Sonic Object

35. Sonic Object Cognition

Rolf Inge Godøy

We evidently have features at different timescales in music, ranging from the sub-millisecond timescale of single vibrations to the timescale of a couple of hundred milliseconds, manifesting perceptually salient features such as pitch, loudness, timbre, and various transients. At the larger timescales of several hundred milliseconds, we have features such as the overall dynamic and timbral envelopes of sonic events, and at slightly larger timescales, also of various rhythmic, textural, melodic, and harmonic patterns. And at still larger timescales, we have phrases, sections, and whole works of music, often lasting several minutes, and in some cases, even hours.

Features at these different timescales all contribute to our experience of music, however the focus in the present chapter is on the salient features of what has been called *sonic objects*, meaning on holistically perceived chunks of musical sound in the very approximately 0.5–5 s duration range. A number of known constraints in the production and perception of musical sound as well as in human behavior and perception in general, seem to converge in designating this timescale as crucial for our experience of music.

The aim of this chapter is then to try to understand how sequentially unfolding and ephemeral sound and sound-related body motion can somehow be transformed in our minds to sonic objects.

35.1 Object Focus

The aim of this chapter is to present theories and tools for research on what we call *sonic objects*. The term *sonic object* can be defined as a fragment of musical sound, typically in the approximately 0.5–5 s duration range, a fragment perceived holistically as a coherent and somehow meaningful unit. A sonic object is arguably the most basic unit in musical experience, capable of making a rich set of perceptually salient sonic and multimodal features present in our minds.

A sonic object may encompass a single tone or chord, a short phrase of several tones and/or chords in succession, a single sound event (e.g., of hitting a tam-tam, of slamming a door, of breaking a bottle), or a more composite but still holistically perceived sound event (e.g., a rapid glissando on a harp or on a washboard, a burst of marbles rolling out onto the floor, a whirl of dry leaves in the wind). The duration limits of a sonic object are determined at one end by the minimal duration necessary to perceive salient features and at the other end by a maximal duration for perceiving the object as a singular and coherent entity, i. e., as not readily divisible into smaller parts.

The term *object* here serves to emphasize the perception of a sound fragment as a coherent entity,

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_35

as something present in our minds in an instant, in a *now-point*, to borrow the expression from *Edmund Husserl* [35.1–3], although sound is something that unfolds in time. How sequentially unfolding sound is transformed into somehow more solid auditory objects in our minds seems still to be largely enigmatic [35.4], however, the point of departure in this chapter is that such *flux-to-solid* transformations evidently do take place in auditory perception and that we may indeed denote salient auditory features by way of object-related concepts and metaphors. Focusing on the *cognition* of sonic objects means trying to provide tools for exploring perceptually salient features of sonic objects in music-related research and in practical contexts for composers, musicians and producers; in short, for anyone working with musical sound.

The idea of sonic objects is generally ascribed to the seminal work of *Pierre Schaeffer* et al. in the 1950s and 1960s and as emerging from practical work in electroacoustic composition of the so-called *musique concrète* [35.5–7]. Before the advent of tape recorders, composers would record sound fragments as loops on phonograph discs, enabling the mixing of sounds by lowering and raising the pickup arm, effectively turning any sound fragment on and off in a mix. Such a loop was called *closed groove* (*sillon fermé*) and when Schaeffer et al. listened to such loops innumerable times, they discovered that their perception of these sound fragments changed, that they tended to shift their attention towards more internal and subtle features of the sound itself, away from the original and more everyday and anecdotal significations of the sound fragments. They called this shifting of attention *reduced listening* (*écoute réduite*), signifying a shift toward perceptually salient sonic features, a shift of focus that eventually lead to a very extensive theory of sonic objects in *Schaeffer*'s monumental *Traité des objets musicaux* [35.5] and related publications [35.6, 7].

This origin is remarkable in that an extensive music theory grew out of practical composition work, and notably out of a radically new way of sound-based composition, very different from traditional note-based Western composition practice. This departure from Western music tradition drove the effort to make a more general and universally applicable music theory, based on a seemingly naïve questioning of the subjective listening experience, a kind of Socratic method of trying to distinguish the various perceptually salient features in auditory experience. This approach permeates the entire project of developing a new music theory and may be termed a *top-down* approach where the point of departure is the overall shapes of any sound, what we could call its *envelopes*, with regards to loudness, pitch features and timbral features and then successively differentiating more and more subfeatures of these main features (e.g., the distribution in time and spectrum of the energy, the amplitude and rate, regularity versus irregularity, of changes in these distributions, etc.) and then, sub-subfeatures of these, etc., progressively exploring more and more details in the sonic objects. It was only at a later stage that *Schaeffer* et al. realized the affinity of this method with classical phenomenological approaches to perception (referring to Husserl and Merleau-Ponty) of taking the subjective mental images of sound as point of departure for investigations, leading *Schaeffer* to remark that they were indeed *doing phenomenology without realizing it* [35.5, p. 262].

The next step of this research program was then seen as establishing *correlations* between the subjective features and the acoustic substrates of the sound, a long-term endeavor that needed to take into account the many nonlinear relationships between the physical signals and the subjective percepts, relationships that were characterized by *anamorphosis*, or *warping*. It should be remembered that psychoacoustics as we know it today was quite different in the 1950s and 1960s when Schaeffer et al. developed their theories and that research in this area was dominated by a more *scientistic* attitude as *Michel Chion* has put it [35.7, p. 30], of regarding human auditory perception as flawed, unreliable and often distorting the *real* features of sound. The attitude of Schaeffer et al. was quite different in naming subjective perception of sound the most important tool for research, as the point of departure for extensive and systematic explorations of musical sound, and only at a later stage going on to map out the correlations between subjective perception and the acoustic signals.

The term *cognition* in the title of this chapter implies a focus on the perceptual output of any sound-producing process, be that in sound synthesis, effects processing, composition, improvisation, or any kind of instrumental or vocal performance. The idea of sonic object cognition is also linked with the method of *analysis-by-synthesis*, meaning exploring something by active creation (and/or imagining) of incrementally different variants in view of finding out what the perceptually salient features (or ingredients) of any sonic object are. Lucidly presented by *J.-C. Risset* as a strategy for exploring timbral features in digital sound synthesis [35.8], this was actually also a strategy of Schaeffer et al. in exploring the contribution of different feature dimensions to the overall subjective impressions of sonic objects (for an instructive example of this, listen to [35.6, CD 2, tracks 90–95]). In other words, the idea of sonic object cognition here denotes a two-stage process of first trying to differentiate (and give names to) what seem to be subjectively salient features of a sonic object and then produce, or imagine, a number of variants of this sonic object where these features are somehow incrementally varied, so as to enable systematic testing of the effect of these features in subjective experience.

In short, sonic object cognition encompasses the capacity to think analytically and practically about fragments of sound in musical contexts and should aspire to combine insights from classical work (e.g., from composition and orchestration) with recent findings (e.g., from musical acoustics, music technology and music perception areas). With such a multitude of elements converging on sonic object cognition, we shall in this chapter first elaborate more on what is meant by the expression *sonic object*, try to demarcate what it is and what it is not, with some considerations of ontology: sonic objects are evidently associated with our experiences of sound, but sound is ephemeral, manifest here and now before vanishing, yet we fortunately usually have some memory trace of what we just heard. Are sonic objects then just as much mental images and in that case, what is the content of such mental images? One answer that has emerged from our own and others' research is that images of sonic objects are closely associated with images of body motion, so we shall then take a look at what is called *motor theory*, as well as associated issues of timescales and duration thresholds in musical experience. This leads to questions of how sonic objects emerge from musical experience by *chunking*, as well as how they are generated and classified. On this background, the remaining sections of this chapter shall focus on features of sonic objects as well as on how we may explore such features by the abovementioned analysis-by-synthesis strategy, before some reflections on prospects and challenges of research on sonic object cognition.

