Introduction
Music experience raises many complex questions that can be analyzed from different points of view. A piece of music can interest a musicologist because of its historical and stylistic relations with the cultural environment in which it has been composed. At the same time, another listener can be attracted to the same piece of music, because of the emotions or the sensations induced by that music. And for the same person, the experience can vary depending on the musical aspects on which she/he is focused. Another way of analyzing music experience is by asking listeners to describe the music using verbal labels. This method of information gathering about music experiences is often used, with many papers published reporting participants’ ratings of various adjectives, for example, while auditioning a piece of music. Many of these studies focus on the emotional experience (for a detailed review, see the volume edited by Sloboda & Juslin, 2010).

In addition to emotions, music can suggest or induce in the listeners a variety of images and sensations (see, e.g., Canazza, De Poli, Rodà, & Vidolin, 2003). Very often, these studies use verbal labels for describing the subjects’ experience when they listen to music. However, verbal descriptions only partially capture the musical experience. Non-verbal associations can be useful in allowing the assessment of non-verbal experiences, too. Such an approach may help overcome some of the limits of communication necessarily imposed by language.

We start from the assumption that the numerous ways to describe music experience do not exclude one another, but rather they may be complementary points of view of the same complex experience. The different metaphors can be suitable for different applications and contexts: for example, emotional aspects of musical experience can be useful in affective human–computer interfaces (Pantic & Rothkrantz, 2003); sensory-motor aspects can be useful in direct interaction and manipulation of contents (Luciani, Florens, & Castagné, 2005); psychological aspects of cross-modal experience may reveal mechanisms concerned with synaesthesia (Simner et al., 2006), and so on. The emerging view is that experiences of music are multi-faceted, consisting of different kinds or possibly interacting qualities, aspects, and dimensions.

A New Taxonomy of Music Expressiveness

As pointed out by Schubert and Fabian (2014), the terms expression and expressiveness can be used interchangeably, but expressiveness can be used in a more specific sense to mean musical expression rather than
emotional and other forms of expression. In particular, Schubert and Fabian presented a categorization (taxonomy) in which expressiveness can be conceptualized as having two kinds of content further divided into two layers respectively. The first important distinction is the differentiation between emotional expression (in either layer) and a self-contained, undefined expressiveness, which is called musical expressiveness or intransitive expressiveness. While emotional expressive content can be communicated and enhanced by expressive intentions of the performer (Juslin, 1997; Juslin & Laukka, 2001), musical expressiveness is characterized by an intransitive quality and a particular interest in performance practice (stylishness), without needing to express a specific emotion. Regarding the two layers denoting each kind of content, we can distinguish between one based on structural elements of the music that are determined by a musical score/composer—the compositional layer—and another based on features that can be added by the performer—the performance layer—such as a nuance in the performance that exaggerates a compositional element (e.g., overdotting a dotted note, extending the length of a suspension or appoggiatura), whether for the purpose of reflecting a performance practice or as a means of personal expression.

We propose that this taxonomy can be further developed by including a third kind of expressiveness content—sensorimotor expressiveness—which can be applied both to the compositional layer and to the performance layer. Sensorimotor expressiveness refers to aspects not covered by musical and emotional expressiveness, since it investigates the domain of cross-modal associations. The starting point is that metaphorical descriptions may offer possibilities to explain and understand aspects of the musical experience otherwise ineffable. According to Lakoff (1993, p. 245), metaphor is not simply a literary device but a basic structure of understanding, and “metaphorical descriptions of music are grounded in embodied experience.” Similarly, Zbikowski (2008, p. 511) considers metaphor “a kind of cross-domain mapping” and considers music a conceptual domain that can be drawn into such mappings. As pointed out by Eitan in two empirical studies, metaphor structures our understanding of music (Eitan & Granot, 2006), and the metaphors used to characterize musical relationships reflect the influence of culture (Eitan & Timmers, 2010).

Our experiments focus on the development and validation of new methods to investigate music perception without verbal measures. These methods are based on sensory qualities, kinetics/energy space, and action metaphors. In the section “Sensorimotor expressiveness described through verbal labels,” we present the results of studies based on the description of sensorimotor expressiveness through verbal labels. Then, the sections “Sensorimotor expressiveness described through action-reaction metaphors” and “Sensorimotor expressiveness described through cross-modal associations” focus on two recently introduced methods for studying sensorimotor expressiveness without using verbal labels, based on action-reaction metaphors and cross-modal associations. Finally, we discuss the results in a broader context and hypothesize new perspectives for future research in this domain.

