Prospective View on Sound Synthesis BCI Control in Light of Two Paradigms of Cognitive Neuroscience

4

Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad,
Jean-Arthur Micoulaud-Franchi and Jean Vion-Dury

Abstract

Different trends and perspectives on sound synthesis control issues within a cognitive neuroscience framework are addressed in this article. Two approaches for sound synthesis based on the modelling of physical sources and on the modelling of perceptual effects involving the identification of invariant sound morphologies (linked to sound semiotics) are exposed. Depending on the chosen approach, we assume that the resulting synthesis models can fall under either one of the theoretical frameworks inspired by the representational-computational or enactive paradigms. In particular, a change of viewpoint on the epistemological position of the end-user from a third to a first person inherently involves different conceptualizations of the interaction between the listener and the sounding object. This differentiation also influences the design of the control strategy

---

M. Aramaki (✉) · R. Kronland-Martinet · S. Ystad
Laboratoire de Mécanique et d'Acoustique (LMA), CNRS UPR 7051, Aix-Marseille University, Centrale Marseille, 31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
e-mail: aramaki@lma.cnrs-mrs.fr

R. Kronland-Martinet
e-mail: kronland@lma.cnrs-mrs.fr

S. Ystad
e-mail: ystad@lma.cnrs-mrs.fr

J.-A. Micoulaud-Franchi · J. Vion-Dury
Laboratoire de Neurosciences Cognitives (LNC), CNRS UMR 7291, Aix-Marseille University, Site St Charles, 3, Place Victor Hugo, 13331 Marseille Cedex 3, France
e-mail: jarthur.micoulaud@gmail.com

J. Vion-Dury
e-mail: jean.vion-dury@ap.hm.fr

E.R. Miranda and J. Castet (eds.), *Guide to Brain-Computer Music Interfacing*,
DOI 10.1007/978-1-4471-6584-2_4

enabling an expert or an intuitive sound manipulation. Finally, as a perspective to this survey, explicit and implicit brain-computer interfaces (BCI) are described with respect to the previous theoretical frameworks, and a semiotic-based BCI aiming at increasing the intuitiveness of synthesis control processes is envisaged. These interfaces may open for new applications adapted to either handicapped or healthy subjects.

4.1 Introduction

In this article, we present different approaches to sound synthesis and control issues and describe how these procedures can be conceptualized and related to different paradigms within the domain of cognitive neuroscience. A special emphasis is put on the notion of intuitive control and how such a control can be defined from the identification of signal invariants obtained both from the considerations of the physical or signal behaviour of the sound-generating sources and the perceptual impact of the sounds on the listeners.

Since the first sounds were produced by a computer in the late 1950s, computer-based (or synthesized) sounds have become subject to an increasing attention for everyday use. In early years of sound synthesis, the majority of applications were dedicated to musicians who learned to play new instruments that generally offered a lot of control possibilities, but required high skills to operate. Due to increasingly powerful computers, new applications linked to communication, virtual reality and sound design have made sound synthesis available for a broader community. This means that synthesis tools need to be adapted to non-expert users and should offer intuitive control interfaces that do not require specific training. The construction of such intuitive synthesis tools requires knowledge about human perception and cognition in general and how a person attributes sense to sounds. Why are we for instance able to recognize the material of falling objects simply from the sounds they produce, or why do we easily accept the ersatz of horse hooves made by the noise produced when somebody is knocking coconuts together? Is the recognition of sound events linked to the presence of specific acoustic morphologies that can be identified by signal analysis? In the approach presented here, we hypothesize that this is the case and that perception emerges from such invariant sound structures, so-called *invariant sound morphologies*, in line with the ecological approach of visual perception introduced by (Gibson 1986). From a synthesis point of view, this theoretical framework is of great interest, since if enables the conception of perceptually optimized synthesis strategies with intuitive control parameters.

Sound synthesis based on the modelling of physical sources is generally divided in two main classes, i.e. physical models and signal models. Physical models aim at simulating the physical behaviour of sound sources (i.e. the physical origin of sounds), while signal models imitate the recorded signal using mathematical representations without considering the physical phenomena behind the sound production. In the case of physical models, an accurate synthesis can only be achieved when physical phenomena linked to the sound production are well described by physics. This is not the case for complex sources (e.g. natural phenomena such as wind, rain, fire, etc.). In the case of signal models, any sound can generally be perfectly resynthesized for instance from the analysis of real sounds, independently of the complexity of the underlying physical phenomena of the sound source. However, the control of such sounds is a difficult issue due to the large number of synthesis parameters that generally are implied in such models and to the impossibility to physically interpret these parameters. The physical and signal models can also be combined to form so-called hybrid models (e.g. Ystad and Voinier 2001). The control of these models requires an expertise and the quality judgment of the control is based on an error function linked to the physical or signal precision between the model and the real vibration. Such controls necessitate a scientific expertise apart from certain cases such as musical applications where the control parameters correspond to physical values controlled by the musician (e.g. pressure, force, frequency, etc.). In this latter case, the musical expertise enables the control.

To propose efficient synthesis models that enable intuitive control possibilities, synthesis models combined with perceptual considerations have been developed lately. Perceptual correlates have been sought by testing the perceptual relevance of physical and/or signal parameters through listening tests (cf. Sect. 4.3.2). In the case of environmental sounds, we have identified such perceptually relevant sound morphologies through several experiments. These experiments have made it possible to identify sound elements, also described as sound “atoms”, specific to given sound categories that enable definition of high-level control parameters for real-time synthesis applications. Such synthesis tools allow users to synthesize auditory scenes using intuitive rather than reflective processes. Intuitive processes appeal on intuition which is a kind of immediate knowledge, which does not require reasoning, or reflective thought. Intuition can also be defined as the knowledge of an evident truth, a direct and immediate seeing of a thought object (Lalande 1926). The quality of the control strategy is in this case based on perceptual judgments and on easily understandable control parameters on the user interface. Therefore, we call this synthesis control, *intuitive control*.

When searching for perceptually relevant sound morphologies, the understanding of attribution of sense of sounds becomes essential. This issue is a natural part of a more general research field called semiotics that consists in studying the general theory of signs. The notion of signs has been addressed since antiquity by the stoic philosophers (Nadeau 1999). Classically, semiotics is divided in syntax, semantics and pragmatics. Semiology is a part of semiotics, which concerns the social life, and dynamic impact of signs, as language (Nadeau 1999). For de Saussure, language constitutes a special system among all semiological facts. In linguistics, for de Saussure, a sign is the association of a signifier (acoustic image) and a signified (the correlated concept) linked together in a consubstantial way (de Saussure 1955). This consubstantial relationship is often difficult to understand. Semiotics span over both linguistic and non-linguistic domains such as music, vision, biology, etc. This means that it is possible to propose a semiotic approach of sounds, without referring to linguistic semiology. Like in de Saussure construction of signs, one can postulate that every natural (environmental) or social sound is linked to the afferent concept in the same consubstantial way. For example, if I hear a bell, I immediately know that it is a bell, and perhaps, but not always, I even manage to imagine the size of the bell, depending on its spectral contents. Except for “abstract sounds”, i.e., sounds for which the sources cannot be easily identified, one can say that each sound can be considered as a non-linguistic sign whose origin can be described using language, in a reflective thought. Previous studies have shown that the processing of both linguistic and non-linguistic target sounds in conceptual priming tests elicited similar relationships in the congruity processing (cf. Sect. 4.5). These results indicate that it should be possible to draw up a real semiotic system of sounds, which is not the linguistic semiology, because phonemes can be considered only as particular cases of sounds.

So far, the identification of signal invariants has made it possible to propose an intuitive control of environmental sounds from verbal labels or gestures. An interesting challenge in future studies would be to propose an even more intuitive control of sound synthesis processes that bypasses words and gestures and directly uses a BCI that records electroencephalographic signals in a BCI/synthesizer loop. This idea is not new and several attempts have already been made to pilot sounds directly from the brain activity. In (Väljamäe et al. 2013), the authors made an exhaustive review in the field of EEG sonification in various applications (medical, neurofeedback, music, etc.) and concluded that the type of mapping strategy strongly depends on the applications. For instance, in the case of musical applications, the mapping is generally determined by artistic choices and does not necessarily mirror a strict semiotic relation. The intuitive BCI-controlled synthesizer that we aim at is intended for a generic context and should enable the identification of brain activity linked to specific signal morphologies that reflect the attribution of sense to a sound.

This paper is organized as follows. In Sect. 4.2, the methodology that leads to intuitive sound synthesis is viewed in the light of representational-computational and enactive perspectives. Then, in Sect. 4.3, two sound synthesis approaches are described and related to the previously presented perspectives. In Sect. 4.4, different control strategies emanating from the different synthesis approaches are described. In Sect. 4.5, some results from experiments supporting the existence of semiotics for non-linguistic sounds are presented. Finally, in Sect. 4.6, a prospective view on a control strategy for synthesis processes based on a BCI is proposed.

