tones and found that listeners tend to focus on particular distinct sound dimensions.³¹ Over studies, one perceptual dimension is consistently related to the increase of sound energy in the initial attack portion of the tone, while another is related to the distribution of frequencies in the long-term spectrum.

However, recent meta-analyses of timbre studies suggest that mechanical properties of the musical instrument and the manner of playing it is reflected in the complex perceptual structure, although listeners are simply asked to focus on the similarity between tones with varying timbre.³² Fig. 2 below shows a re-plot of two of the perceptual dimensions found by McAdams et al. As can be seen, sounds that are perceived as being more similar also have similar source properties.³³ Different object properties such as the manner of excitation or the material of the instrument body can be identified as regions in the perceptual space. In a sound source perception study supporting this, McAdams et al. used a physical synthesis model of a xylophone bar that allowed the authors to control the mechanical and geometrical properties of the sounding object explicitly. The authors asked listeners to rate the similarity between sounds from simulated objects varying in mass, viscoelastic damping, and length and found an accurate perceptual representation of these physical parameters, even though listeners were not asked to attend to them.

30 P.M. Morse & K.U. Ingard, *Theoretical Acoustics* (New York: McGraw-Hill Book Company, 1968).

31 E.g. J.M. Grey & J.W. Gordon, "Perceptual effects of spectral modifications on musical timbres", *Journal of the Acoustical Society of America* 63 (1978), S. McAdams et al., "Perceptual scaling of synthesized musical timbres", *Psychological Research* 58 (1995), P. Iverson & C.L. Krumhansl, "Isolating the dynamic attributes of musical timbre", *Journal of the Acoustical Society of America* 95 (1993).

32 B. Giordano, *Sound source perception in impact sounds* (PhD Thesis, University of Padova 2005), B. Giordano & S. McAdams, "Sound source mechanics and musical timbre: Evidence from previous research", *Music Perception* 28 (2010).

33 J. Hjortkjær, *Towards a Cognitive Theory of Musical Tension* (PhD Thesis, University of Copenhagen 2011), p. 242.

This suggests that listeners pick up object properties from tones and that these are *implicitly* reflected in perception. The perceptual importance, for instance, of the attack portion of a tone found by timbre studies was also noticed by Schaeffer who argued that sound objects could be classified according to qualitatively different forms of attack.³⁴ This is mirrored by the ‘sluggishness’ of auditory cortex neurons mentioned above, where neurons may respond precisely to sound only at their onset and seemingly throw away the precise temporal response to ongoing amplitude variations found at the brain stem level.³⁵ Temporal evolution of the attack, however, may be informative about the manner in which an object is manipulated. As can be seen in figure 2, tones produced by continuant excitation (blowing or bowing) and tones produced by an impulse on the instrument (plucked or struck) cluster in different parts of the perceptual space. The first perceptual dimension correlating with the amplitude rise time effectively categorizes the different types of actions, although there is no identification task related to them.

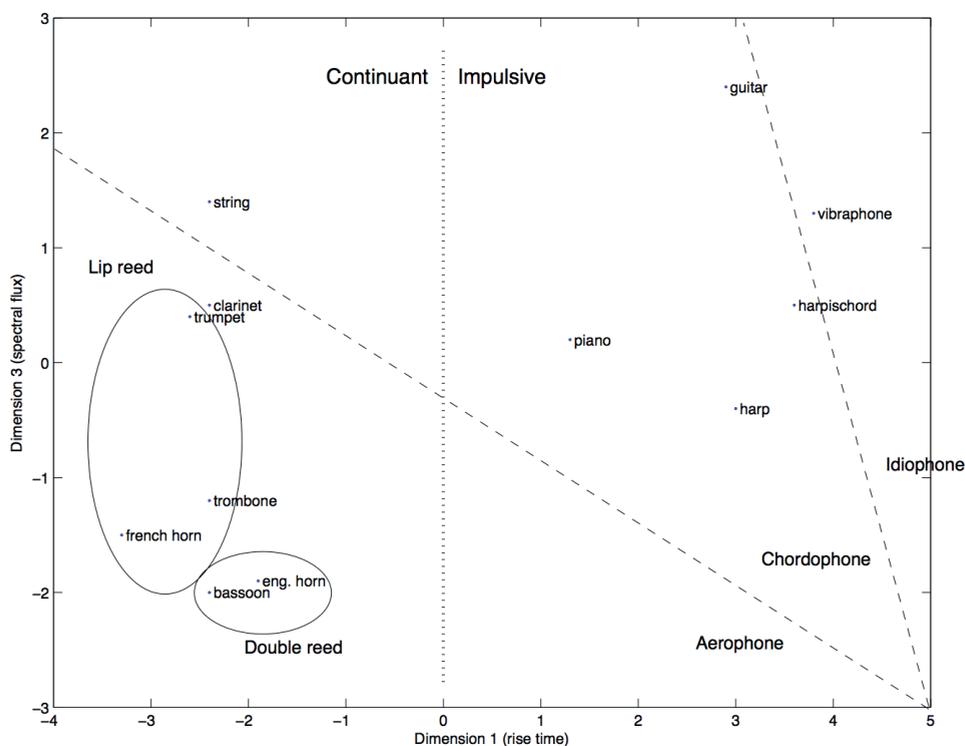


Fig. 2. Perceptual timbre dimensions 1 and 3 reported by McAdams et al.³⁶ Different mechanical properties of the instruments, such as excitation mode and instrument family, appear as different regions in the perceptual space.

