Research article

Embodied responses to musical experience detected by human bio-feedback brain features in a Geminoid augmented architecture

Rosario Sorbello, Salvatore Tramonte, Carmelo Calì, Marcello Giardina, Shuichi Nishio, Hiroshi Ishiguro, Antonio Chella

INTRODUCTION

Emotions
Cognitive architecture
Social robot
Android robot
Geminoid
Brain Computer Interface (BCI)
Human Humanoid Interaction (HHI)

Keywords:
ARTICLE INFO

Human Humanoid Interaction (HHI)
Brain Computer Interface (BCI)
Geminoid
Android robot
Social robot
Cognitive architecture
Emotions

ABSTRACT

This paper presents the conceptual framework for a study of musical experience and the associated architecture centred on Human-Humanoid Interaction (HHI). On the grounds of the theoretical and experimental literature on the biological foundation of music, the grammar of music perception and the perception and feeling of emotions in music hearing, we argue that music cognition is specific and that it is realized by a cognitive capacity for music that consists of conceptual and affective constituents. We discuss the relationship between such constituents that enables understanding, that is extracting meaning from music at the different levels of the organization of sounds that are felt as bearers of affects and emotions. To account for the way such cognitive mechanisms are realized in music hearing and extended to movements and gestures we bring in the construct of tensions and of music experience as a cognitive frame. Finally, we describe the principled approach to the design and the architecture of a BCI-controlled robotic system that can be employed to map and specify the constituents of the cognitive capacity for music as well as to simulate their contribution to music meaning understanding in the context of music experience by displaying it through the Geminoid robot movements.

INTRODUCTION

Music is as much stably diffused across history and cultures as it is widely debated with respect to its cognitive function, mechanism and adaptive functions. Theoretical biology, philosophy of mind and of perception, psychology, neuro-sciences, cognitive sciences and artificial intelligence have been taking part in the research that aims at settling the questions about the nature, function and meaning of music or selected properties of musical composition and understanding. Is there a capacity for music that includes the cognitive resources to process and understand music has it an adaptive evolutionary value as Jackendoff and Lerdahl (2006) suggested or is it a spandrel as assessed by Bigand (1990).

Is it even the cultural specialization of what Pinker (1997) has called an "auditory cheesecake", that is to say an apparent cognitive function that in reality does not respond to any biological cause or effect so that it could vanish from our species without consequences unlike language, vision, social reasoning and physical know-how. If musical capacity is admitted, is it decomposable into distinct cognitive resources? Are those cognitive resources music-specific, specialized from other cognitive functions (language, auditory perception) or even reducible to the interaction of general-purpose cognitive capacities? The research to settle those questions meets the issues that are discussed in the long-established disciplines of music theory, musicology and ethnomusicology as well as in music learning and education fields. In connection with these issues this paper is organized as follows. In the first section we discuss some theoretical and experimental considerations on the biological foundations of music and the cognitive structures.

In the second section, we extend the discussion of music specificity to emotion expression and arousal. In the third section, we describe the architecture we designed to simulate how the constituents of musical capacity act in musical experience and test the validity of their models using the Humanoid Robot Geminoid.

The specificity of music cognition: Biological foundations and musical grammar

Pinker (1997) draws on the assumption that no every feature that occurs during the evolution must satisfy a function, or in general
increase fitness, in order to argue that music is an activity assembled from the cognitive functions that accomplish biological goals while it is functionally meaningless. As cheesecakes, hardly the result of evolution, pack many agreeable sensory stimuli together unlike anything in the natural world, so that the mind experiences the patterns of sight, taste, smell at once by which environment has stimulated the cognitive functions that contributed to fitness, so music borrows its compositional and design features that give listeners pleasure from language, auditory heuristics that serve as inverse acoustics, emotional calls, attention to the attributes and changes that signal safe or unsafe habitats.