4.2 Two Conceptions on the Way We Interact with the Surrounding World

Sound synthesis that integrates perceptual effects from the morphology of their signal in order to enable intuitive control to the end-user brings forward the following questions: How do I attribute a meaning to a perceived sound (related to the semiotics)? What effect does this sound have on me? These questions induce a change in our position with respect to the sound from a third-person position (observer) in more traditional synthesis approaches where only acoustic considerations are taken into account, to a first-person position (implied) in the perceptual synthesis processes. This corresponds to a change from a representational to a neurophenomenological point of view in the field of cognitive neuroscience (Varela 1996). We here adopt a similar change of viewpoint to investigate the phenomenon of sound perception as it was seminally studied in (Petitmengin et al. 2009).

Classically, in the standard paradigm of cognitive neuroscience, there is, on one hand, the physical object and on the other hand, the subject that perceives this object according to his/her mental representation of the physical reality. From this conception of representation proposed by Descartes, a representational-computational paradigm has been developed. This paradigm involves the existence of a *correct* representation of the physical world and assumes that the perception of the object is all the more adequate when the subject's mental representation matches the physical reality, considered as the reference (Varela 1989). Less classically, in the neurophenomenological paradigm of cognitive sciences, it is the interaction between the subject and the object, which enables the subject to perceive an object. F. Varela called this interaction: *enaction* (Varela 1989; Varela et al. 1991). In the enactive paradigm, the mind and the surrounding world are mutually imbricated. This conception is inspired from the phenomenological philosophy of Husserl, who called this interaction a noetic–noematic correlation (Husserl 1950). He posited that there was a link between intentional content on the one hand, and extra-mental reality on the other, such that the structure of intentionality of the consciousness informs us about how we perceive the world as containing particular objects. In a certain manner, and quite caricatured, the physical reality is no more the reference, and the subject becomes the reference. The perception of the object is all the more adequate when the subject's perception makes it possible to efficiently conduct an action to respond to a task. As Varela puts it (Varela et al. 1991):

> The enactive approach underscores the importance of two interrelated points: 1) perception consists of perceptually guided action and 2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided.

and concludes:

> We found a world enacted by our history of structural coupling.

In 1966, P. Schaeffer, who was both a musician and a researcher, published the "Traité des objets musicaux" (Schaeffer 1966), in which he reported more than ten years of research on electroacoustic music. He conducted a substantial work that was of importance for electroacoustic musicians. With a multidisciplinary approach, he intended to carry out fundamental music research that included both Concrete[1] and traditional music. Interestingly, he naturally and implicitly adopted a phenomenological approach to investigate the sound perception in listening

---

[1] The term "concrete" is related to a compositional method which is based on concrete material, i.e., recorded or synthesized sounds, in opposition with "abstract" music which is composed in an abstract manner, i.e., from ideas written on a score, and becomes "concrete" afterwards.

experiences. In particular, he introduced the notion of *sound object*. The proposition of P. Schaeffer naturally conducts the acoustician from the representational-computational paradigm to the enactive paradigm, since P. Schaeffer in line with the phenomenological viewpoint stresses the fact that sound perception is not only related to a correct representation of the acoustic signal. This is also coherent with later works of Varela and the conception of perception as an enactive process, where the sound and the listener constitute a unique imbricated system. The perception of the sound is modified by the intentionality of the subject directed towards the sound, which can induce an everyday listening, which is a source-oriented kind of listening, or musical (or acousmatic) listening, which involves the perception of the quality of the sound (Gaver 1993a, b). Thus, sound synthesis should not be limited to the simulation of the physical behaviour of the sound source. In other words, it is the sound object given in the process of perception that determines the signal to be studied, meaning that perception has to be taken into account during the signal reconstruction process.

In the work of P. Schaeffer, morphology and typology have been introduced as analysis and creation tools for composers as an attempt to construct a music notation that includes electroacoustic music and therefore any sound. This typological classification is based on a characterization of spectral (mass) and dynamical (facture) profiles with respect to their complexity and consists of 28 categories. There are nine central categories of “balanced” sounds for which the variations are neither too rapid and random nor too slow or non existent. Those nine categories include three facture profiles (sustained, impulsive or iterative) and three mass profiles (tonic, complex and varying). On both sides of the balanced objects in the table, there are 19 additional categories for which mass and facture profiles are very simple/repetitive or vary a lot. This classification reveals perceptually relevant sound morphologies and constitute a foundation for studies on intuitive sound synthesis.

Based on these previous theoretical frameworks from cognitive neuroscience, we suggest that the control of sound synthesis can be discussed in the framework of the representational-computational and the enactive points of view. In the approach inspired by the representational-computational framework, we consider that the user controls physical or signal parameters of the sound with the idea that the more actual (with respect to the physical reality) the parameter control, the better the perception. The physical or signal properties of sounds are considered as the reference for the sound control. In the approach inspired by the enactive framework, we consider that the user is involved in an interactive process where he/she controls the sound guided by the perceptual effect of his/her action. The idea is that the more recurrent (and intuitive) the sensorimotor manipulation, the better the perception. The sound control enables the perception to become a perceptually guided action. This is an enactive process because the sound influences the control effectuated by the subject and the control action modifies the sound perception. The sound as perceived by the subject is thus the reference for the sound control. Such enactive framework formed a theoretical basis for a recent research community centred on the conception of new human–computer interfaces (Enactive Network) and in a natural way, led to numerous interactive applications in musical contexts (*Journal of New Music Research, special issue* "*Enaction and Music*" 2009). A general review on fundamental research in the field of enactive music cognition can be found in (Matyja and Schiavio 2013).

4.3 Sound Synthesis Processes

To date, two approaches to synthesize sounds could be highlighted: sound synthesis based on the modelling of physical sources (from either physical or signal perspectives) and sound synthesis based on the modelling of perceptual effects. Interestingly, these synthesis approaches could be linked to the two paradigms related to our perception of the surrounding world (i.e. approaches inspired by the representational-computational and the enactive paradigms, cf. Fig. 4.2) described in the previous section.

4.3.1 Two Approaches for Sound Synthesis

4.3.1.1 Modelling the Physical Sources

In the case of sound synthesis based on the modelling of physical/vibrating sources, either the mechanical behaviour or the resulting vibration of the sound source is simulated.

Physical synthesis models that simulate the physical behaviour of sound sources can either be constructed from the equations describing the behaviour of the waves propagating in the structure and their radiation in air (Chaigne 1995) or from the behaviour of the solution of the same equations (Karjalainen et al. 1991; Cook 1992; Smith 1992; Bilbao 2009). Physical models have been used to simulate a large number of sound sources from voice signals to musical instruments. Several synthesis platforms based on physical modelling are now available, such as Modalys that is based on modal theory of vibrating structures that enable the simulation of elementary physical objects such as strings, plates, tubes, etc. These structures can further be combined to create more complex virtual instruments (http://forumnet.ircam.fr/product/modalys/?lang=en)n.d). Cordis-Anima is a modelling language that enables the conception and description of the dynamic behaviour of physical objects based on mass-spring-damper networks (http://www-acroe.imag.fr/produits/logiciel/cordis/cordis_en.htmln.d). Synthesis models for continuous interaction sounds (rolling, scratching, rubbing, etc.) were proposed in previous studies. In particular, models based on physical modelling or physically informed considerations of such sounds can be found (Gaver 1993a; Hermes 1998; van den Doel et al. 2001; Pai et al. 2001; Rath and Rocchesso 2004; Stoelinga and Chaigne 2007). In particular, Avanzini et al. (2005) developed a physically based synthesis model for friction sounds. This model generates realistic sounds of continuous contact between rubbed surfaces (friction, squeaks, squeals, etc.). The parameters of the model are the tribological properties of the contact condition (stiffness, dynamic or static friction coefficients, etc.) and the dynamic parameters of the interaction (mainly the velocity and the normal force). Also, a synthesis technique based on the modal analysis of physical objects (finite element modelling of each object for precomputation of shapes and frequencies of the modes) was proposed by (O'Brien et al. 2002) in the context of interactive applications. Note that this approach presents a limitation when the physical considerations involve complex modelling and can less easily be taken into account for synthesis perspectives especially with interactive constraints.

Signal synthesis models that simulate the resulting vibration of the sound source are based on a mathematical modelling of the signal. They are numerically easy to implement and can be classified in three groups as follows:

- Additive synthesis: The sound is constructed as a superposition of elementary sounds, generally sinusoidal signals modulated in amplitude and frequency (Risset 1965). For periodic or quasi-periodic sounds, these components have average frequencies that are multiples of one fundamental frequency and are called harmonics. The amplitude and frequency modulation (FM) laws should be precise when one reproduces a real sound. The advantage of these methods is the potential for intimate and dynamic modifications of the sound. Granular synthesis can be considered as a special kind of additive synthesis, since it also consists in summing elementary signals (grains) localized in both the time and the frequency domains (Roads 1978).
- Subtractive synthesis: The sound is generated by removing undesired components from a complex sound such as noise. This technique is linked to the theory of digital filtering (Rabiner and Gold 1975) and can be related to some physical sound generation systems such as speech (Flanagan et al. 1970; Atal and Hanauer 1971). The advantage of this approach is the possibility of uncoupling the excitation source and the resonance system. The sound transformations related to these methods often use this property to make hybrid sounds or crossed synthesis of two different sounds by combining the excitation source of a sound and the resonant system of another (Makhoul 1975; Kronland-Martinet 1989).
- Global (or non-linear) synthesis: The most well-known example of such methods is audio FM. This technique updated by Chowning (1973) revolutionized commercial synthesizers. The advantages of this method are that it calls for very few parameters, and that a small number of numerical operations can generate complex spectra. They are, however, not adapted to precise signal control, since slight parameter changes induce radical signal transformations. Other related methods such as waveshaping techniques (Arfib 1979; Le Brun 1979) have also been developed.