35.2 Ontologies

Sonic objects, defined as fragments of musical sound in the approximately 0.5–5 s duration range, will in most cases have multiple significations and multiple features, i. e., be several things at once. To enhance our knowledge of sonic objects, as well as to avoid misunderstandings, we need to clarify *what-is-what*, or clarify the *ontologies*, of sonic objects. As a first step, we may take a look at Schaeffer's analysis of listening, an analysis that is the basis for the subsequent ontological differentiations of sonic objects. Briefly stated, Schaeffer distinguished four components in listening (see [35.9, pp. 129–133], for a more extensive presentation):

1. *Listen* (*écouter*), which denotes the basic capacity for discerning different sounds in our environment.
2. *Hear* (*ouïr*), i. e., the basic physiological capacity for sensing sound.
3. *Hark* (*entendre*), meaning the intentional focus on some sound(s).
4. *Understand* (*comprendre*), denoting the transition from basic listening to understanding the significations of the sounds.

In everyday situations, these four components may variably be active or not, e.g., if I hear the squeaking of a door when I am expecting a visitor, I would understand this sound as indicating the arrival of my visitor and probably not be so much interested in the sonic features of the squeak. But being sensitive to such squeaking noises, I could direct my attention towards the squeak, making a note that I should lubricate the hinges to get rid of the squeak, or I could even be intrigued by the brass instrument-like timbral features of the squeak, its pitch and dynamical envelope; in short, start to focus on the squeak as a sonic object. Shifting my attention away from the contextual and/or causal signification of the squeak to its sonic features, is typically an act of the *reduced listening* mentioned earlier.

Furthermore, that we, in the case of electroacoustic music as well as in several other listening situations, only hear the sound and don't see the source of the sound, was by Schaeffer called *acousmatic listening*, alluding to the account of Pythagoras allegedly hiding behind a screen when teaching so that his pupils should not be distract by seeing him. The principle of acousmatic experience signifies a general divide between the production and the perception of musical sound, encouraging us to focus on the actually perceived features and to disregard whatever generative scheme is behind the sounds that we hear.

This divide between production and perception could also be seen as a critique of some prominent 20th century Western music that advocated generative schemes (e.g., dodecaphony, integral serialism, various algorithmic composition schemes and also more recent sonification schemes). It could be generalized to a model consisting of a *control space* and a *morphology space* as was suggested by *morphodynamical* theory [35.9–11]. The control space is concerned with the control variables input to any generative process, e.g., the oscillator frequencies, amplitudes, envelope shapes and modulation index in a frequency modulation (FM) synthesis model, and the morphology space is concerned with the perceptual output of any model, e.g., the timbral output of the FM synthesis model. As anyone familiar with FM synthesis probably has experienced, there may sometimes be a linear, seemingly coherent relationship between the input variables and the perceived output features, however, there may also often be a nonlinear relationship between incremental changes in control input values and perceived changes in output features, e.g., that small input value changes may result in disproportionately large and seemingly uncorrelated changes in output features.

Such discrepancies between changes in control input and perceived output may be a general risk when transferring features from one domain to another: it is not obvious that we readily perceive the relationships between any two domains, e.g., when sonifying visual or numerical information [35.12]. Such discrepancies of input and output are then a matter of mistaken ontological identity, encountered in various Western 20th century compositional schemes when an ordering scheme from one domain is uncritically transferred to another domain [35.9], and may be generalized as mistaken *mappings*. Being careful in making perceptually pertinent mappings is one of the main issues of interactive sonic design and of new instruments of musical expression [35.13], and such mappings should be based on careful analysis of the salient perceptual feature dimensions involved in sonic objects, not on more abstract numerical relationships.

Furthermore, given the acousmatic principle of disregarding the source of any sonic object, the next stage becomes an ontological analysis of the sonic object as such. First, we can list what Schaeffer indicated that the sonic object is *not* [35.7, pp. 34–35]:

- The sonic object is not the sounding body.
- The sonic object is not the physical signal.
- The sonic object is not a fragment of a recording.
- The sonic object is not a symbol notated in a score.
- The sonic object is not a state of the mind.

Furthermore, Schaeffer made an extensive differentiation of salient features that may be present in parallel in sonic objects (more on this later), noting that our intentional focus may wander from one feature or set of features to another feature or set of features, thus making it difficult or even futile to try to pin down exactly what a sonic object might be at any moment. *Schaeffer*'s conclusion after various ontological considerations and differentiations was that the sonic object is an *intentional unit*, meaning that it has several aspects, both sequentially and in parallel, that are kept together in our minds by our active mental focus [35.5, p. 263].

In the decades after the publication of Schaeffer's *Traité* in 1966, various psychoacoustic and neurocognitive research has focused on what we could broadly call *auditory object perception*. Work on so-called *auditory scene analysis* [35.14] has demonstrated a number of low-level signal features that contribute to our perceptual judgments of sonic objects, such as spectral coherence and qualitative discontinuities, documenting the effects of various gestalt principles, as well as the interaction of these low-level features with more high-level schema-based factors in our ability to discern sonic objects in listening [35.15]. Other recent neurophysiological research seems to confirm the basic gestalt-related principle of so-called *exclusive allocation* [35.16], suggesting similarities with visual domain object criteria, something that is also pursued in mapping out the different brain processes assumed to be involved here, as well as suggesting that there are cross-modal elements at work in auditory object perception [35.4]. This last point is of particular interest to our subsequent sections of multimodal features of sonic objects, in particular body motion and haptic features that clearly help us to grasp the ephemeral features of sound as more solid objects in our minds.

This research on auditory objects may be summarized as concerning the question of *coherence* in auditory perception and as also suggesting that there are both similarities and links between the different sense modalities at work here. Although there are still many unanswered questions, one candidate for such sonic object coherence is the motor schemas involved in auditory perception, an idea that has been much in focus in the so-called *motor theory of perception* and which will be an important element in the present chapter.

35.3 Motor Theory

The gist of the motor theory of perception is that we mentally (and sometimes even overtly) simulate body motion associated with whatever it is that we are perceiving, e.g., that when we hear ferocious drumming we mentally simulate energetic hand motion, or when we hear soft string music, we mentally simulate slow protracted bowing.

In our context of sonic object cognition and Schaeffer's research strategy of disregarding sources and ordinary significance, it is also clear that Schaeffer's strategy does not exclude a motor theory approach to how we perceive and classify sonic objects [35.17]. The point of motor theory is that sensations of body motion and postures are fundamental and ubiquitous in human cognition, hence also is the basis for sonic object perception, meaning that sonic objects may be perceived in terms of body motion and body posture schemas and not only as *pure sound*.

There are several links between sound and motion in music and we have in the past decades seen a growing number of research projects and publications in this area (see Chap. 38 this volume, and [35.18] for overviews). We have in the past decades also seen a general trend in the direction of understanding body motion as essential in most areas of human cognition [35.19], such as in social contexts [35.20] and also in abstract thought [35.21]. Motor cognition may be seen as integral to several modalities, e.g., in vision with active simulated tracing of what we are seeing [35.22], or in hearing with active simulation of the sound-producing vocal apparatus motion as suggested by the motor theory in linguistics [35.23, 24]. The motor theory perspective has been progressively better supported with the arrival of so-called noninvasive brain observation methods [35.25–27]. Also various behavioral studies have clearly demonstrated the strong links between body motion and sound perception, e.g., the effects of seeing lip motion on the perceived sound in the so-called *McGurk effect* [35.28].

What is essential here is the overall energy and motion trajectory images associated with sonic objects, meaning that such images are generic and transferable from one sound source to another as long as the overall envelopes are reasonably similar. This is the deeper and more general understanding of the reduced listening: departing from the everyday and anecdotal source and signification (e.g., crashing car, breaking twig, marble tumbling down a staircase, etc.) to a more generic energy-trajectory envelope [35.17].