**Sensorimotor Expressiveness Described Through Verbal Labels**

In the context of music performance studies, the term expressive intention has been introduced by Gabrielsson and co-workers (Gabrielsson & Juslin, 1996), with reference to the various emotion characters that a musician tries to express by performing a music score. Traditionally, expressive intentions in music refer to the emotional expression content, introduced in the previous section. Indeed, Gabrielsson’s work is focused exclusively on emotions, more precisely on discrete emotions, because listeners find it quite natural to attach general emotional labels to pieces of music. Nevertheless, listeners find it quite natural to use other kinds of labels to describe pieces of music. As an example, the rank of the most frequent tags freely attached to pieces of music by the users of Last.fm was investigated. Last.fm is a widely used online music repository and streaming resource (http://www.last.fm). The company has also released a public application programming interface (API; http://www.last.fm/api), which allows everyone to access the metadata of the repository and to count the frequency of the user-defined tags. Our investigation revealed that happy and sad are frequently used and are classified...
Giovanni De Poli et al.

in the 65th and 62nd places, respectively (on a total of about 500,000 listed tags), but other basic emotions such as fear, tender, and solemn were not so common and are ranked over the 2,000th place. At the same time, many other words are frequently used to tag music pieces: excluding the tags directly related to a particular musical genre (e.g., rock, jazz, or classical), we found frequently used adjectives such as hard (35th), cool (36th), heavy (54th), soft (106th), and dark (121st).

In several different experiments (e.g., Baraldi, De Poli, & Rodà, 2006; Canazza, De Poli, Rinaldin, & Vidolin, 1997), De Poli and colleagues extended the concept of the performer’s expressive intentions beyond emotions by including labels with sensorial connotations. Consequently, the studies about expressive intentions in music can refer also to the third kind of content (see the section “A new taxonomy of music expressiveness”), namely sensorimotor expressiveness. The authors asked several musicians to play a short piece of music several times, changing each performance in order to convey a different expressive intention, suggested by one of the adjectives: bright, dark, light, heavy, soft, hard, and natural. The natural performance was defined as “an interpretation in accordance with musical practice” but without any particular expressive intention. It provided a useful baseline to measure the performance variability due to other expressive intentions. The experiments involved different musical instruments (clarinet, violin, saxophone, and piano) and musical pieces, mainly from the Western classical repertoire but also some popular and jazz melodies. Then, several groups of listeners were asked to rate the recorded performances along a set of continuous verbal scales, denoting qualities the listeners felt the stimulus expressed.

The results showed that the performers’ intentions and the listeners’ impressions generally agreed. The listeners’ ratings were analyzed by means of multivariate techniques: two quite distinct expressive factors were observed, one related to the kinetics parameters and the other related to the energy of the pieces. From a musical point of view, the first factor depicts rapid tempo (bright and light pieces) against slow (heavy pieces) or moderate (soft pieces) tempo. The second factor, on the other hand, is mainly correlated with energy-related parameters such as intensity. The performances inspired by different expressive intentions are organized in a two-dimensional space named Kinetics-Energy space, and presents a point of departure from the commonly used two-dimensional emotion space (Eerola & Vuoskoski, 2011; Schubert, 1999). As will be shown in the next sections, other experiments carried out by De Poli and co-workers showed that, under certain conditions, music pieces are arranged by listeners along these two axes, which have consequently been interpreted as two basic dimensions for describing sensorimotor expressiveness.