In some cases, both approaches (physical and signal) can be combined to propose hybrid models, which have shown to be very useful when simulating certain musical instruments (Ystad and Voinier 2001; Bensa et al. 2004).

4.3.1.2 Modelling the Perceptual Effects

In the case of sound synthesis based on the modelling of perceptual effects, the sound generation is not merely based on the simulation of the physical or signal phenomena. This approach enables the synthesis of any kind of sounds, but it necessitates the understanding of the perceptual relevance of the sound attributes that characterize the sound category in question. Concerning environmental sounds, several studies have dealt with the identification and classification of such sounds (Ballas 1993; Gygi and Shafiro 2007; Gygi et al. 2007; Vanderveer 1979). A hierarchical taxonomy of everyday sounds was proposed by Gaver (1993b) and is based on three main categories: sounds produced by vibrating solids (impacts, deformation, etc.), aerodynamic sounds (wind, fire, etc.) and liquid sounds (drops, splashes, etc.). This classification related with the physics of sound events and has shown to be perceptually relevant. Hence, the perceptual relevance of these categories encourages the search for invariant sound morphologies specific to each category. This notion is developed in the next section.

4.3.2 Invariant Sound Morphologies

The invariant sound morphologies associated with the evocation of sound attributes can either be linked to the physical behaviour of the source (Giordano and McAdams 2006), to the signal parameters (Kronland-Martinet et al. 1997) or to timbre qualities based on perceptual considerations (McAdams 1999). This means that different synthesis approaches can be closely related, since in some cases, physical considerations and in other cases, signal variations might reveal important properties to identify the perceived effects of the generated sounds. In particular for environmental sounds, several links between the physical characteristics of actions (impact, bouncing, etc.), objects (material, shape, size, cavity, etc.) and their perceptual correlates were established in previous studies (see Aramaki et al. 2009; Aramaki et al. 2011 for a review). In summary, the question of sound event recognition was subject to several inquiries (e.g. Warren and Verbrugge 1984; Gaver 1993a, b) inspired by Gibson's ecological approach (Gibson 1986) and latter formalized by McAdams and Bigand (1993). This led to the definition of structural and transformational invariants linked to the recognition of the object's properties and its interaction with the environment, respectively.

Sounds from impacted objects: Impact sounds have been largely investigated in the literature from both physical and perceptual points of view. Several studies revealed relationships between perceptual attributes of sound sources and acoustic characteristics of the produced sound. For instance, the attack time has been related to the perception of the hardness of the mallet that was used to impact the resonant object, while the distribution of the spectral components (described by inharmonicity or roughness) of the produced sound has been related to the perceived shape of the object. The perceived size of the object is mainly based on the pitch. A physical explanation can be found in the fact that large objects vibrate at lower eigenfrequencies than small ones. Finally, the perception of material seems to be linked to the damping of the sound that is generally frequency-dependent: high frequency components are damped more heavily than low frequency components. In addition to the damping, the density of spectral components, which is directly linked to the perceived roughness, was also shown to be relevant for the distinction between metal versus glass and wood categories (Aramaki et al. 2009, 2011).

Sounds from continuous interactions: Based on previous works described in Sect. 4.3, invariant sound morphologies related to the perception of interactions such as rubbing, scratching and rolling were investigated (Conan et al. 2013a, b; Thoret et al. 2013). An efficient synthesis model, init. An efficient synthesis model, initially proposed by (Gaver 1993a) and improved by (van den Doel et al. 2001), consists in synthesizing the interaction sounds by a series of impacts that simulates the successive micro-impacts between a plectrum and the asperities of the object's surface. Therefore, it has been highlighted that a relevant sound invariant morphology allowing the discrimination between rubbing and scratching interactions was the temporal density of these impacts, i.e., the more (respectively, the less) impacts that occur, the more the sound is associated to rubbing (respectively, to scratching) (Conan et al. 2012). For the rolling interaction, it has been observed, from numerical simulations based on a physical model, that the temporal structure of the generated impact series follows a specific pattern. In particular, the time intervals between impacts and associated amplitudes are strongly correlated. Another fundamental aspect supported by physical considerations is the fact that the contact time of the impact depends on the impact velocity. This dependency also seems to be an important auditory cue responsible for the evocation of a rolling interaction (Conan et al. 2013).

These studies related to such interaction sounds led us to address the perceptual relation between the sound and the underlying gesture that was made to produce the sound. Many works highlighted the importance of the velocity profile in the production of a movement and its processing may be involved at different levels of perception of a biological movement both in the visual and in the kinaesthetic system ((Viviani and Stucchi 1992; Viviani et al. 1997; Viviani 2002) for a review). Based on these findings, we investigated whether the velocity profile, in the case of graphical movements, was also a relevant cue to identify a human gesture (and beyond the gesture, the drawn shape) from a friction sound. Results from a series of perceptual experiments revealed that the velocity profile transmits relevant information about the gesture and the geometry of the drawn shape to a certain extent. Results also indicated the relevance of the so-called 1/3-power law, defined from seminal works by Viviani and his colleagues and translating a biomechanics constraint between the velocity of a gesture and the local curvature of the drawn shape, to evoke a fluid and natural human gesture through a friction sound (cf. Thoret et al. 2013, 2014 for details and review).

Other environmental sounds: For other classes of environmental sounds such as wave or aerodynamic sounds, physical considerations generally involve complex modelling and signal models are then useful. From a perceptual point of view, these sounds evoke a wide range of different physical sources, but interestingly, from a signal point of view, common acoustic morphologies can be highlighted across these sounds. We analysed several signals representative of the main categories of environmental sounds as defined by Gaver and we identified a certain number of perceptually relevant signal morphologies linked to these categories (Gaver 1993a, b). To date, we concluded on five elementary sound morphologies based on impacts, chirps and noise structures (Verron et al. 2009). This finding is based on a heuristic approach that has been verified on a large set of environmental sounds. Granular synthesis processes based on these five sound "atoms" then enabled the generation of various environmental sounds (i.e. solid interactions, aerodynamic or liquid sounds). Note that this atom dictionary may be completed or refined in the future without compromising the proposed methodology.

A first type of grain is the "tonal solid grain" that is defined by a sum of exponentially damped sinusoids. Such a grain is well adapted to simulate sounds produced by solid interactions. Nevertheless, this type of grain cannot alone account for any kind of solid impact sounds. Actually, impact sounds characterized by a strong density of modes or by a heavy damping may rather be modelled as an exponentially damped noise. This sound characterization stands for both perceptual and signal points of view, since no obvious pitch can be extracted from such sounds. Exponentially damped noise constitutes the second type of grain, the so-called "noisy impact grain". Such a grain is well adapted to simulate crackling sounds. The third type of grain concerns liquid sounds. From an acoustic point of view, cavitation phenomena (e.g. a bubble in a liquid) lead to local pressure variations that generate time-varying frequency components such as exponentially damped linear chirps. Hence, the so-called "liquid grain" consists of an exponentially damped chirp signal. Finally, aerodynamic sounds generally result from complicated interactions between solids and gases and it is therefore difficult to extract useful information from corresponding physical models. A heuristic approach allowed us to define two kinds of aerodynamic grains: the "whistling grain" (slowly varying narrow band noise) and the "background aerodynamic grain" (broadband filtered noise). Such grains are well adapted to simulate wind and waves.

By combining these five grains using an accurate statistics of appearance, various environmental auditory scenes can be designed such as rainy ambiances, seacoast ambiances, windy environments, fire noises, or solid interactions simulating solid impacts or footstep noises. We currently aim at extracting the parameters corresponding to these grains from the analysis of natural sounds, using matching pursuit like methods. Perceptual evaluations of these grains will further allow us to identify or validate signal morphologies conveying relevant information on the perceived properties of the sound source.

4.4 Control Strategies for Synthesis Processes

The choice of synthesis model highly influences the control strategy. Physical synthesis models have physically meaningful parameters, which might facilitate the interpretation of the consequence of the control on the resulting sound. This is less so for signal models obtained from mathematical representations of sounds.

Perceptual considerations might, however, be combined with these models to propose intuitive control strategies as described in the following sections.