The basis for mental presence of the sonic object could be called *motor-mimetic cognition* [35.29–31]. The idea of motor-mimetic cognition is that any feature of sound, and of musical experience in general, can be traced as a shape. This is a strong version of motor theory and one that can be used here to explore features of sonic objects. Specifically, we see this motor-mimetic behavior in sound imitating activities, such as in onomatopoeias and in *beatboxing*, *scat singing*, but also clearly evident in so-called *air instrument performance*, where people with little or no musical training seem to be able to simulate sound-producing body motion [35.32] and in so-called *sound-tracing*, where people are asked to trace the body motion shape they believe best reflect salient features of sonic objects they are hearing [35.33].

We have various kinds of music-specific body motion, all of which may variably contribute to the motor images of sonic objects; body motion that may be categorized as follows [35.18]:

- *Sound-producing* body motion. This includes various kinds of *excitatory motion* such as hitting, stroking, bowing and blowing, as well as *modulatory motion* that modifies the sound, such as the left hand motion on string instruments, the closing and opening of mutes on brass instruments, etc. In addition, various kinds of so-called *ancillary motion* belong in this main category: motion that is not strictly sound-producing but still necessary in order to avoid fatigue or strain injury, or to help shaping the musical expression and also *communicative motion* in relation to fellow musicians or the audience.
- *Sound-accompanying* body motion. This includes dancing, walking, gesticulating, tracing, etc., often reflecting sound-producing body motion, but often more vaguely the energy and affect suggested by the musical sound.

Also, there are a number of more global, i.e., nonlocal, music-related body motion sensations, such as calm–agitated, fast–slow, smooth–rugged, regular–irregular, etc., where there are also often correlations with affective features. Hence, we should next consider some elements of timescales and duration thresholds in relation to sonic objects.

35.4 Timescales and Duration Thresholds

It is well known that we have a timescale range from approximately 20–20 000 events per second for perceiving pitch as well as stationary timbral and loudness features of sounds [35.34]. In our context of sonic objects, the sub-20 Hz region is of particular interest, as this is where we have the nuances and fluctuations of pitch, loudness and timbre, including microtextural elements such as trills and tremolos, as well as various slower elements such as envelopes of pitch, timbre and loudness of tones, and sensations of motion (gait) and rhythmical and melodic patterns.

Also in the sub-20 Hz region we have feature duration thresholds, meaning how long an excerpt we have to hear in order to get some impression of salient features. It has been suggested that we may perceive salient features of sound in fragments as short as 350 ms [35.35], but for other musical features such as rhythmical, melodic and harmonic patterns, we have to hear longer fragments, i. e., approaching the upper limit for sonic object duration of approximately 5 s. The question then becomes that of trying to see what is *inherent* in the sound as necessary minimum duration for the recognition of salient features and what is more a matter of the peculiarities of our perceptual and cognitive processing. Hence, we should try to determine duration thresholds from two points of view:

1. That which captures most salient information, i. e., neither too short nor too long, e.g., sufficiently long for perceiving salient basic dynamic, pitch-related and timbral features, as well as more high-level features of rhythm, texture, melody, harmony, style, genre, sense of gait and affect, but not so long as to seem being redundantly repetitious.
2. That which is intrinsically in accordance with our cognitive constraints and/or predispositions. This includes what are typical human action durations, optimal durations in motor control, durations of short-term memory stores and of attention spans and timescales for awareness.

Both points of view seem to be relevant and they sometimes seem to converge with respect to duration, perhaps by adjustment of salient information timescales in human communication with ecological attention timescales, as suggested by *Ernst Pöppel* [35.36]. What seems to be clear is that we have rather different qualitative features at different timescales, and for this reason we have in our own research on sonic objects and music-related body motion applied the following main timescale categories:

- *Micro timescale*, typically in the sub-0.5 s range and including continuous and quasistationary features such as pitch, dynamics and stationary timbral features. At this timescale, we may also have some fast fluctuations in the sound such as rapid tremolos and trills, what we shall with Schaeffer call *grain*.
- *Meso timescale*, typically in the 0.5–5 s range, encompassing whole sonic objects. This timescale includes most salient musical features such as rhythmical, textural, melodic, timbral, etc. patterns, readily enabling identification of genre, style, sense of gait and affect.
- *Macro timescale*, typically in the above-5 s range, including whole phrases, sections and movements and usually consisting of concatenations of sonic objects of the meso timescale. The macro timescale is also that of narratives or dramaturgy, however, it seems not to be so well researched in terms of actual perceptual features.

One crucial feature of the meso timescale is that sequences of elements (tones, sounds) at this timescale give rise to sonic objects with qualitative new features not present at smaller timescales, e.g., a melodic contour is present at the meso timescale as a sonic object because it is *present as one extended and coherent object with a shape and not primarily as a sequence of individual tones*. Similarly, body motion at this meso timescale is also readily perceived and conceived as forming coherent motion objects with shapes, and details of such object formation are found in the topic of chunking.

35.5 Chunking

We may think of music as a continuous stream of auditory, visual, proprioceptive, haptic, etc. sensations, yet it seems that we make sense of such streams by segmenting them into somehow meaningful units, into what we call *chunks*. This process of chunking is the basis for the perception of sonic objects, hence one of the main topics in this chapter. Chunking here means not just a segmentation of a continuous stream, but also a *transformation* of otherwise sequential events or entities into qualitative new and larger units, making the formerly single events/entities totally fuse, or *disappear*, into the new chunk unit in the sense of belonging exclusively to this new unit by the so-called *exclusive allocation* principle of gestalt theory [35.14].

There are two main approaches to chunking in music, *endogenous* and *exogenous* [35.37]:

- *Endogenous chunking* is based on qualitative discontinuities in the signal such as that between sound and silence, between different pitches, timbres or levels of loudness or various repeated patterns, e.g., metrical, melodic or harmonic. Yet this may not always work, either because of insufficient discontinuities, e.g., protracted and unchanging or slowly changing sounds, or because of competing discontinuities, e.g., in a series of staccato tones alternating between sound and silence, necessitating the projection of endogenous chunking schemas onto the sounds.

- *Exogenous chunking* is based on the projection of schemas onto what we perceive, schemas based on what we have acquired in previous experience and in particular, various motor schemas acquired from experiences of sound and motion production.

In cases of exogenous chunking, i. e., when applying schemas from previous experience onto sensory input, we can see that such schemas are often constraint-based in their origin. We should thus consider some constraints at work in traditional (i. e., pre-electronic) means for sound production. This includes various features of the instrument and room acoustics (reverberation) that may shape our experience of chunks, e.g., by incomplete damping of sounds resulting in a smearing that glues otherwise distinct sonic events together in qualitative new and larger chunks. But we also have constraint-based chunking originating in our organism, both of attention [35.36], and related to motor control [35.38, 39]. Specifically, we have the following chunk-inducing constraints in musical performance:

- *Phase transition*: There are thresholds of grouping dependent on duration and rate of events [35.40], typically causing singular events to fuse into higher-level units with increase in event rate (e.g., singular impulses fuse into a tremolo with acceleration). Conversely, units may split into lower-level units with decrease in event rate (e.g., a tremolo split into a series of impulses with deceleration).
- *Coarticulation*: The contextual fusion of otherwise distinct motion and sound by the effector (finger, hand, arm, vocal apparatus) always being in a context of having just made some sound and is about to make another sound in the immediate future and that there is a corresponding contextual smearing of the resultant sound because of incomplete damping between the sound events [35.41].
- *Goal points*: That human motion control is goal directed [35.42], typically resulting in so-called *key-postures* at salient moments in time, i. e., at downbeats and other accented points in the music, resulting in chunks being centered on these goal points [35.43].

It seems that various research converges in documenting motion chunking in human behavior in general and rhythmic gestalts in particular [35.39]. As noted by *Klapp* et al. in connection with polyrhythmic patterns [35.44, p. 318]:

> *The limitation to only one motor Gestalt may be analogous to limits that arise with visual patterns such as the Necker cube. That figure can be perceived in only one of its configurations at any given instant. In either configuration, however, all of the lines of the cube are perceived simultaneously as one pattern. Thus, the Gestalt is not restricted in terms of the number of lines that can be perceived. Instead, the limit is that only one organization can be activated. Similarly, the limit in concurrent motor actions is assumed not to lie in the number of muscles that can be controlled, but, instead, the limit is that only one action pattern can be active.*

In other words, it seems that even rather complex patterns of motion may be conceived and perceived as a chunk in motor control.