If the mind is a neural computer, whose design is the outcome of the natural selection, music is a by-product that uses the toolbox of combinatorial algorithms for causal and probabilistic reasoning on plants, animals, objects and people. On the contrary Peretz (2006), argued that congenital amusia disorders point to the biological foundations of music. Subjects who present such disorders are unable from the birth to perceive melodic contours or pitch changes, to recognize melody, to perceive dissonance, to sing in tune, to hear a sung or a played “wrong” notes, because mistuned or out-of-key. Those deficits are also dissociated from language. Subjects affected by amusia can recognize the lyrics of songs, which are difficult to be recognized when hummed, and recall the lyrics of familiar song that they cannot sing in tune. Amusic subjects can rely upon the general strategy to employ the surrounding pitch distributions to learn novel syllabic structures. It is likely that music features cognition does not depend on general learning processing, rather on processing that implies computational domain and input code specificity. Domain specificity does not mean brain localization. It is also consistent with neural network architecture. In this connection, a network of neural circuits from the right inferior frontal-temporal area to the right auditory area has been hypothesized to underlie the encoding system, whose lesions are causally related to amusia as reported by Schuppert, Münte, Wieringa, and Altenmüller (2000) and Stewart, von Kriegerstein, Warren, and Griffiths (2006).

Similar specificity of music cognition emerges from the research of the Cognitive Sciences into the processing systems and representation structures of the features of sounds that are relevant to music. This research addresses the question: what does it mean to “hear” a piece of music? Jackendoff and Jerod (2006). Hearing music in proper sense implies understanding it, hence hearing music means using cognitive resources to recognize the structures embedded in the stream of musical sounds to reconstruct the rules that account for the organization of pitch, rhythm, meter, dynamics, duration. As for language, one can build a formal model of these rules, which allow the listeners deriving the structures, the relation of dependence and inclusion from the linear order of sounds. The model is equivalent to the grammar of music and it amounts to the cognitive principles underlying the “capacity for music”. The model of Jerod and Jackendoff (1983) consists of a set of “preference rules” that decide the grouping of notes, i.e. the segmentation of surface sound events into units in which notes belong to one another and the cognate determination of boundaries between them, along with the metric distribution of accents, the reduction of sounds unfolding in time to the dominant ones given their harmonic relations, the prediction of the degree of tension and relaxation perceived for each sound given its distance from the harmonic or metric function assigned by such dominance relations. Such a model accounts also for rhythm as the ruled combination of grouping and meter.

The rules of reduction and the pattern of tension and relaxation can be mapped onto a tree structure that mirrors the melodic and harmonic structure of the musical piece embedded in the sound surface. The rules work at many hierarchical levels so that they allow recursive structures. Supporting evidence that listeners actually extract the structures as predicted by this model has been provided by Palmer and Krumhansl (1987), Bigand (1990) for tonal music. The rules of the grammar are connected with the cognitive organization of the tonal pitch space. Krumhansl (2001) puts forth a cognitive model in which perceived musical pitch structure depends on the position of tones, chord and keys in the tonal hierarchy. It is noteworthy that even grouping is founded on principles of perceptual segmentations that account for the order that one can extract from sounds in the environment as it is shown by the experimental work of Bregman (1994). This shows that the domain specificity of a cognitive component may emerge from specialization. It can be even assumed that the perceptual common basis for specialization is connected also with the evidence on the common general principles which Carterette and Kendall (1999) have been found to underlie tonal scale systems across cultures.

The affective constituent of musical meaning: emotion and kinematics

The question of what does it mean hearing music cannot be thoroughly addressed if emotions or affective responses recognized or elicited by it are not taken into account. Scherer and Zentner (2001) brought in the distinction between perceived and aroused musical emotions. One may indeed recognize the emotion or the affective state conveyed by the musical piece without feeling it as well as one may feel an emotion contingently upon listening to music. Sloboda and Justl (2001) and Gabrielsson (2009) underlined the mechanisms and the capacity involved in those two cases may be different. Indeed there is evidence that the relation between perceived and aroused emotions is not a one-to-one mapping. Zentner, Grandjean, and Scherer (2008) asked subjects to rate which emotions among a large data set they more frequently either perceived or felt when listening to music. They found that some emotions are more likely to be recognized than others as response to music when perceived rather than felt.