4.4.1 Control of Synthesis Parameters

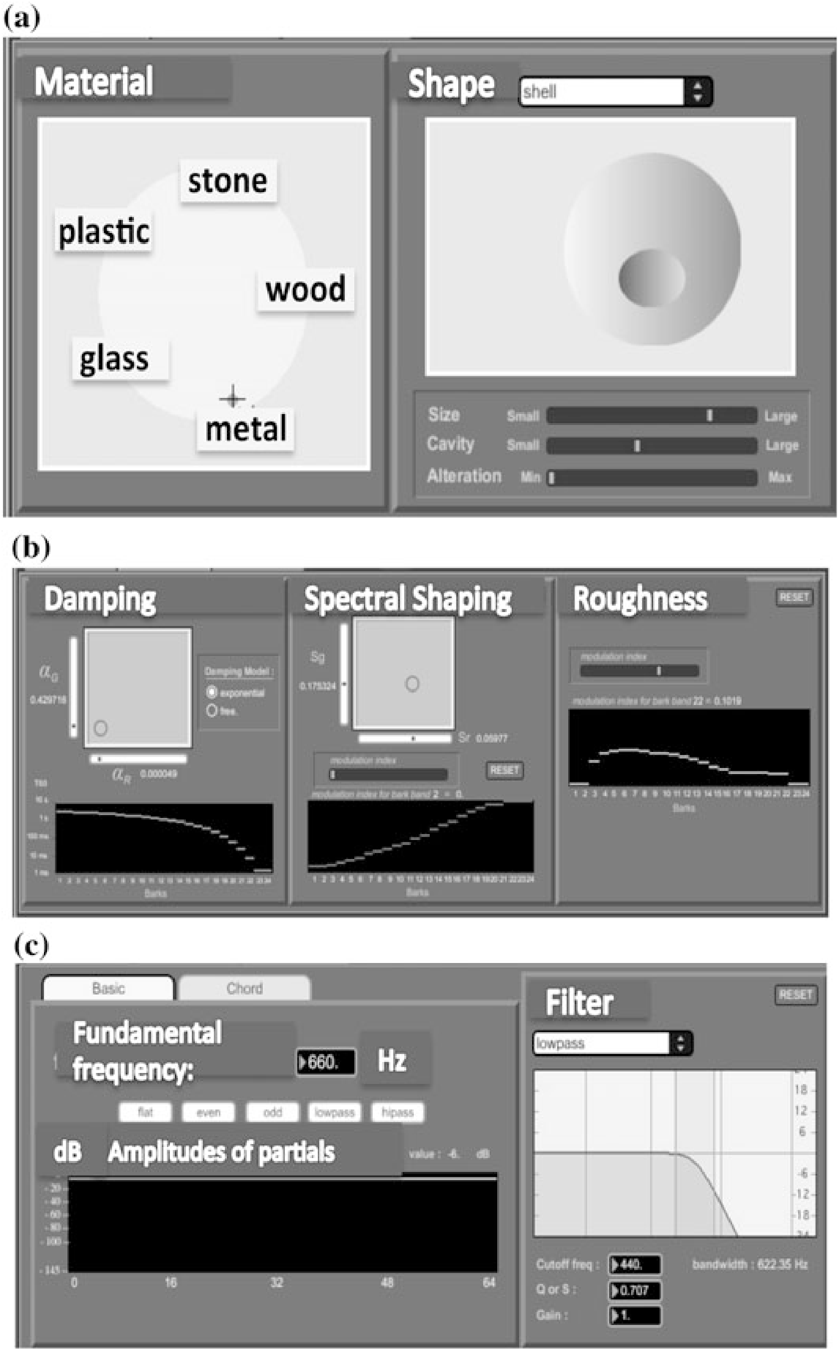
Although physical models can produce high-quality sounds that are useful for instance for musical purposes, this approach is less adapted to environmental sounds, when the physics of such sound sources is not sufficiently well understood or the existing models are not real-time compatible. In such cases, signal models that enable the simulation of the sound vibrations through mathematical models are useful. The control of these models consists in manipulating physical or signal parameters. Practically, these approaches might involve the control of physical variables (for instance, characterizing the tribological or mechanical properties of the source) or a high number of synthesis parameters (up to a hundred) that are generally not intuitive for a non-expert user. This means that a certain scientific (or musical) expertise is needed to use such models (expert control). In fact, the calibration of the control of these models is based on an *error function* that reveals the difference between the model and the actual physical sound vibration (cf. Fig. 4.2).

4.4.2 Control of Perceptual Effects

Common to all the previous approaches described in Sect. 4.4.1 is the lack of perceptual criteria. Actually, since the timbre of the resulting sound is generally related to the synthesis parameters in a non-linear way, the control process can quickly become complicated and non-intuitive. The design of a control of perceptual effects may lead to the definition of an intuitive control strategy. In particular, based on the identification of invariant sound morphologies (cf. Sect. 4.3.2), control processes mediating various perceptual evocations could be designed. In line with the previous definitions of structural and transformational invariants, the framework of our control strategy is based on the so-called {*action/object*} paradigm, assuming that the produced sound can be defined as the consequence of an action on an object. This approach supports the determination of sound morphologies that carry information about the action and the object, respectively.

Here we present several synthesis tools that we have developed for generating and intuitively controlling sounds. These synthesis models make it possible to relevantly resynthesize natural sounds. In practice, we adopted hierarchical levels of control to route and dispatch the parameters from an intuitive to the algorithmic level. As these parameters are not independent and might be linked to several signal properties at a time, the mapping between levels is far from being straightforward.

Sounds from impacted objects: We have developed an impact sound synthesizer offering an intuitive control strategy based on a three-level architecture (Aramaki et al. 2010a) (cf. Fig. 4.1). The top layer gives the user the possibility to define the impacted object using verbal descriptions of the object (nature of the



Shape: shell

Fundamental frequency: 660. Hz

value : -6. dB

Filter: lowpass

Cutoff freq : 440.

bandwidth : 622.35 Hz

Q or S : 0.707

Gain : 1.

Fig. 4.1 **a** *Top layer* (semantic labels describing the perceived material and shape of the object), **b** *middle layer* (acoustic descriptors) and **c** *bottom layer* (synthesis parameters of the signal model) designed for the control of the impact sound synthesizer

perceived material, size and shape, etc.) and the excitation (force, hardness of the mallet, impact position, etc.). The middle layer is based on perceptually relevant acoustic descriptors linked to the invariant sound morphologies (cf. Sect. 4.3.2). The bottom layer consists of the set of synthesis parameters (for expert users). Two mapping strategies are implemented between the layers (we refer to (Aramaki et al. 2010a) for more details). The first one focuses on the relationships between verbal