In summary, we may conclude that sonic objects represent the convergence of chunking factors from the signal (constraints and qualitative discontinuities in both sound and motion), the sense of motion (internal sense of effort and proprioception, i. e., action gestalts, goal points, coarticulation, phase transitions, etc.) and attention constraints; hence, the convergence of both the exogenous and the endogenous factors of chunking in musical experience.

35.6 Sound Generation

In the beginning of the *musique concrète*, the sound fragments used were typically so-called *found objects*, meaning ready-made fragments of sound from the environment, (human, animal, mechanical, etc., but also from more conventional instrumental or vocal sources), hardly processed beyond the initial cutting before the subsequent concatenations into musical compositions. As for the boundaries (i. e., start and stop points) of these fragments, Schaeffer noted that these could be either *natural* in the sense of occurring at some qualitative discontinuity in the sound, e.g., between sound and silence, or they could be *artificial* in the sense of being cut out of a context. In the latter case, the effect of the cutting would in turn become an integral part of the sonic object, e.g., as a steep attack on an otherwise smooth, sustained sound.

Although the acousmatic attitude and the strategy of reduced listening encourages us to focus on intrinsic perceptual features of sound, we know that such features are often the result of peculiarities of the sound source. We should then also consider some aspects of sound generation in general, including more traditional instrumental and vocal sound generation as well as sound synthesis and processing, which contribute to our perception of sonic objects. As suggested by auditory scene analysis research [35.14], we tend to perceive sounds as emanating from a source because of spectrotemporal coherence, i. e., that the spectral components tend to fuse into one sonic object because of low-level gestalt principles such as synchrony of onsets and motion of partials. Yet also, there are cases when schemas from past experience come into play in identifying sound sources, i. e., where there is great spectral variation over the entire pitch range and different playing modes of an instrument (e.g., the full range of a piano), yet where the instrument is usually perceived as *the same* in spite of such significant differences.

Also, we should remember that motor theory suggests that production schemas are projected onto what we hear, largely independent of the source and it may be argued that this also applies to acousmatic listening [35.17]. The point is that motor theory in general applies to any perceptual feature by its relation to body motion kinematics, haptics, proprioception and effort, something that was in fact suggested also by Schaeffer in the three basic dynamic envelopes of sonic objects, regardless of source:

1. *Sustained*: Continuous transfer of energy from the body to the instrument such as in bowing or blowing, resulting in a more or less stable and continuous sound.
2. *Impulsive* A spike or sudden peak of effort such as in hitting or kicking, resulting in a sudden attack in the sound followed by a longer or shorter decay.
3. *Iterative*: A rapid back and forth motion as in a tremolo or a trill, resulting in fast ripple-like features in the sound.

These main categories of sound-producing motion and the corresponding resultant sounds are subject to phase transitions dependent on rate and duration of the events: shortening a sustained motion turns it into a impulsive motion, and conversely, lengthening an impulsive motion turns it into a sustained motion, and slowing down an iterative motion turns it into a series of impulsive motions, etc.

Additionally, there are also body postures projected onto the sound by way of past knowledge of sound-producing motion, e.g., narrow positioning of hands versus spread positioning of hands on a keyboard, or vocal tract shapes of open, closed, pointed, etc. The motor theory perspective suggests that in general, there is a projection of assumed body motion onto whatever it is that we are hearing and that this in particular goes for overall sense of effort, including sense of velocity, acceleration and jerk in the music; in short, that music has what could be called rich gestural affordances [35.45], and that all these sensations of body motion evoked by the sound contributes to our perception and subsequent *objectification* of the music.

In summary: Schemas of sound production may reflect basic body motion schemas, regardless of the sound source in question, meaning that sonic object cognition is largely determined by body motion shapes, in turn based on a number of constraints. In the case of very long and stationary or slowly evolving sounds, the energy envelopes may tend to be perceived as *non-human* in that there are no rests in effort and also that the duration may be above the usual thresholds of attention, c.f. *Pöppel*'s suggestions of the perceptual present and the tendency to shift attention in experiences lasting significantly longer than approximately 3 s [35.36].

35.7 Constraints and Idioms

Needless to say, the scope of musical expression is vast, if not infinite, but there are also a number of constraints in musical sound production. The physics of musical instruments and the ergonomics of sound production together make up a number of constraints on music making, constraints that in turn also influence our perceptions of sonic objects.

The basic scheme of sound production is that of energy transfer from the human body to the instrument by way of hitting, stroking, plucking, bowing, blowing, etc. and the response of the musical instrument to any such excitatory body motion. The energy dissipation of an instrument is then a shaping factor of sonic objects, providing its overall envelopes of pitch, timbre and loudness, as well as a number of transients and fluctuations in the course of the sonic object. Additionally, the room acoustics, e.g., the reverberation and other resonant features, contribute to the features of the sonic object. The patterns of excitation and subsequent energy dissipation are highly characteristic of different instruments and people seem to possess quite extensive knowledge about these patterns [35.46].

The combined physical and ergonomic constraints of any instrument typically result in what is commonly called *idioms*. Idioms are sound-producing motions, on any instrument or the human voice, that are considered particularly well-sounding and also comfortable or easy to perform, hence based on successful combinations of instrument physics and ergonomics. Idioms may variably so contribute to images of sonic objects and are of particular importance in the more composite sonic objects we find in orchestration (more on this later).

The most basic constraint of music performance is that *all sonic events are embedded in action trajectories* and that there is no immediate transition from one sound-producing event to another, i. e., that all effector motion (by fingers, hands, arms, lips, tongue, feet) takes time and that all effector motion (regardless of speed) makes for continuous trajectories. In sum, we have the following constraints of sound-producing body motion contributing to sonic object cognition:

- *Effort constraints*, i. e., there are limits to endurance and musicians need to take rests when performing to avoid fatigue and strain injury.
- *Speed constraints*, related to the previous constraint, but partly also a matter of motor control.
- *Phase transitions*, i. e., the transitions based on rate, speed and duration of basic motion-sound categories mentioned earlier.
- *Coarticulation*, based on the fact that human motion cannot be instantaneous and that there always is a transition from one effector position to another effector position, hence there always are both so-called *spillover effects* and *anticipatory effects* in performance, leading to a contextual smearing of both motion and sound.
- *Goal points*, meaning that performance motion is organized hierarchically and that these hierarchies are also (variably so) reflected in the shaping of the sonic objects.

In summary, there are a number of constraints on traditional sound production in music, constraints that can be summarized as concerning envelopes of energy and as forming schemas that also seem to extend to electronic music.

35.8 Sound Synthesis

Although available technologies for sound synthesis and sound processing present great possibilities for generating a large variety of sonic objects, they also present substantial challenges of how to control the parameters involved in this generation (see Chap. 38 for more on this). The main point in our context is then how to access perceptually salient sonic object features in sound synthesis and processing. (The division between *synthesis* and *processing* may not always be so clear cut, e.g., convolving two signals may be a case of cross-synthesis or source-filter synthesis, but may also be a case of effects processing, e.g., adding reverb to a sound or simulating various other room acoustic coloring of sound.)

This is initially much dependent on the type of synthesis model being used. The main classification into so-called *signal models* and *physical models* may be useful here: physical models shall in principle incorporate some element of energy input (excitation) and energy dissipation (resonance), albeit often in a highly simplified form. On the other hand, the signal models are inherently more abstract in the sense that there is really no simulation of real-world excitation and resonance except for what we may put into the control parameters, e.g., by designating envelopes that approximate expected behavior of real-world sonic events such as of plucking strings, hitting membranes, or stroking metal plates.