34 Schaeffer, *Traité*, p. 226f.

35 Nelken, "Processing of complex stimuli".

36 McAdams et al., "Perceptual scaling of synthesized musical timbres".

Towards a neurophenomenology of sound objects

The auditory system has a remarkable ability to extract information about the sound world around us. Abstraction of properties that belongs to objects is an integral function of the auditory system and may begin already in early sensory processing. This also questions the idea that continuous parameters of the 'sound itself' (it's duration, intensity, frequency content, etc.) is the 'raw material' of listening, and that the perception of objects is a process of association *from* these parameters. It suggests instead that object-level properties are the immediate target of perception.

This is in apparent contrast with Schaeffer's phenomenological definition of sound objects as a reduction of sound sources to the 'sound itself'. It is, however, not in contrast with the way in which phenomenology has traditionally conceived perceptual objects. The phenomenological 'reduction' of the causal source of perceptual objects does not reduce objects to lower-level perceptual representations (e.g. the geometrical shape of a visual object, the pitch contour of a sound object). On the contrary, phenomenology has traditionally turned towards 'the thing itself'. As Merleau-Ponty writes:

The form of objects is not their geometrical shape: it stands in a certain relation to their specific nature, and appeals to all our other senses as well as sight. The form of a fold in linen or cotton shows us the resilience or dryness of the fibre, the coldness or warmth of the material. Furthermore, the movement of visible objects is not the mere transference from place to place of coloured patches which, in the visual field, correspond to those objects. In the jerk of the twig from which a bird has just flown, we read its flexibility or elasticity, and it is thus that a branch of an apple-tree or a birch are immediately distinguishable. One sees the weight of a block of cast iron which sinks in the sand, the fluidity of water and the viscosity of syrup. In the same way, I hear the hardness and unevenness of cobbles in the rattle of a carriage, and we speak appropriately of a 'soft', 'dull', or 'sharp' sound.³⁷

We experience aspects of the same physical 'thing' in many senses and it is these high-level properties of the object (their 'specific nature') that are immediately accessible in perception. This agrees with the view of reverse hierarchies in sensory processing, suggesting a fast pre-attentive organization of the perceptual field into gross object categories while reverse processing allows us to scrutinize sensory features in more detail. Early representations of object features may be multisensory,³⁸ and properties of the 'sound object' are abstracted across the senses even though we may only experience it in sound, as Merleau-Ponty also points out. In particular, recent research has suggested a tight coupling between auditory and motor representations that allows us to im-

37 M. Merleau-Ponty *Phenomenology of Perception* (New Jersey: Routledge & Kegan Paul Ltd. 1945/1962), p. 229. See also L. Windsor, "Using auditory information for events in electroacoustic music", *Contemporary Music Review* 10 (1994).

38 C.E. Schroeder & J. Foxe, "Multisensory contributions to low-level, 'unisensory' processing", *Current Opinion in Neurobiology* 15 (2005).

mediately grasp the action involved manipulating an object from the sound it makes and to understand the possible use of the object in behavior (its 'affordance').³⁹

Auditory cognitive neuroscience today is only beginning to understand more abstract properties of sound perception. A large number of studies in the past decades have examined brain networks involved in the perception of traditional musical parameters (melody, rhythm, harmony, tonality), but less is still known about the mechanisms involved in recognizing timbre or sonorous features of musical sounds. Whether or not a particular kind of music is based on pitch structures, we spontaneously hear musical sounds as 'soft', 'dull' or 'sharp' as we do with objects in the physical world. We may recognize a melody as a structural whole as it unfolds over time, but musical timbre allows us to recognize a wealth of information within milliseconds.⁴⁰ Paradoxically, the neural mechanisms involved in recognizing abstract 'semantic' properties of musical sounds may tap into basic mechanisms in auditory processing reflecting the sensitivity of the auditory system towards sound objects.

Schaeffer launched an immensely important project in music research. He recognized a need to describe how sound objects appear in perception in order to allow composers to explore sounds as the material of musical ideas. However, it is unclear what the formal perceptual parameters of this description should be. In particular, it is not clear that 'sound objects' can be defined meaningfully without considering objects as physical things. New insights into the biology of the auditory system allows us to expand the Schaefferian project by considering the embodied context in which sound perception takes place and the ways in which the perceiving organism is oriented towards an environment.

39 J.E. Warren, R.J.S. Wise & J.D. Warren, "Sounds do-able: auditory-motor transformations and the posterior temporal plane", *TRENDS in Neurosciences* 28 (2005).

40 C. Krumhansl, "Plink: 'Thin slices' of music", *Music Perception* 27 (2010).