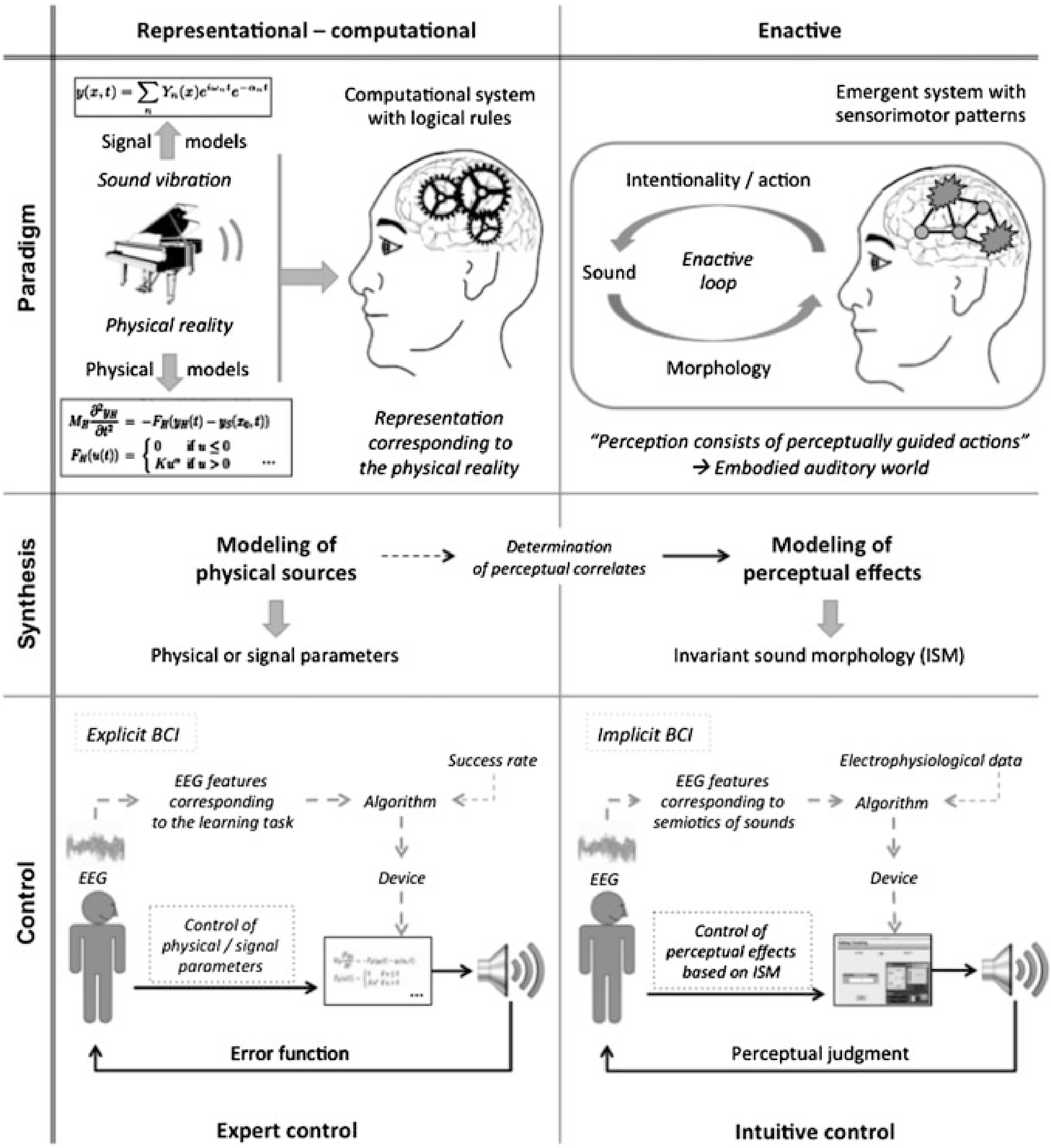


Fig. 4.2 General didactic synopsis including two approaches inspired by the representational-computational and enactive paradigms from cognitive neuroscience, the associated viewpoints for sound synthesis (modelling of physical sources and modelling of perceptual effects) and sound control (expert and intuitive control). A prospective view on the use of BCI in the context of sound synthesis control is also illustrated

descriptions of the sound source and the sound descriptors (damping, inharmonicity, roughness, etc.) characterizing perceptually relevant sound morphologies. The second one focuses on the relationships between sound descriptors and synthesis parameters (damping coefficient, amplitude and frequency of the components).

Sounds from continuous interactions: Control strategies for the synthesis processes of such sounds have recently been developed. In particular, an intuitive control strategy adapted to a non-linear friction sound model (producing phenomena such a creaky door, a singing glass or a squeaking wet plate) has been proposed. Inspired from Schelleng's diagrams, the proposed control is defined from a flexible physically informed mapping between a dynamic descriptor (velocity, pressure), and the synthesis parameters, and allows coherent transitions between the different non-linear friction situations (Thoret et al. 2013). Another intuitive control strategy dedicated to rolling sound synthesis has also been proposed (Conan et al. 2013). This strategy is based on a hierarchical architecture similar to that of the impacted object sounds (cf. previous paragraph). The high-level controls that can be manipulated by the end-user are the characteristics of the rolling ball (i.e. size, asymmetry and speed) and the irregularity of the surface. The low-level parameters (e.g. impacts' statistics, modulation frequency and modulation depth) are modified accordingly with respect to the defined mapping. Recently, a control strategy enabling to perceptually morph between the three continuous interactions, i.e. rubbing, scratching and rolling, was designed. For that purpose, we developed a synthesis process that is generic enough to simulate these different interactions and based on the related invariant sound morphologies (cf. Sect. 4.3.2). Then, a perceptual "interaction space" and the associated intuitive navigation strategy were defined with given sound prototypes considered as anchors in this space (Conan et al. 2013).

Finally, in line with the *action/object* paradigm, the complete synthesis process has been implemented as a source-filter model. The resulting sound is then obtained by convolving the excitation signal (related to the nature of the interaction) with the impulse response of the resonating object. The impulse response is implemented as a resonant filter bank, which central frequencies correspond to the modal frequencies of the object.

Immersive auditory scenes: An intuitive control of the sound synthesizer dedicated to environmental auditory scenes was defined. The control enables the design of complex auditory scenes and included the location and the spatial extension of each sound source in a 3D space so as to increase the realism and the feeling of being immersed in virtual scenes. This control is particularly relevant to simulate sound sources such as wind or rain that are naturally diffuse and wide. In contrast with the classical two-stage approach, which consists in first synthesizing a monophonic sound (timbre properties) and then spatializing the sound (spatial position and extension in a 3D space), the architecture of the proposed synthesizer yielded control strategies based on the overall manipulation of timbre and spatial attributes of sound sources at the same level of sound generation (Verron et al. 2010).

The overall control of the environmental scene synthesizer can be effectuated through a graphical interface where the sound sources (selected among a set of available sources: fire, wind, rain, wave, chimes, footsteps, etc.) can be placed around the listener (positioned in the centre of the scene) by defining the distance and the spatial width of each source. The sources are built from the elementary grains defined previously in Sect. 4.3.2. A fire scene is for instance built from a combination of a whistling grain (simulating the hissing), a background aerodynamic grain (simulating the background combustion) and noisy impact grains (simulating the cracklings). The latter grains are generated and launched randomly with respect to time using an accurate statistic law that can be controlled. A global control of the fire intensity, mapped with the control of the grain generation (amplitude and statistic law), is then designed. A rainy weather sound ambiance can be designed with a rain shower, water flow and drops, each of these environmental sounds being independently spatialized and constructed from a combination of the previous grains (see Verron et al. 2009 for more details). In case of interactive uses, controls can be achieved using either MIDI interfaces, from data obtained from a graphical engine or other external data sources.

4.5 Evidence of Semiotics for Non-linguistic Sounds

To propose an even more intuitive control of sound synthesis that directly uses a BCI, a relationship between the electroencephalogram (EEG) and the nature of the underlying cerebral processes has to be investigated. We here present results of several experimental studies aiming at supporting the existence of semiotics for non-linguistic sounds. In these studies, we used either synthetic stimuli using analysis/transformation/synthesis processes or sounds of a specific kind called "abstract" sounds promoting acousmatic listening (cf. Sect. 4.2). The participants' responses and reaction times (RTs) provided objective measurements to the processing of stimulus complexity.

Electrophysiological data: When appropriate, we also investigated the neural bases of the involved brain processes by analysing the EEG with the method of event-related potentials (ERP) time-locked to the stimulus onset during the various information processing stages. The ERP elicited by a stimulus (a sound, a light, etc.) are characterized by a series of positive (P) and negative (N) deflections relative to a baseline. These deflections (called components) are defined in terms of their polarity, their maximum latency (relative to the stimulus onset), their distribution among several electrodes placed in standard positions on the scalp and by their functional significance. Components P100, N100 and P200 are consistently activated in response to the auditory stimuli (Rugg and Coles 1995). Several late ERP components (N200, P300, N400, etc.) are subsequently elicited and associated with specific brain processes depending on the experimental design and the task in hand.

4.5.1 Perceptual Categorization of Sounds from Impacted Materials

In this experiment, we studied the perception of sounds obtained from impacted materials, in particular, wood, metal and glass (Aramaki et al. 2010a; Aramaki et al. 2010b; Aramaki et al. 2011). For this purpose, natural sounds were recorded, analysed, resynthesized and tuned to the same chroma to obtain sets of synthetic sounds representative of each category of the selected material. A sound-morphing process (based on an interpolation method) was further applied to obtain sound continua simulating progressive transitions between materials. Although sounds located at the extreme positions on the continua were indeed perceived as typical exemplars of their respective material categories, sounds in intermediate positions, which were synthesized by interpolating the acoustic parameters characterizing sounds at extreme positions, were consequently expected to be perceived as ambiguous (e.g. to be neither wood nor metal). Participants were asked to categorize each of the randomly presented sounds as wood, metal or glass.

Based on the classification rates, we defined "typical" sounds as sounds that were classified by more than 70 % of the participants in the right material category and "ambiguous" sounds, those that were classified by less than 70 % of the participants in a given category. Ambiguous sounds were associated with slower RTs than typical sounds. As might be expected, ambiguous sounds are therefore more difficult to categorize than typical sounds. This result is in line with previous findings in the literature showing that non-meaningful sounds were associated with longer RTs than meaningful sounds. Electrophysiological data showed that ambiguous sounds elicited more negative ERP (a negative component, N280, followed by a negative slow wave, NSW) in fronto-central brain regions and less positive ERP (P300 component) in parietal regions than typical sounds. This difference may reflect the difficulty to access information from long-term memory. In addition, electrophysiological data showed that the processing of typical metal sounds differed significantly from those of typical glass and wood sounds as early as 150 ms after the sound onset. The results of the acoustic and electrophysiological analyses suggested that spectral complexity and sound duration are relevant cues explaining this early differentiation. Lastly, it is worth noting that no significant differences were observed on the P100 and N100 components. These components are known to be sensitive to sound onset and temporal envelope, reflecting the fact that the categorization process occurs in later sound-processing stages.