Compared with traditional musical instruments and the human voice, there are then no transfers of energy, no motion constraints and and no idioms, involved in sound synthesis (although for the expert listener there may be detectable peculiarities of the synthesis models being used, in a sense resembling *idioms*, e.g., of the often-heard FM and granular synthesis models). This does not preclude that we may project various schemas from past musical experiences onto new kinds of sonic objects (so-called *anthropomorphic projection*), or that the composition is organized more in accordance with traditional score arrangements of sounds (as is sometimes the case in multitrack music production). What seems to be an inherent tendency in our perception is then that of grouping sounds into events and sources as if they were originating from real-world sources and that we project various motor schemas onto sonic objects that have really no body motion element in their origins.

However, the desire to introduce more nuances and expressiveness into the domain of sound synthesis and processing has in the past couple of decades lead to much effort in developing new means for real-time control, including new interfaces of various kinds (e.g., New Interfaces for Musical Expression (NIME) conferences), as well as extending traditional musical instruments with new control possibilities. The latter development has the advantage of exploiting musicians' already acquired skills, as well as to explore the sonic territory stretching from that of traditional instruments into new sounds.

Besides the ergonomics, one main issue here is that of mapping, i. e., of how body motion is assigned to the control parameters of whatever sound synthesis or processing model is being used. Different schemes of mapping have been explored [35.13], however one crucial question here is actually that of trying to determine what are the perceptually salient feature dimensions of sonic objects and how these may be correlated with body motion features.

35.9 Feature Taxonomy

Schaeffer's work suggests that any sound whatsoever, natural or synthesized, may be considered a sonic object provided it fulfills the criteria of having a suitable duration and capable of being perceived as a holistic and coherent entity. The challenge then becomes that of developing some kind of conceptual apparatus that enables us to diagnose salient perceptual features of sonic objects, i. e., to differentiate these features and compare their relative variable values (e.g., degree of dynamic fluctuation, degree of stability of perceived pitch, etc.).

The project of developing a feature taxonomy becomes a multistage process of firstly discerning and naming some feature that we believe is salient in our subjective image of a sonic object, and secondly to distinguish further what may be subfeatures of this main feature and sub-subfeatures of these features, etc., as far as we see useful for our analytic and/or practical purpose at any time. This means applying a number of metaphorical labels and try to define what they refer to in the sonic object (e.g., the fast dynamic fluctuation as the *grain* of the sonic object). The next step will then be to see these metaphor-labeled features as dimensions in a multidimensional feature space, where each dimension may have a minimum and a maximum value (e.g., between minimum and maximum amplitude of the grain fluctuations) and where any sonic object may be positioned in such a multidimensional feature space.

Such a scheme is very general, applicable to any sonic object and in no way limited to our traditional Western music theory. It is not constrained by a symbolic notation system and all features may be taken into consideration. In this sense, Schaeffer's feature taxonomy is concerned with *concrete* and not with *abstract*, features [35.7, p. 39] and [35.9, p. 216]. There are of course categories at work here, e.g., for pitch and for timbre, but the point is that the feature space of sonic objects is much more extensive than that afforded by traditional Western music theory and its associated categories. As we know, e.g., Western categories of pitch are not universal and other musical cultures (e.g., in India) have more nuanced concepts for pitch. In this sense, the symbolic versus subsymbolic divide is bypassed by Schaeffer's taxonomy and we could add also the symbolic versus suprasymbolic divide as well, in the sense that several fused tone events (or sound events in the case of nonpitched sounds) may very well form a new and larger sonic object with its own overall salient features. In other words, this may include features beyond those of mainstream analysis (including music information retrieval (MIR) and its so-called chroma vectors (i. e., pitch extraction) and other pattern retrieval schemes), features that so far are not well captured by symbolic queries but which are highly significant in musical experiences, e.g., phrase shapes and other meso level emergent features.

In the decades since the publication of Schaeffer's *Traité* in 1966, there has been significant progress in detecting and representing a large number of previously unnamed features in psychoacoustics [35.47, p. 326]. Also software development, e.g., like the *MIRtoolbox* and the *Timbre Toolbox* and *Praat*, has made it much easier now to work directly with sonic object features such as spectral shapes, spectral centroid, spectral flux, formantic shapes, etc.

However, as argued earlier, sonic object cognition will not only include acoustic features, but also corresponding motion features, i. e., trajectories, contours, postures, etc.; that is, various perceptually salient features that may be correlated with acoustic features, as was the long-term goal of Schaeffer's research program.

The main principle of Schaeffer's feature taxonomy is that any feature whatsoever may be given a shape-related metaphor labeling. This gives us a very general and systematic conceptual apparatus for dealing with features in a lucid manner, as well as bridging the gap towards quantitative data of sound and motion.

35.10 Shape Cognition

Sound, as well as associated sensations of body motion, are ephemeral, yet may evidently also result in more solid mental images in our memory. That we now have the technologies for capturing, replaying and making various representations of the ephemeral is indeed one of the most significant developments of our times, as it was, albeit to a much more modest extent, in the late 1940s with the beginning of the *musique concrète*. However, these possibilities of freezing sound and motion also raises new questions of what kind of knowledge we may represent by these various technology-based schemes, and what are the fundamental knowledge goals of our research, questions that point to *shape cognition* as a very basic epistemological and also quite pragmatic paradigm for representing sonic object features.

The term *shape cognition* designates our capacity for thinking and representing sound and motion features visually, as geometric objects in two or more dimensions and as distinct from symbolic representations. As for the symbolic representations, ranging from common-practice Western notation to more recent schemes such as MIDI (musical instrument digital interface) and MusicXML (music extensible markup language), one major challenge is to overcome the discrete bias that prevents us from having good notions of sonic objects, i. e., we need representations that highlight the actual continuous sonic unfolding and make the individual Western pitches and durations fused into coarticulated, coherent sonic objects, as well as to represent the many significant nuances of musical sound. We have in the course of recent decades seen some interesting attempts to more closely represent the holistic features of sonic objects as shapes:

- Enhanced spectral representations of actual musical sound, from the pioneering work of *Robert Cogan* several decades ago [35.48] to the presently readily available software, e.g., the earlier mentioned toolboxes.
- Animations of various kinds that represent the unfolding of sonic objects, be that based on spectral images, on MIDI data, or conventional notation.
- Subjective sketches (also with a signal basis e.g., in the *Acousmographe* software) and various graphical scores [35.49].
- Various body-motion-based shape renderings of sonic objects, including the mentioned *sound tracings* [35.33] and *air instrument performances* [35.32], as well as possibilities for extracting shape information from music-related video and motion capture material of musical performances and other music-related body motion [35.50].

The overall rationale here is that any feature may in principle be traced as a shape, hence making even the most ephemeral sensation *solidly* present for scrutiny and further differentiation. Morphodynamical theory has given us an extensive and well-reflected basis for shape cognition [35.9–11, 51]. In the words of *René Thom* [35.51, p. 6]:

> [...] *the first objective is to characterize a phenomenon as shape, as a spatial shape. To understand means first of all to geometrise.*

In sum, shape cognition offers a top-down approach to sonic object design in that we start with the overall shapes for loudness, pitch- and timbre-related features, something that we following Schaeffer may call the *typology* of sonic objects, then going on to subfeatures and sub-subfeatures of these shapes and into the more internal features, what we may call the *morphology* of the sonic object.

However, shape cognition and shape representation is a long-term project where we have substantial challenges of developing schemes for graphics, animations, feature choice, etc. But for now, whenever confronted with some feature(s) of sonic objects we want to study, we can start by drawing them subjectively, with pencil on paper or just mentally and then successively formalize such sketches into graphs and animations.

35.11 Typology and Morphology of Sonic Objects

A good point of departure for more systematic shape cognition of sonic objects could be the typological and morphological classification scheme of Schaeffer. The *typology* can be summarized as denoting the overall shape, or envelope, of any sonic object, with regards to its dynamic (loudness), and timbre- and pitch-related content. The *morphology* is on the other hand primarily concerned with that which goes on within these overall shapes. Schaeffer's typology and morphology scheme implements a kind of *form-content* distinction, with the advantage of being a general, flexible, top-down differentiation scheme applicable to any sonic object, regardless of origin.