Sensorimotor Expressiveness Described Through Action-Reaction Metaphors

There is a long tradition of studying gestures associated with music (Godøy & Leman, 2009). Embodied approaches to music (e.g., Franěk, van Noorden, & Režný, 2014; Leman, 2007) consider body movements as effective for communicating emotions and expressive intentions, feelings, and ideas both for the performer and the listener. When listening to music, people can mirror the expressive aspects of the music in the form of actual movement patterns (see also, Frances & Bruchenschweitzer, 1983). Thus examining the connections between music and movements can be fruitful in understanding expressiveness from a non-verbal point of view. Free and spontaneous human movements in response to music were studied by Camurri, Castellano, Ricchetti, and Volpe (2006), to investigate the correlations with the emotional characterization of music excerpts, and by Amelynck, Maes, Martens, and Leman (2014), who analyzed commonality and individuality of these movements. Clynes also examined micromovements in response to music via a hardware interface referred to as a Sentograph. The Sentograph captures finger movements from which one can deduce expressive properties in terms of predictable finger movement patterns, and Clynes spent much of his work in interpreting these patterns in terms of emotional expressions (Clynes, 1973; Clynes & Nettheim, 1982). However, his work is a clear precursor to modern conceptions of responding to music non-verbally.
In order to understand sensorimotor expressiveness without using verbal labels, we focused on the associations between music and human movements from the point of view of an action-reaction paradigm, by which we mean interactions with other objects through contact, pressure, and (usually small amounts of) motion. This kind of interaction can be simulated by haptic devices (i.e., devices that measure a user’s fingertip position, pressure, and motion and exert a precisely controlled force vector response to the fingertip). Such devices enable users to interact with and feel a wide variety of virtual objects (Massie & Salisbury, 1994).

To investigate this kind of interaction, we started from the basic mechanical concepts of Friction, Elasticity, and Inertia, grouped henceforth into the so-called FEI metaphor (De Poli, Mion, & Rodà, 2009). From a physical perspective, ideal Friction dissipates energy and acts as a scaling factor for the input action, while ideal Elasticity and Inertia store Energy: in particular, Inertia stores kinetic energy and opposes changes in movement while Elasticity stores potential energy and opposes changes in forces. To control these parameters, we designed three elementary haptic reference stimuli synthesized by means of a Phantom Omni haptic device (http://www.sensible.com/haptic-phantom-omni.htm). The tip of the haptic device, when moved by the user’s fingers, produces respectively a frictional, elastic, or inertial force reaction. From the user’s point of view, these reactions provide the impression of interacting with real objects (e.g., the elastic force response simulates the behavior of a spring, whereas the frictional response simulates the movement of an object in a viscous fluid).

The objectives of our experiments (De Poli et al., 2009) were to: (a) determine whether the FEI metaphor is able to describe some aspects of the musical experience both at compositional and performance levels and (b) provide evidence for relationships between the FEI metaphor and other metaphors, in particular the emotional and the sensorial. In the experiments, participants were asked to listen to various musical excerpts and to associate each one to one of the three haptic responses. While most research on music expression deals with emotions, expressive intention is a broader concept that includes emotions as well as sensorimotor aspects. Thus, we studied expressions both from the emotional and the sensorial domain, in order to derive a more general organization (and possibly interpretation) of similarities among the expressive intentions of the two domains. We used as stimuli simple melodies played by a performer according to the categories Happy-Sad (high and low valence), Angry-Calm (high and low arousal) for the affective domain, and according to the categories Hard-Soft (high and low energy) and Light-Heavy (high and low kinetics) for the Kinetics-Energy space. The results of the experiments showed that subjects are able to associate each haptic response to differentially expressive musical stimuli. In particular, statistically significant relationships between Elasticity and the happy/light stimuli, between Inertia and the sad/calm stimuli, and between Friction and the hard/anger/heavy stimuli were found. Haptic responses are therefore reliable tools that use physical interaction to describe expressive characteristics of music. Furthermore, the same clusters were also found in a previous study (Mion & De Poli, 2008) where acoustic cues were used to group music performances inspired by different emotional and sensory expressive intentions. This result supports the hypothesis that the three clusters correspond to different expressive categories that can be characterized both from an acoustic and perceptual point of view.

Moreover, the FEI metaphor is not strictly related to musical contents, so that the results could be applied also to other non-musical contexts. For example, the approach of the action-reaction metaphor can be explored in artistic forms based on gestural dynamics, such as painting or dancing (Luciani, 2009).

Sensorimotor Expressiveness Described Through Cross-Modal Associations

As evident from the previous sections, the capacity of music to elicit a rich number of often shared sensations and images can be attributed in part to cross-modal correspondences, which may be intrinsic attributes of our perceptual system’s organization (Marks, 1978). Many studies investigated the
relationship between music versus colors, tastes, vision, and odors, suggesting that people can exhibit consistent cross-modal responses in different sensory modalities (Ward & Mattingley, 2006).