4.5.2 Conceptual Priming for Non-linguistic Sounds

In language, a comprehensible linguistic message is for instance conveyed by associating words while respecting the rules of syntax and grammar. Can similar links be generated between non-linguistic sounds so that any variation will change the global information conveyed? From the cognitive neuroscience point of view, one of the major issues that arises from this question is whether similar neural networks are involved in the allocation of meaning in the case of language and that of sounds of other kinds. In a seminal study using a priming procedure, Kutas and Hillyard (Kutas and Hillyard 1980) established that the amplitude of a negative ERP component, the N400 component, increases when final sentence words are incongruous (e.g. *The fish is swimming in the river/carpet*). Since then, the N400 has been widely used to study semantic processing in language. In recent studies, priming procedures with non-linguistic stimuli such as pictures, odours, music and environmental sounds have been used (e.g. Holcomb and McPherson 1994; Castle et al. 2000; Koelsch et al. 2004; Daltrozzo and Schön 2009; Van Petten and Rheinfelder 1995; Orgs et al. 2006). Although the results of these experiments mostly have been interpreted as reflecting some kind of conceptual priming between words and non-linguistic stimuli, they may also reflect linguistically mediated effects. For instance, watching a picture of a bird or listening to a birdsong might automatically activate the verbal label "bird". Therefore, the conceptual priming cannot be taken to be purely non-linguistic because of the implicit naming induced by the processing of the stimulus. Such conceptual priming might imply at least language, generation of auditory scenes, and mental imaging, at various associative (non specific) cortex area levels. This might probably activate large neural/glial networks using long-distance synchronies, which could be investigated by a synchronous EEG activity measurement (Lachaux et al. 1999).

The aim of our first conceptual priming study (Schön et al. 2010) was to attempt to reduce as far as possible the likelihood that a labelling process of this kind takes place. To this end, we worked with a specific class of sounds called "abstract sounds", which physical sources cannot be easily recognized, meaning that verbal labelling is less likely to take place (Merer et al. 2011). We then conducted conceptual priming tests using word/sound pairs with different levels of congruence between the prime and the target. Subjects had to decide whether or not the prime and the target matched. In the first experiment, a written word was presented visually before the abstract sound, and in the second experiment, the order of presentation was reversed. Results showed that participants were able to assess the relationship between the prime and the target in both presentation orders (sound/word vs. word/sound), showing low inter-subject variability and good consistency. The presentation of a word reduced the variability of the interpretations of the abstract sound and led to a consensus between subjects in spite of the fact that the sound sources were not easily recognizable. Electrophysiological data showed the occurrence of an enhanced negativity in the 250–600-ms latency range in response to unrelated as compared to related targets in both experiments and the presence of a more fronto-central distribution in response to word targets and a more centro-parietal distribution in response to sound targets.

In a subsequent study (Aramaki et al. 2010b), we avoided the use of words as primes or targets. Conceptual priming was therefore studied using impact sounds (also used in the categorization experiment previously presented), as both primes and targets. As described in Sect. 4.5.1, these impact sounds were qualified as either typical or ambiguous with respect to a material category depending on their score in the categorization experiment. 3° of congruence were investigated through various combinations of typical and ambiguous sounds as prime and target: related, ambiguous and unrelated. The priming effects induced in these conditions were compared with those observed with linguistic sounds (spoken words) in the same group of participants. Results showed that N400-like components were also activated in a sound–sound design. This component may therefore reflect a search for meaning that is not restricted to linguistic meaning. Moreover, ambiguous targets also elicited larger N400-like components than related targets for both linguistic and non-linguistic sounds. These findings showed the existence of similar relationships in the processing of semiotics of both non-linguistic and linguistic target sounds. This study clearly means that it is possible to draw up a real language for non-linguistic sounds.

4.6 Towards a Semiotic-Based Brain Computer Interface (BCI)

BCIs provide a link between a user and an external electronic device through his or her brain activity, independently of the voluntary muscle activity of the subject. Most often BCIs are based on EEG recordings that allow for non-invasive measurements of electrical brain activity. As substitutional devices, BCIs open interesting perspectives for rehabilitation, reducing disability and improving the quality of life of patients with severe neuromuscular disorders such as amyotrophic lateral sclerosis or spinal cord injury (Wolpaw et al. 2002). Such interfaces, among many other possibilities, enable patients to control a cursor, to select a letter on a computer screen, or to drive a wheelchair. In addition to medical and substitutional applications, BCIs as enhancing devices can be used with healthy subjects. For example, in the field of video games, BCIs could capture the cognitive or emotional state of the user through the EEG to develop more adaptive games and to increase the realism of the gaming experience (Nijholt 2009). To date, two approaches to BCI could be highlighted: “explicit (or active) BCI” and “implicit (or passive) BCI” (George and Lécuyer 2010). These two classes of BCI could be linked with the two approaches inspired from the paradigms of cognitive science (described in Sect. 4.2) and the two approaches for sound synthesis (described in Sect. 4.3).

4.6.1 Explicit BCI

The explicit BCI is based on the principles of *operant conditioning*, the basic learning concept in experimental psychology, which assumes that the probability of occurrences of an animal or human behaviour is a function of a positive or negative reinforcement during the subject’s learning process (Micoulaud-Franchi et al. 2013). Thus, the explicit BCI requires a learning period (George and Lécuyer 2010). In practice, the subject intentionally tries to control his/her cognitive activity to change his/her EEG activity and control an external electronic device. The EEG signal is recorded, processed in real time to extract the information of interest (e.g. spectral power EEG, slow cortical potential or ERP). This information is related to a cognitive activity that the subject intentionally produces. This information is further transmitted to the external electronic device using specific mapping that leads to the control of the device in the desired direction. The positive reinforcement (and the success rate) is determined by the capacity of controlling the external electronic device to achieve a given task.

This configuration fits with traditional neurofeedback therapeutics where the subject learns to intentionally control EEG through visual or auditory positive reinforcement, without any control of external device (Micoulaud-Franchi et al. 2013). In this context, the positive reinforcement could be an increase of a number of points, an advance of an animation on a computer screen, or a modification of a sound. When the EEG is related to symptoms of a disease, it has been shown that neurofeedback techniques can have a therapeutic effect, as is the case with attention deficit disorder with hyperactivity (Micoulaud-Franchi et al. 2011) or epilepsy (Micoulaud-Franchi et al. 2014).

4.6.2 Implicit BCI

In contrast with explicit BCI, the implicit BCI is not based on the principle of operant conditioning. The feedback in implicit BCI is used to optimize the interaction with an external device by directly modulating the brain activity and the cognitive activity of the subject (George and Lécuyer 2010). Implicit BCI does not require a learning period. In practice, the subject does not have to try to control intentionally his EEG. The EEG signal is recorded, processed in real time to extract the information of interest (e.g. power spectral EEG or ERP) corresponding to the subject's cognitive activity, and transmitted to the external electronic device to modulate and optimize the interaction between the device and the user.

This configuration fits with some non-traditional neurofeedback therapeutics that do not require specific cognitive tasks and are supposed to directly modulate the brain activity of the subject in order to optimize brain dynamics, although this remains largely hypothetical. Thus, unlike traditional neurofeedback approaches presented in the previous section, these non-traditional neurofeedback approaches have a very low level of therapeutic and clinical evidence (Micoulaud-Franchi et al. 2013).

4.6.3 Towards an Intuitive Control Using Semiotic-Based BCI

From the two approaches inspired by previous theoretical frameworks from cognitive neuroscience (Sect. 4.2), we propose a prospective view on a sound synthesis control strategy based on BCI. We reflect on whether EEG BCI would be helpful to increase the intuitiveness of control with the sound synthesizer. For a didactic perspective, we suggest to describe explicit and implicit BCI, respectively, from the representational-computational and from the enactive points of view.

We stress that in the explicit BCI, the user controls the external electronic device (positive reinforcement) as if it was an external object. In some way, there is a gap between the information of interest extracted from the recorded EEG activity and the positive reinforcement. The information feedback could be given to the subject by any kind of signal. The positive reinforcement mainly is useful for the learning process and for determining a success rate and is close to an error function (Sect. 4.4.1). We think that in many cases, explicit BCI does not permit to create recurrent sensorimotor patterns (from the enactive point of view) that enable action to be guided by the direct perception of the stimulus, which could be a limitation in the intuitiveness of BCI controllability.

We stress that in the Implicit BCI, the user and his/her brain is involved in an enactive process. In some way, there is a direct link between the information of interest extracted from the recorded EEG and the feedback. This feedback is not a positive reinforcement as defined by the operant-conditioning model. In fact, the aim of the feedback is not to inform the subject about the cognitive strategies that he/she develops during the learning process, but to directly influence the brain activity (and thus the EEG). Any kind of feedback cannot be used, but only those with the desired effect on the brain and the cognitive activity in order to enhance the interaction and the intuitiveness of the system.

Therefore, in the context of sound synthesis, a control strategy involving the use of explicit or implicit BCI would necessitate different mapping strategies. From a conceptual point of view, we stress that explicit and implicit BCI involve different levels of semiotic relation, i.e., the relation between the feedback and the meaning that the subject attributes to a sound. These two scenarios are discussed in the following paragraphs.