The typology can be seen as a first sorting of sonic objects in that the perceptually most prominent features are taken into account. The typology starts with the three general dynamic envelope categories we have presented earlier and that correspond clearly with body motion categories:

- *Sustained*
- *Impulsive*
- *Iterative*.

This overall dynamic envelope classification is followed by a similarly coarse classification with regards to pitch:

- *Tonic stable*, meaning having a clearly perceivable pitch and that the pitch is relatively stable during the course of the sonic object.
- *Tonic varying*, meaning having a clearly perceivable pitch, but that this changes during the course of the sonic object, e.g., has an upward glissando.
- *Nontonic*, meaning inharmonic or noise-based sounds without any clearly perceivable pitch.

This results in a 3×3 typological matrix, a first and rather coarse but still useful classification of sonic objects. It should also be mentioned that prior to this typology, some other criteria of selection have been applied, i. e., first of all of length and of variability, criteria that are summarized in the notion of the *suitable object* [35.17].

Moving on to the more *internal* features of the sonic object, these are classified according to the principles of the morphology of Schaeffer's theory, here with two main dimensions of the morphology:

- *Gait*: Slower fluctuations in the sound, e.g., as made by a series of slower repeated tone onsets on an instrument, by slower opening–closing of mutes on brass instruments, by slow up and down glissandi on string instruments – in short, any motion in the sound slower than that which is typically perceived as tremolos or trills.
- *Grain*: Fast fluctuations in the sound.

All kinds of grain fluctuations within the sonic objects can be differentiated with our own terms:

- Tremolo, active shake, e.g., as on a violin with rapid back-and-forth bow motion or with *flatterzunge* on a brass instrument.
- Tremolo, passive response, e.g., washboard, maracas, or the grainy sound of a deep double bass or deep bassoon tone.
- Vibrato and/or trills, possible on several instruments.
- Spectral flux by active modulation made e.g., by rapid opening–closing of mutes on brass instruments.
- Spectral flux as passive response to excitation on various instruments, e.g., by scraping a tam-tam with a metal rod.

Note that different terminologies are possible and that the main point is thinking about objects holistically and qualifying their feature dimensions top-down, from the overall to the microlevel. This also goes for Schaeffer's concept of *mass*, i. e., spectral content of sonic objects, and this may be related to psychoacoustic elements, such as the following:

- Spectral shape of any kind, both quasistationary and changing in the course of the sonic object
- Spectral centroid, or the *focus* of the mass
- Correlating body posture shapes with spectral shapes, both of the vocal apparatus and other effectors.

There are a number of categories, subcategories, sub-subcategories, etc., here, but the important thing to keep in mind is that this is all in view of being a practical tool for sonic diagnosis, i. e., it should be considered a *questionnaire* and not a *balance sheet* [35.7, pp. 92–93].

35.12 Singular, Composed, Composite and Concatenated Objects

Although one of the key features of a sonic object is that it should be perceived and conceived holistically as a unit, it is also possible to consider variants of sonic objects not only along feature dimensions (as in the preceding section on the typology and morphology), but also by adding and/or expanding sonic objects. For this reason, it could be useful to make distinctions between singular and more complex sonic objects, distinctions that are in fact often encountered in orchestration and multitrack music production:

- Sonic objects may be *singular*, i. e., one basic and *simple* sound, e.g., of a tone from a piano, a flute, or a plucked string. As we know, such sonic events may in themselves contain several components, typically attack transients in the beginning of the sound, various fluctuations in the course of the sound and in the decay parts.
- Singular sonic objects may also be expanded by superposition, typically by adding other sonic objects to the attack point, resulting in what Schaeffer called *composed*, sonic objects. This is often found in both instrumental and electroacoustic composition, e.g., as in the example from Lutoslawski's *Jeux Venitiens* in Fig. 35.1a, where there is a loud percussive attack object added to the soft attack of the sustained string instrument chord.
- Sonic objects may be extended in the temporal direction, i. e., temporally smeared by adding prefixes and/or suffixes, what Schaeffer called *composite* sonic objects, frequently encountered as various ornaments and other figures in different kinds of music and also associated with coarticulation, i. e., the fusion of rapid tone events and sound-producing motion into higher-level sonic objects [35.41]. In Fig. 35.1b, we see an example of this from Messian's *Regard de la Vierge* where a rush of demisemiquavers and the final quavers fuse together to one single composite sonic object.
- Sonic objects may of course be *concatenated* to form longer stretches of music, as is evident in various kinds of collage-type composition, not only in electroacoustic music and disc jockey scratching music, but just as much in more traditional instrumental music (e.g., Schnitke, Maxwell-Davies, etc., or it may be more homogeneous, e.g., as in works by Messiaen, Stravinsky, etc.). This is what happens at what we above defined as the macro timescale and this is also an area that is not well researched in the field of music perception, but it would not be unreasonable to assume that there are new emergent contextual effects at the macro timescale not present at the meso and micro timescales.

In general, questions about what happens in long stretches of music are not addressed by classical sonic object research. From the cognitive science research referred to above, we could guess that attention is not constant, that it may tend to fluctuate, similar to gestalt flips when looking at the Necker cube, as suggested by *Ernst Pöppel* [35.36]. What does seem to be both quite clear and also feasible to document in sound- and body motion-based research, are the overall global features of longer stretches as done by MIR software (e.g., global indicators of loudness, harmonicity, roughness, spectral flux, etc.) and in body motion software (quantity of motion, mean square jerk, etc.), features that probably are quite good indicators of overall subjective sensations of effort and affect in musical experience.

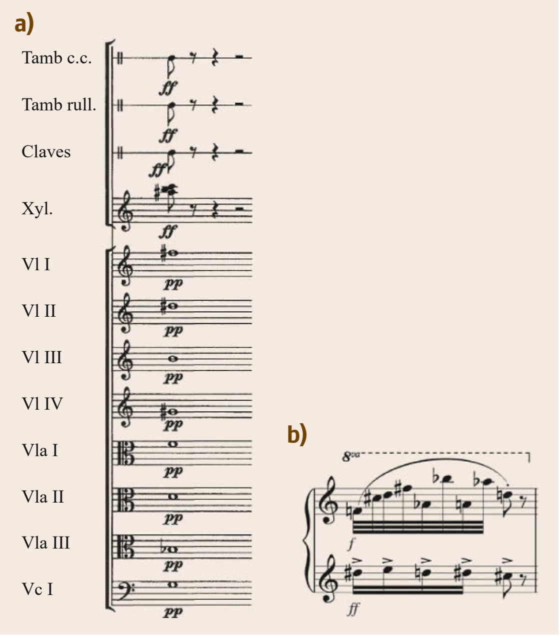

Fig. 35.1 (a) A *composed* sonic object from the first movement of Lutoslawski's *Jeux Venitiennes* consisting of a sustained and soft string chord with an added loud percussion attack. (b) A composite sonic object from Messian's *Regard de la Vierge* where there is a coarticulated prefix with a rush of demisemiquavers towards the final quavers (reproduced with permission from [35.31])

35.13 Textures, Hierarchies, Roles and Translations

The term *texture* can here be defined as the distribution of sound components within any sonic object. All sonic objects have internal textures, but notably, these textures may vary between totally stationary sounds (e.g., perfectly harmonic sounds) and complex, inharmonic, noise-dominated, or *chaotic* sounds, with most sonic objects found somewhere between these extremes. This is a wider definition of texture than usually encountered in more traditional music analysis, however, the traditional textural categories of homophony, polyphony, heterophony and monody, may be included in this definition as subcategories. This wide definition of texture, besides being in line with the morphological principles of Schaeffer, also reflects the theories and aesthetics of *Xenakis* of various statistical distributions of sound grains [35.52], and also reflects the textural focus in other late-20th century compositions e.g., by Ligeti and Lutoslawski.