According to Spence (2011), cross-modal correspondences between different combinations of modal and amodal features are mediated not only by innate factors but also by experience and statistical factors, so they need to be considered alongside semantic, spatiotemporal, developmental, and affective dimensions. In particular, Spence distinguishes among structural, statistical, and semantic cross-modal correspondences and hypothesizes that structural correspondences may bear fruitful comparison with synaesthesia, since they can reflect an unlearned aspect of perception, whereas other correspondences may reflect the internalization of the statistics of the natural environment or may appear semantically mediated. For example, Ward and Simner (2003) have reported an unusual case of developmental synaesthesia in which speech sounds induce an involuntary sensation of taste that is subjectively located in the mouth. The findings from this case study provide important implications for the understanding of the mechanisms of developmental synaesthesia, since they emphasize the role of learning and conceptual knowledge.

In this complex and multifaceted research field, sensory scales, which we present as non-verbal bipolar scales taken from the visual, gustatory, haptic, and tactile domains, can represent a reliable tool for investigating the relationship between sensory perception, emotions, and musical stimuli. Starting with the semantic differential approach, sensory scales were adopted by Da Pos et al. (2013), who hypothesized that asking a participant to express a judgment on the warmness or coldness of a stimulus is different from asking her/him to express this evaluation by directly immersing her/his hands into warm or cold water. If we consider, for example, the word blue from the semantic point of view, its meaning can be connected to sadness (I feel blue) or to a musical style (blues) or to the color blue. On the other hand, the large and problematic number of semantic associations can be reduced if we select those sensorial concepts that are less ambiguous. We propose that sensory scales are indicative of two kinds of cognitive processes: (a) the synaesthetic (or categorical correspondence), in which the connection between hearing sound as colors, textures, etc. is direct and (usually) involuntary and (b) the metaphoric, in which sensory modality terms are employed to enrich verbal descriptions.

The sensory domains covered by our sensory scales were visual, tactile, haptic, and gustatory. From the visual domain we adopted the scales maluma-takete, the couple of pseudo-words invented by Köhler (1929) in order to show the non-arbitrary mapping between speech sounds and the visual shape of objects, and orange-blue, the visualization of the two colors in two cards. These two colors are not perfectly complementary, but they were chosen as representatives of presumably warm and cold colors, respectively.

From the tactile domain, we selected the scales smooth-rough (sheets of sandpaper in different grit sizes), hard-soft (a piece of wood and a cylinder of polystyrene foam), cold-warm (one cup of cold water and one cup of warm water), and tense-relaxed (iron wire covered with cloth and rubber band covered with cloth).

From the gustatory domain, we selected the scale bitter-sweet: a bitter and a sweet substance. In an experiment conducted by Crisinel and Spence (2010), cross-modal associations between tastes and words were investigated. Results showed that sugar was associated with the soft-sounding nonsense words “maluma” and “lula,” but not with “bobolo” and “bouba,” suggesting that the presence of the consonants “m” and “l” was an important factor in participants’ choice. Our experiments replicated the maluma coupling with sweet and with soft.

The haptic scale we employed was heavy-light: a dark plastic bottle full of liquid and the same bottle without liquid. The dull color of the bottle did not allow participants to discriminate between the full and the empty bottle.

In Murari, Rodà, Canazza, De Poli, and Da Pos (2015), we designed two experiments (one with musical excerpts in major tonality and one with musical excerpts in minor tonality) in order to investigate the difference between Osgood’s semantic differential (Osgood, Suci, & Tannenbaum, 1957) and
a new sensory differential based on the direct perception of continuous measures. The same procedure was applied by Da Pos and Pietto (2010) in the field of normal and iridescent colors and in the study of unique hues. Musical excerpts in minor and major tonality had already been categorized in experiments by Bigand, Vieillard, Madurell, Marozeau, and Dacquet (2005) by means of multidimensional scaling and by Rodà, Canazza, and De Poli (2014) by means of cluster analysis. The employment of sensory scales was considered a way of obtaining more refined results able to go beyond music emotions.

In Murari et al. (2015), we planned to directly compare sensory scales and their verbal equivalent scales in a setting in which participants were divided randomly into two groups to avoid order effects. One group completed the set of sensory scales first (SV-group) and the other group completed the set of verbal scales first (VS-group). The order in which the rating task was completed (verbal scales first versus sensory scales first) impacted on the ratings. In particular, the results of factor analysis performed on the SV and the VS group separately revealed a relatively stronger juxtaposition between sensory and verbal scales in the SV group, suggesting that sensory scales can be influenced by verbal labels, and that conversely, the absence of those labels still makes an important contribution to explaining variance in response. Moreover, factor analysis was performed on the two groups of participants combined, and this also suggested an even more accentuated juxtaposition of sensory and verbal scales, since they grouped into different factors. Results showed that asking a participant to evaluate a piece of music by merging her/his hands into warm or cold water is different from verbally stating if that piece of music is warm or cold. The Chopin excerpt (Prelude no. 22), for example, was rated cold only in the sensory scale condition, and the Mozart excerpt (Adagio from Concerto K 488) was rated bitter only in the sensory scale condition.