In the case of explicit BCI as defined above, the subject would have to control his/her cognitive activity to change his/her EEG and thus to control a specific parameter of the sound synthesizer. No semiotic relation between the EEG, the effect of the synthesized sound on the EEG, and the sound perception is therefore needed. In other words, the subject has to do something that is not necessarily related to the semiotics of the perceived synthesized sound to control the synthesizer. More so, an external algorithm is used to interpret the information of interest extracted from the EEG and to control the electronic device. For example, paying attention to a target to produce a P300 component that will be processed by the BCI and arbitrarily associated with a control parameter according to the output of the algorithm and to a success rate (Fig. 4.2). This situation that necessitates a certain expertise acquired during a learning period seems to be quite close to sound synthesis based on the physical or signal modelling of sound vibrations (Sect. 4.3).

In the case of implicit BCI as defined above, the aim would be to enhance the quality and the intuitiveness of the sound synthesizer by taking into account the EEG induced by the sound. Thus, a strict semiotic relation between the EEG and the influence of sounds on the EEG should be known. In other words, we need to understand the neural bases of sound semiotics ("electrophysiological data" in Fig. 4.2) to implement this information in an implicit BCI process dedicated to the sound synthesizer. We propose to call it "semiotic-based BCI". In this context, the results obtained from previous EEG experiments presented in Sect. 4.5 constitute an interesting starting point for the design of such a mapping strategy. This approach seems to be quite close to sound synthesis based on the modelling of perceptual effects, which does not necessitate a learning period (Sect. 4.3). This intuitive control implies that perceptual and cognitive aspects are taken into account in order to understand how a sound is perceived and interpreted. As shown in Fig. 4.2, a loop is thus designed between perception and action through the intuitive control of the sound synthesizer (Sect. 4.2). Implicit BCI offers the possibility of a second loop, between the sound effect on the EEG and the sound synthesizer that is likely to optimize the sound effect on both the perceptual judgment and the Implicit BCI.

4.7 Conclusion

To date, the design of a control strategy of sound synthesis processes that uses a BCI is still a challenging perspective. As discussed in (Väljamäe et al. 2013), a synthesis control of sounds directly from the brain through the measurement of its cerebral activity is still in its early stages. In particular, the mapping between electrophysiological signal features and synthesis parameters is generally validated on the basis on different metrics depending on applications. However, the definition of such metrics implies a given conception on the way we interact with the surrounding world.

To broach this issue, we introduced two conceptual approaches inspired from the representational-computational and the enactive paradigms from cognitive neuroscience. In light of these paradigms, we revisited the existing main approaches for synthesis and control of sounds. In fact, the viewpoints adopted to synthesize sounds are intricately underpinned by paradigms that differ in the epistemological positions of the observer (from a third or a first-person position) and have a substantial consequence on the design of a control strategy (cf. Figure 4.2). On one hand, synthesis processes based on the modelling of physical sources (from either the mechanical behaviour or the resulting vibration) are controlled by physical or signal parameters. This approach is based on the existence of a correct representation of the physical world and introduces the notion of an error function between the model and the physical reality as a quality criterion. Therefore, it requires a certain expertise from the end-user. On the other hand, synthesis processes based on the modelling of perceptual effects involve the identification of invariant sound morphologies specific to given perceptual attributes of the sound source. This approach assumes the emergence of an embodied auditory world from an enactive process. The perceptual judgments are considered as a quality criterion for the model, leading to the design of a more intuitive control.

By associating these conceptual and pragmatic considerations, we proposed a prospective view on the methodology to be used to design a BCI control. For the sake of illustration, we treated limited aspects of BCIs by addressing explicit BCI from the representational-computational point of view and implicit BCI from the enactive point of view. Actually, we are aware that the frontier between explicit and implicit BCI might be difficult to establish and less didactic than what this article presents. Indeed, the implicit communication channel might sometimes be used in an explicit way (George and Lécuyer 2010), and inversely brain plasticity can enable the participant to make use of the training experienced from the explicit BCI to generate implicit recurrent sensorimotor patterns (Bach-y-Rita and Kercel 2003). With current apparatus performances, the rate of transfer information between the BCI and the device is quite limited and the final task has to be defined accordingly. While this technique may represent a restricted interest for healthy users (in some cases, it would be easier to directly control the device manually), it constitutes a relevant medium for medical applications and can be used as a substitutional device for diseases. In the implicit BCI, the control is included in an optimization system in which the electrophysiological data supplies further information about the way the user perceives the sound (beyond verbal labels or gestures for instance). In contrast with the explicit BCI, this configuration is well adapted to intuitive synthesis control. Therefore, we suggested a “semiotic-based BCI” founded on identified links between the brain activity and invariant signal morphologies reflecting the attribution of sense to a sound that may enhance the interactivity and the intuitiveness of the system.

4.8 Questions

1. What are the characteristics of the representational-computational paradigm of perception?
2. What are the characteristics of the enactive paradigm of perception?
3. What is the difference between physical and signal sound synthesis models?
4. What are the main limitations of the use of physical models for sound synthesis?
5. How can the invariant sound morphologies be determined?
6. Which invariant sound morphologies are related to the perception of material in an impact sound?
7. Which aspects should be taken into account in the design of a control strategy based on a representational-computational or an enactive paradigm?
8. What are the characteristics of explicit (or active) BCI?
9. What are the characteristics of implicit (or passive) BCI?
10. What is the purpose of offering intuitive control of sound synthesis processes using BCI?

References

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2009) Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad S, Kronland-Martinet R, Jensen K (eds) Computer music modeling and retrieval—genesis of meaning of sound and music, vol 5493., LNCSSpringer, Berlin, Heidelberg, pp 1–17

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2011) Controlling the perceived material in an impact sound synthesizer. IEEE Trans Audio Speech Lang Process 19(2):301–314

Aramaki M, Gondre C, Kronland-Martinet R, Voinier T, Ystad S (2010a) Imagine the sounds: an intuitive control of an impact sound synthesizer. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K (eds) Auditory display, vol 5954., Lecture notes in computer scienceSpringer, Berlin, Heidelberg, pp 408–421

Aramaki M, Marie C, Kronland-Martinet R, Ystad S, Besson M (2010b) Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. J Cogn Neurosci 22(11):2555–2569

Arfib D (1979) Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. J Audio Eng Soc 27:757–768

Atal BS, Hanauer SL (1971) Speech analysis and synthesis by linear prediction of the speech wave. J Acoust Soc Am 50(2B):637–655

Avanzini F, Serafin S, Rocchesso D (2005) Interactive simulation of rigid body interaction with friction-induced sound generation. IEEE Trans Speech Audio Process 13(5):1073–1081

Bach-y-Rita P, Kercel W (2003) Sensory substitution and the human-machine interface. Trends in Cogn Sci 7:541–546

Ballas JA (1993) Common factors in the identification of an assortment of brief everyday sounds. J Exp Psychol Hum Percept Perform 19(2):250–267

Bensa J, Jensen K, Kronland-Martinet R (2004) A hybrid resynthesis model for hammer-strings interaction of piano tones. EURASIP J Appl Sig Process 7:1021–1035

Bilbao S (2009) Numerical sound synthesis: finite difference schemes and simulation in musical acoustics. Wiley, Chichester, UK

Castle PC, van Toller S, Milligan G (2000) The effect of odour priming on cortical EEG and visual ERP responses. Int J Psychophysiol 36:123–131

Chaigne A (1995) Trends and challenges in physical modeling of musical instruments, In: Proceedings of the international congress on acoustics', Trondheim, Norway

Chowning J (1973) The synthesis of complex audio spectra by means of frequency modulation. J Audio Eng Soc 21:526–534

Conan S, Aramaki M, Kronland-Martinet R, Thoret E, Ystad S (2012) Perceptual differences between sounds produced by different continuous interactions. Proceedings of the 11th Congrès Français d'Acoustique. Nantes, France, pp 409–414

Conan S, Aramaki M, Kronland-Martinet R, Ystad S (2013) Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012). Lecture notes in computer science, vol 7900. Springer, Berlin, Heidelberg, chapter Intuitive Control of Rolling Sound Synthesis

Conan S, Thoret E, Aramaki M, Derrien O, Gondre C, Kronland-Martinet R, Ystad S (2013) Navigating in a space of synthesized interaction-sounds: rubbing, scratching and rolling sounds. In: Proceedings of the 16th international conference on digital audio effects (DAFx-13), Maynooth, Ireland

Cook PR (1992) A meta-wind-instrument physical model, and a meta-controller for real-time performance control. In: Proceedings of the international computer music conference, pp 273–276

Daltrozzo J, Schön D (2009) Conceptual processing in music as revealed by N400 effects on words and musical targets. J Cogn Neurosci 21:1882–1892

de Saussure F (1955) Cours de linguistique générale. Payot, Paris

Flanagan JL, Coker CH, Rabiner LR, Schafer RW, Umeda N (1970) Synthetic voices for computer. IEEE Spectr 7:22–45

Gaver WW (1993a) How do we hear in the world? Explorations of ecological acoustics. Ecol Psychol 5(4):285–313

Gaver WW (1993b) What in the world do we hear? An ecological approach to auditory source perception. Ecol Psychol 5(1):1–29

George L, Lécuyer A (2010) An overview of research on "passive" brain-computer interfaces for implicit human-computer interaction. In: International conference on applied bionics and biomechanics ICABB 2010—workshop W1 Brain-Computer Interfacing and Virtual Reality, Venezia, Italy

Gibson JJ (1986) The ecological approach to visual perception, Lawrence Erlbaum Associates

Giordano BL, McAdams S (2006) Material identification of real impact sounds: effects of size variation in steel, wood, and plexiglass plates. J Acoust Soc Am 119(2):1171–1181

Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. Percept Psychophys 69(6):839–855

Gygi B, Shafiro V (2007) General functions and specific applications of environmental sound research. Front Biosci 12:3152–3166

Hermes DJ (1998) Synthesis of the sounds produced by rolling balls. Internal IPO report no. 1226, IPO, Center for user-system interaction, Eindhoven, The Netherlands

Holcomb PJ, McPherson WB (1994) Event-related brain potentials reflect semantic priming in an object decision task. Brain and Cogn 24:259–276. http://forumnet.ircam.fr/product/modalys/?lang=en (n.d.). http://www-acroe.imag.fr/produits/logiciel/cordis/cordis_en.html (n.d.)