The point with the wide definition of texture is to accommodate textural features of any sonic object and also to see that textures may sometimes be explicitly made (e.g., as in a tremolo on a piano) or may sometimes be emergent features of a singular sound, e.g., as in the case of the tremolo-like grain sound of a deep double bass tone. This view of texture is again very much in line with the acousmatic principle of Schaeffer's theory of disregarding the source and focusing on the perceived sonic object features.

Texture encompasses patterns in the sub-20 Hz region and may in speed range from very rapid tremolos and trills (i. e., close to the approximately 20 Hz continuous sound threshold) of *grain*, to the very slow, tentatively down to 30 BPM [35.53], and even slower of *gait*, ultimately also to completely stationary sounds.

In addition to the mentioned categories of grain and gait, there are various subcategories and dimensions, mainly specifying various kinds of spectral distributions and patterns of change. We could think of texture as situated in a multidimensional feature space with each dimension extending from some minimum to some maximum value, e.g., as in the following:

- Dense–spread
- Thick–thin
- Synchronous onsets–asynchronous onsets
- Short tones–long tones
- Many sustained tones–few or no sustained tones
- Wet–dry
- Little or no melodic movement–much melodic movement
- Small intervals in melodic lines–large intervals in melodic lines etc.

A division between foreground melody and background accompaniment is needless to say a very common texture in Western music. Added to this is the phenomenon of simultaneous different speeds in many musical textures, i. e., that the foreground is moving faster than the background, however, there are also deviations from this ordering. What is necessary then is to make an analysis of what could be called *role hierarchies* in sonic objects (and in longer musical excerpts as well). Such an analysis may typically reveal:

- Foreground, i. e., that which is the focus of attention at any moment.
- Background, i. e., that which is more in the periphery of attention.
- Role groups, meaning that complex sonic textures may often be composed of a smaller number of main roles (e.g., foreground + background), but that these main roles in turn are divided into subroles, as in the example in Fig. 35.2 below.
- In some cases, also totally fused, e.g., *Ligeti's Atmosphères* with the entire ensemble dedicated to one

Fig. 35.2 An excerpt from Nikolai Rimsky-Korsakov's *Golden Cockerel Suite* containing different textural roles. See main text for details

singular sonic object, where the role of each instrument is more like that of being a partial in a complex spectrum.

Roles differentiated into subroles, sub-subroles, etc., are in cases of what may be considered successful orchestration, matched with suitable instrumental idioms. As an example, consider the fragment from the opening of Nikolai Rimsky-Korsakov's *Golden Cockerel Suite* in Fig. 35.2. Here we can observe a clear role organization:

- The foreground is made up by the English horn, transferred to the oboe in the third measure, doubled in the harp, the harp changing to harmonics in the second measure. This foreground may be regarded as a series of *impulsive* sounds and at that, *composite* with the attacks in the English horn doubled with the plucked attacks in the harp.
- Against this ascending foreground, we have a descending chromatic line in the flute, celesta and violin I and II, a line of basically sustained sounds that are in fact also a series of *composite* sonic objects with the sustained flute sounds as the core, with the celesta as attacks coloring and with the tremolos in the violins adding a *grain* texture to this.
- The *pp* in the strings enhances the white noise nature of string sound, making the *pp* tremolo on the cymbal, also with a soft white noise cloud, melt into this hushed background of the strings. Together, strings and cymbal clearly have the role of sustained sound, equivalent to adding a *wet* reverb signal to the sound.

In summary, we have in this excerpt three main roles and these main roles have in turn subroles, each of which matches the instrumental idioms well and, on the other hand, results in a rich and colorful orchestral sound as a whole.

To further study textural elements in music, we should also consider transfers of a musical idea/intention from one setting to another, as is done in orchestration and arrangements and which may be collectively called *musical translation*. But what is a musical idea and what is the gist of that which is translated? Besides a practical matter of making and evaluating different arrangements or orchestration versions of tunes or other musical compositions, translating sonic objects from one setting to another is also an encouragement to make us reflect on textural features and role functions in music, including on the role of idioms in music. As we know, it is usually more difficult to translate highly idiomatic expressions from one language to another and attempts at making a more literary translation, as close as possible word-by-word to the original, will often lead to quite strange, or even ridiculous, results. A translation that renders the more general meaning will usually be considered more useful as well as more true to the intentions of the original expression. Similar considerations apply in music, meaning that a task of translation could proceed as follows:

1. An analysis of textural elements
2. Assignments of textural elements to roles, including voice leading
3. Assignments of roles to instruments according to optimal match with idioms
4. Combinations of roles into textures, also taking optimal acoustic (spectral) distribution into account.

In sum, sonic object cognition may enhance our capacity to think analytically about orchestration or instrumentation from a combined perceptual-generative (including motor perspective) point of view, combining the best from the classics, e.g., the main principles of Rimskij-Korsakoff of idiom use, of spectral distribution, fusion, voice-leading, etc., with present schemes for sonic object feature analysis and synthesis, as well as more systematic research on similarity in music.

35.14 Analysis-by-Synthesis

Analysis-by-synthesis can be defined as the systematic exploration of features at all timescales, micro, meso, and macro, by producing and perceptually evaluating a series of sonic objects with incrementally different features. With a foundation in *J.-C. Risset*'s ideas [35.8], we may expand the idea of analysis-by-synthesis to include any activity where people are somehow engaged in producing a number of variant versions of a sonic object and evaluating these variant sonic objects in view of finding the most appropriate, suitable, pleasing, well-sounding, etc. sonic objects for the occasion or task at hand. This will then mean that most musical practice, rehearsal, studio production and composition work has some element of analysis-by-synthesis in that there is a trial-and-error, incremental tweaking of parameters going on. Analysis-by-synthesis is then de facto a familiar phenomenon in musical practice, however, this may have eluded much of music theory and analysis.

The strategy of analysis-by-synthesis applied to sonic objects is then a process with the following steps:

1. Determining some feature dimension(s)
2. Incrementally varying the value(s) of this feature's dimensions
3. Evaluating the subjective perceptual results of these variations.

It may turn out that incremental value changes are perceived as just that, as a gradual change in some hue, but it may also turn out that an incremental change suddenly results in a very distinct qualitative change. Examples of this are encountered in sound synthesis, e.g., the bowed-plucked transition with a gradual decrease in the attack time duration, or a significant change in timbre with FM synthesis.

In summary, the working strategy here is to explore categories, categorical thresholds, intercategorical and intracategorical variation of sonic objects through such analysis-by-synthesis, thus progressively building up our knowledge of sonic objects.

35.15 Summary

The focus on sonic objects, both in the works of Pierre Schaeffer and various later research [35.46, 54], is motivated by the belief that the sonic object is a highly significant element of musical experience, that it has a privileged status in our understanding of music and as a tool for music creation as well. The point is that in spite of significant advances in musical acoustics, psychoacoustics, music cognition, music technology and music information retrieval during the last couple of decades, there are still many unnamed salient features in musical sound and we definitely still need the top-down approach as was presented by Schaeffer et al. half a century ago. In other words, the concept of *object* has great potential as the core of now fast growing insights from a number of music-related research areas.

Needless to say, there are serious challenges of representation, however there is work going on that progressively gives us better signal-based representations, i. e., more selective and focused visual representations that enable in-depth, systematic explorations of sonic object features as shapes. In particular, various graphical and animation representations of combined sonic and body motion features are promising in offering us insights into how rich aesthetic sensations emerge from the ephemeral.