Such results confirm that sensory scales allow an alternate understanding of the musical experience, since they index different aspects of the musical experience that are not accessible to natural language. The consistency of results confirms that the sensory differential tool created for our experiments is reliable and provides a novel way of allowing qualities of music to be differentiated. Moreover, sensory scales can provide fruitful results in the fields of cross-modal associations and synaesthesia.

Discussion and Conclusions

We propose that the taxonomy of expressiveness in music proposed by Schubert and Fabian (2014) should be augmented by including sensorimotor content in addition to emotions and musical content. The different experiments described in this chapter show that by approaching the musical experience from an action-based metaphor perspective, both haptic and sensory, new insights can be gleaned. Such approaches have the advantage of being applicable to different contexts (color perception, artistic forms based on gestural dynamic, such as painting or dancing), to move beyond verbalization of emotions, and to mitigate some of the other problems associated with the use of verbal labels.

As pointed out by Karbusicky (1987, p. 433), “thought in music occurs primarily in asemantic shapes and formulas.” Any attempt to interpret these shapes and formulas through language or linguistic theory would ultimately fail to capture the substance of musical thought. Karbusicky’s notion of purely musical metaphor is exemplified by the fact that disparate musical material can be brought and combined together to generate new meaning in infinite ways. In particular, the sensory differential provided by sensory scales, as compared to Osgood’s semantic differential, may serve as a useful tool to offer a more complete perspective not only about the meanings of affect terms but also in order to present an alternative, low-dimensional response space and to investigate perceptual aspects of synaesthesia and cross-modality.

With regard to the way in which ratings of musical excerpts can be collected, it would be interesting to develop a blended modality instrument in which collection of emotional responses could be augmented with sensory scales continuously in real time as music unfolds and at the end of the piece (for a review of how this has been done for emotion in music, see Schubert, 2010). Data can be
collected by sensorimotor devices, and the comparison of results deriving from the different methods can inspire new and deeper insights into essential responses to music (see, e.g., Nielsen, 1987).

From a technological point of view, the understanding of metaphors for describing the different aspects of music (affective, sensorial, or physical) supports the development of applications for the interaction with musical contents: the relations between acoustic cues and metaphors can be employed to enhance the interface of devices that use gestural interaction with musical content, such as portable music players or musical video games. Moreover, research on sensory scales can foster the development of innovative interfaces to browse audio digital collections. These new devices will allow users to interrelate in a spontaneous and even expressive way with interactive multimedia systems, relying on a set of advanced musical and gestural content processing tools adapted to the profiles of individual users, adopting descriptions of perceived qualities, or making expressive movements.

From a theoretical perspective, there is much research to be done. For example, one of the studies we reported suggests that sensory scales can produce meanings that are distinct from their verbal label equivalents (selecting an actual color rather than the verbal label for the color, or feeling a warm bottle rather than reading the word *warm*). But some words are more likely to map onto sensory experience better than others, and some sensory experiences are likely to map onto words better than others. Identifying when these mappings do and do not occur will provide important data in the understanding of cognitive processing of the stimulus under consideration, and in further developing the tools to measure the experiences. Verbal scales are easy and efficient to deploy, whereas haptic and sensory scales are more labor and resource intensive, so while investigations of music from a multi-sensory perspective continue, we need to continue efforts in determining the efficiency of the new measures. Furthermore, we need to find out how important the multi-sensory experiences are in the sum total of the musical experience. Do people care if music tastes bitter or sweet (sensory) versus sad or joyous (emotional)?

It should also be noted that music provides an important kind of stimulus for investigating multi-sensory experience, because unlike language it is not restricted to denotative meaning, and unlike some more mundane stimuli (colors, single tones, etc.) music listening is considered by many to be a highly meaningful, all-encompassing activity.

The discovery of new ways of reporting musical experience promises more ground to be made in music psychology and philosophy, in particular because we are able to gradually hone in on the otherwise ineffable aspects of music, what some may consider to be a holy grail and even untouchable (see, e.g., Dickie, 1997).

**References**


84
Beyond Emotion