Husserl E (1950) Idées directrices pour une phénoménologie, Gallimard. J New Music Res, special issue “enaction and music” (2009), 38(3), Taylor and Francis, UK

Karjalainen M, Laine UK, Laakso T, Vilimtiki V (1991) Transmission-line modeling and real-time synthesis of string and wind instruments. In: I. C. M. Association (ed) Proceedings of the international computer music conference, Montreal, Canada, pp 293–296

Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici A (2004) Music, language and meaning: brain signatures of semantic processing. Nat Neurosci 7(3):302–307

Kronland-Martinet R (1989) Digital subtractive synthesis of signals based on the analysis of natural sounds. A.R.C.A.M. (ed), Aix en Provence

Kronland-Martinet R, Guillemain P, Ystad S (1997) Modelling of natural sounds by time-frequency and wavelet representations. Organ Sound 2(3):179–191

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–204

Lachaux JP, Rodriguez E, Martinerie J, Varela F (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208

Lalande A (1926) Vocabulaire technique et critique de la philosophie, Edition actuelle, PUF n quadrige z 2002

Le Brun M (1979) Digital waveshaping synthesis. J Audio Eng Soc 27:250–266

Makhoul J (1975) Linear prediction, a tutorial review. In: Proceedings of the IEEE, vol 63. pp 561–580

Matyja JR, Schiavio A (2013) Enactive music cognition: background and research themes. Constr Found 8(3):351–357. http://www.univie.ac.at/constructivism/journal/8/3/351.matyja

McAdams S (1999) Perspectives on the contribution of timbre to musical structure. Comput Music J 23(3):85–102

McAdams S, Bigand E (1993) Thinking in sound: the cognitive psychology of human audition, Oxford University Press, Oxford

Merer A, Ystad S, Kronland-Martinet R, Aramaki M (2011) Abstract sounds and their applications in audio and perception research. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K (eds) Exploring music contents, vol 6684., Lecture notes in computer scienceSpringer, Berlin, Heidelberg, pp 176–187

Micoulaud-Franchi JA, Bat-Pitault F, Cermolacce M, Vion-Dury J (2011) Neurofeedback dans le trouble déficit de l’attention avec hyperactivité : de l’efficacité à la spécificité de l’effet neurophysiologique. Annales Médico-Psychologiques 169(3):200–208

Micoulaud-Franchi JA, Cermolacce M, Vion-Dury J, Naudin J (2013) Analyse critique et épistémologique du neurofeedback comme dispositif thérapeutique. le cas emblématique du trouble déficit de l’attention avec hyperactivité’, L’évolution psychiatrique

Micoulaud-Franchi J, Lanteaume L, Pallanca O, Vion-Dury J, Bartolomei F (2014) Biofeedback et épilepsie pharmacorésistante : le retour d’une thérapeutique ancienne ? Revue Neurologique 170(3):187–196

Nadeau R (1999) Vocabulaire technique et analytique de l’épistémologie, PUF

Nijholt A (2009) BCI for games: a ‘state of the art’ survey. In: Stevens SM, Saldamarco SJ (eds) LNCS, vol 5309. Springer, Berlin, pp 225–228

O’Brien JF, Shen C, Gatchalian CM (2002) Synthesizing sounds from rigid-body simulations. In: Press A (ed) The ACM SIGGRAPH 2002 symposium on computer animation, pp 175–181

Orgs G, Lange K, Dombrowski J, Heil M (2006) Conceptual priming for environmental sounds and words: An ERP study. Brain Cogn 62(3):267–272

Pai DK, van den Doel K, James DL, Lang J, Lloyd JE, Richmond JL, Yau SM (2001) Scanning physical interaction behavior of 3D objects. In: Proceedings of SIGGRAPH 2001, computer graphics proceedings, annual conference series, pp 87–96

Petitmengin C, Bitbol M, Nissou JM, Pachoud B, Curalucci H, Cermolacce M, Vion-Dury J (2009) Listening from within. J Conscious Stud 16:252–284

Rabiner LR, Gold B (1975) Theory and application of digital signal processing. Prentice Hall, Englewood Cliffs, NJ

Rath M, Rocchesso D (2004) Informative sonic feedback for continuous human–machine interaction—controlling a sound model of a rolling ball. IEEE Multimedia Spec Interact Sonification 12(2):60–69

Risset JC (1965) Computer study of trumpet tones. J Acoust Soc Am 33:912

Roads C (1978) Automated granular synthesis of sound. Comput Music J 2(2):61–62

Rugg MD, Coles MGH (1995) Electrophysiology of mind. Event-related brain potentials and cognition, number 25. In: ‘Oxford Psychology’, Oxford University Press, chapter The ERP and Cognitive Psychology: Conceptual issues, pp 27–39

Schaeffer P (1966) Traité des objets musicaux, du Seuil (ed)

Schön D, Ystad S, Kronland-Martinet R, Besson M (2010) The evocative power of sounds: conceptual priming between words and nonverbal sounds. J Cogn Neurosci 22(5):1026–1035

Smith JO (1992) Physical modeling using digital waveguides. Comput Music J 16(4):74–87

Stoelinga C, Chaigne A (2007) Time-domain modeling and simulation of rolling objects. Acta Acustica united Acustica 93(2):290–304

Thoret E, Aramaki M, Gondre C, Kronland-Martinet R, Ystad S (2013) Controlling a non linear friction model for evocative sound synthesis applications. In: Proceedings of the 16th international conference on digital audio effects (DAFx-13), Maynooth, Ireland

Thoret E, Aramaki M, Kronland-Martinet R, Velay J, Ystad S (2014) From sound to shape: auditory perception of drawing movements. J Exp Psychol Hum Percept Perform

Thoret E, Aramaki M, Kronland-Martinet R, Ystad S (2013) Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012), number 7900. In: Lecture notes in computer science, Springer, Berlin, Heidelberg, chapter reenacting sensorimotor features of drawing movements from friction sounds

Väljamäe A, Steffert T, Holland S, Marimon X, Benitez R, Mealla S, Oliveira A, Jordà S (2013) A review of real-time EEG sonification research. In: Proceedings of the 19th international conference on auditory display (ICAD 2013), Lodz, Poland, pp 85–93

van den Doel K, Kry PG, Pai DK (2001) Foleyautomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of SIGGRAPH 2001, computer graphics proceedings, annual conference series, pp 537–544

Van Petten C, Rheinfelder H (1995) Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. Neuropsychologia 33 (4):485–508

Vanderveer NJ (1979) Ecological acoustics: human perception of environmental sounds, PhD thesis, Georgia Inst. Technol

Varela F (1989) Invitation aux sciences cognitives. Seuil, Paris

Varela F (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3:330–335

Varela F, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA, USA

Verron C, Aramaki M, Kronland-Martinet R, Pallone G (2010) A 3D immersive synthesizer for environmental sounds. IEEE Trans Audio Speech Lang Process 18(6):1550–1561

Verron C, Pallone G, Aramaki M, Kronland-Martinet R (2009) Controlling a spatialized environmental sound synthesizer. Proceedings of the IEEE workshop on applications of signal processing to audio and acoustics (WASPAA). New Paltz, NY, pp 321–324

Viviani P (2002) Motor competence in the perception of dynamic events: a tutorial. In: Prinz W, Hommel B (eds) Common mechanisms in perception and action. Oxford University Press, New York, NY, pp 406–442

Viviani P, Redolfi M, Baud-Bovy G (1997) Perceiving and tracking kinaesthetic stimuli: further evidence for motor-perceptual interactions. J Exp Psychol Hum Percept Perform 23:1232–1252

Viviani P, Stucchi N (1992) Biological movements look uniform: evidence of motor-perceptual interactions. J Exp Psychol Hum Percept Perform 18:603–623
Warren WH, Verbrugge RR (1984) Auditory perception of breaking and bouncing events: a case study in ecological acoustics. J Exp Psychol Hum Percept Perform 10(5):704–712
Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain-computer interfaces for communication and control. Clin Neuro physiol 113:767–791
Ystad S, Voinier T (2001) A virtually-real flute. Comput Music J 25(2):13–24