References

35.1 E. Husserl: *On the Phenomenology of the Consciousness of Internal Time, 1893–1917* (Kluwer, Doredrecht 1991), English translation by John Barnett Brough

35.2 R.I. Godøy: Thinking now-points in music-related movement. In: *Concepts, Experiments and Fieldwork. Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Peter Lang, Frankfurt 2010) pp. 245–260

35.3 R.I. Godøy: Sound-action awareness in music. In: *Music and Consciousness*, ed. by D. Clarke, E. Clarke (Oxford Univ. Press, Oxford 2011) pp. 231–243

35.4 T.D. Griffiths, J.D. Warren: What is an auditory object?, Nat. Rev. Neurosci. **5**(11), 887–892 (2004)

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

35.6 P. Schaeffer: *Solfège de l'objet Sonore* (INA/GRM, Paris 1998), with sound examples by Reibel, G. and Ferreyra, B. 1967

35.7 M. Chion: *Guide des Objets Sonores* (INA/GRM Buchet/Chastel, Paris 1983)

35.8 J.-C. Risset: Timbre analysis by synthesis: Representations, imitations and variants for musical composition. In: *Representations of Musical Signals*, ed. by G. De Poli, A. Piccialli, C. Roads (MIT Press, Cambridge 1991) pp. 7–43

35.9 R.I. Godøy: *Formalization and Epistemology* (Scandinavian Univ. Press, Oslo 1997)

35.10 J. Petitot: *Morphogenèse du Sens I* (Presses Universitaires de France, Paris 1985)

35.11 J. Petitot: *Forme in Encyclopædia Universalis* (Encyclopædia Universalis, Paris 1990)

35.12 A.R. Jensenius, R.I. Godøy: Sonifying the shape of human body motion using motiongrams, Empir. Musicol. Rev. **8**(2), 73–83 (2013)

35.13 A. Hunt, M. Wanderley, M. Paradis: The importance of parameter mapping in electronic instrument design, J. New Music Res. **32**(4), 429–440 (2003)

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

35.15 J.K. Bizley, Y.E. Cohen: The what, where and how of auditory-object perception, Nat. Rev. Neurosci. **14**, 693–707 (2013)

35.16 I. Winkler, T.L. van Zuijen, E. Sussman, J. Horváth, R. Näätänen: Object representation in the human auditory system, Eur. J. Neurosci. **24**(2), 625–634

(2006)
35.17 R.I. Godøy: Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus, Organ. Sound **11**(2), 149–157 (2006)
35.18 R.I. Godøy, M. Leman: *Musical Gestures: Sound, Movement and Meaning* (Routledge, New York 2010)
35.19 V. Gallese, T. Metzinger: Motor ontology: The representational reality of goals, actions and selves, Philos. Psychol. **16**(3), 365–388 (2003)
35.20 M. Wilson, G. Knoblich: The case for motor involvement in perceiving conspecifics, Psychol. Bull. **131**(3), 460–473 (2005)
35.21 V. Gallese, G. Lakoff: The brain's concepts: The role of the sensory-motor system in conceptual knowledge, Cogn. Neuropsychol. **22**(3/4), 455–479 (2005)
35.22 A. Berthoz: *Le Sense du Mouvement* (Odile Jacob, Paris 1997)
35.23 A.M. Liberman, I.G. Mattingly: The motor theory of speech perception revised, Cognition **21**, 1–36 (1985)
35.24 B. Galantucci, C.A. Fowler, M.T. Turvey: The motor theory of speech perception reviewed, Psychon. Bull. Rev. **13**(3), 361–377 (2006)
35.25 J. Haueisen, T.R. Knösche: Involuntary motor activity in pianists evoked by music perception, J. Cogn. Neurosci. **13**(6), 786–792 (2001)
35.26 E. Kohler, C. Keysers, M.A. Umiltà, L. Fogassi, V. Gallese, G. Rizzolatti: Hearing sounds, understanding actions: Action representation in mirror neurons, Science **297**, 846–848 (2002)
35.27 M. Bangert, E.O. Altenmüller: Mapping perception to action in piano practice: A longitudinal DC-EEG study, BMC Neuroscience **4**, 26 (2003)
35.28 H. McGurk, J. MacDonald: Hearing lips and seeing voices, Nature **264**, 746–748 (1976)
35.29 R.I. Godøy: Imagined action, excitation and resonance. In: *Musical Imagery*, ed. by R.I. Godøy, H. Jorgensen (Swets and Zeitlinger, Lisse 2001) pp. 237–250
35.30 R.I. Godøy: Motor-mimetic music cognition, Leonardo **36**(4), 317–319 (2003)
35.31 R.I. Godøy: Images of sonic objects, Organ. Sound **15**(1), 54–62 (2010)
35.32 R.I. Godøy, E. Haga, A.R. Jensenius: Playing air instruments: Mimicry of sound-producing gestures by novices and experts. In: *GW2005, LNAI 3881*, ed. by S. Gibet, N. Courty, J.-F. Kamp (Springer, Berlin, Heidelberg 2006) pp. 256–267
35.33 K. Nymoen, R.I. Godøy, A.R. Jensenius, J. Torresen: Analyzing correspondence between sound objects and body motion, ACM Trans. Appl. Percept. **10**(2), 9 (2013)
35.34 B.C.J. Moore: *Hearing* (Academic, San Diego 1995)
35.35 R. Gjerdingen, D. Perrott: Scanning the dial: The rapid recognition of music genres, J. New Music Res. **37**(2), 93–100 (2008)
35.36 E. Pöppel: A hierarchical model of time perception, Trends Cogn. Sci. **1**(2), 56–61 (1997)
35.37 R.I. Godøy: Reflections on chunking in music. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (Peter Lang, Frankfurt 2008) pp. 117–132
35.38 N. Hogan, D. Sternad: On rhythmic and discrete movements: Reflections, definitions and implications for motor control, Exp. Brain Res. **181**, 13–30 (2007)
35.39 S.T. Klapp, R.J. Jagacinski: Gestalt principles in the control of motor action, Psychol. Bull. **137**(3), 443–462 (2011)
35.40 H. Haken, J. Kelso, H. Bunz: A theoretical model of phase transitions in human hand movements, Biol. Cybern. **51**(5), 347–356 (1985)
35.41 R.I. Godøy: Understanding coarticulation in musical experience. In: *In: Sound, Music and Motion Lecture Notes in Computer Science*, ed. by M. Aramaki, M. Derrien, R. Kronland-Martinet, S. Ystad (Springer, Berlin 2014) pp. 535–547
35.42 S.T. Grafton, A.F. Hamilton: Evidence for a distributed hierarchy of action representation in the brain, Hum. Mov. Sci. **26**, 590–616 (2007)
35.43 R.I. Godøy: Quantal elements in musical experience. In: *Sound–Perception–Performance. Current Research in Systematic Musicology*, ed. by R. Bader (Springer, Berlin, Heidelberg 2013) pp. 113–128
35.44 S.T. Klapp, J.M. Nelson, R.J. Jagacinski: Can people tap concurrent bimanual rhythms independently?, J. Motor Behav. **30**(4), 301–322 (1998)
35.45 R.I. Godøy: Gestural affordances of musical sound. In: *Musical Gestures: Sound, Movement and Meaning*, ed. by R.I. Godøy, M. Leman (Routledge, New York 2010) pp. 103–125
35.46 D. Rocchesso, F. Fontana: *The Sounding Object* (Edizioni di Mondo Estremo, Firenze 2003)
35.47 U. Zölzer: *DAFX: Digital Audio Effects* (Wiley, Chichester 2011)
35.48 R. Cogan: *New Images of Musical Sound* (Harvard Univ. Press, Cambridge 1984)
35.49 B. Schäffer: *Introduction to Composition* (PWM, Warsaw 1976)
35.50 A.R. Jensenius: Some video abstraction techniques for displaying body movement in analysis and performance, Leonardo: J. Int. Soc. Arts Sci. Technol. **46**(1), 53–60 (2013)
35.51 R. Thom: *Paraboles et Catastrophes* (Flammarion, Paris 1983)
35.52 I. Xenakis: *Formalized Music* (Pendragon, Stuyvesant 1992)
35.53 L. Van Noorden: The functional role and bio-kinetics of basic and expressive gestures in activation and sonification. In: *Musical Gestures: Sound, Movement and Meaning*, ed. by R.I. Godøy, M. Leman (Routledge, New York 2010) pp. 154–179
35.54 F. Delalande, M. Formosa, M. Frémiot, P. Gobin, P. Malbosc, J. Mandelbrojt, E. Pedler: *Les Unités Sémiotiques Temporelles: Éléments Nouveaux d'analyse Musicale* (Marseille, Édition MIM-Documents Musurgia 1996)