As regards perceived emotions, Gabrielsson (2003) found that subjects consistently agree that music expresses emotions and their judgments on the emotions, which are perceived in musical pieces, are reliable, robust and only marginally affected by musical training, age and gender. Furthermore, those judgments are to some extent systematically coupled with certain combination of musical features, so that they could be predicted with reasonable accuracy. Justl (2000) reports that tempo, timing, tone attacks and decays, timbre, accent, rhythm, pitch and pitch contour, tonality, harmony, dynamics, mode, articulation are simple musical features that convey emotions like happiness, sadness, anger, fear and tenderness. More complex features like harmonic progression, melody and musical form need further experimental scrutiny.

In fact there is evidence that the relation between perceived emotions and simple musical features is correlational. Simple features must be associated in configurations to convey one emotion and the same feature may enter different configurations. However the lack of a one-to-one mapping is not contrary to the evidence of robust musical expression of emotions. Justl (2000) presents a model of additive combination of features with redundancy fits the reliability of emotion perception in music.

Perceived emotions can be studied as discrete or continuous phenomena. In the first case, attention is paid to basic emotions which, according to Plutchik (1994), are joy, sadness, anger, fear. In the second case, emotions are represented as something that arise and evolve continuously. Perceived emotions are decomposed in dimensions like activation or valence and are modelled according to theories that map them onto a conceptual space as that put forth by Russell (1980). Schubert (2004) employs a two-dimensional space to investigate the relation between features like melodic contour, loudness, tempo, timbre, texture and emotions perceived in four pieces of Romantic music.

As regards aroused emotions too, Gabrielsson (2003) and Sorbello et al., (2017) points out that most subjects reports that they repeatedly feel emotions during music listening. It is highly significant that some landmark studies in this respect have found evidence of consistent agreement of physiological indicators of affective responses, emotions classifications and the pattern of tension and relaxation expected on the
theoretical grounds of the preference of the musical grammar. Sloboda (1991) asked subjects to identify musical pieces that in the past induced in them somatic reactions associated with emotions. It was then found that musical features like appoggiatura or melodic and harmonic progression, syncopation were associated for a third of subjects respectively with tears or heart reactions, while unexpected harmony promoted shivers. For the remaining subjects, a robust correlation was found between somatic responses and the overall musical character of the piece.

Krumhansl (1997) and Krumhansl and Greer (2000) made physiological measures on subjects that listened to musical excerpts selected for their contrasting characteristics or to Mozart Piano Sonata in E-flat major K. 282. In the first case measures were compared with the results of an emotion classification task. In the second one measures were compared with the tension generated by the consistency of the sounds with the root in the second one. A response of higher tension was expected as sounds consistency decreased. The correlation found was significant in both cases and the change in the physiological and classification responses could be considered musically specific.

From the results presented in the previous sections the conclusion may be drawn that emotions are perceived or aroused while listening to music because the musical features contain information that is processed by subjects according to a structure that to some extent is correspondent to the changes of state associated with emotions. Perceived and aroused emotions involve the feeling of tensions and relaxations, which have as somatic an amplitude, speed, direction as the corresponding physical parameters motions have.

As early as Michotte (1950) emotions are considered as a change of state felt by subjects with reference to something internally or externally sensed, which can be mapped as the variations of specifiable parameters in space and time. Such a change can be then either perceived or even recognized in motions and actions as it was first proved by Heider and Simmel (1944) who tested the attribution of intentions, emotions and agency to moving geometric shapes. A growing amount of studies is being devoted to the function of performers movements in effectively communicating emotions by emphasizing musical features Gabrielson (2003), Juslin and Sloboda (2001) found that particular gestures match the acoustical features that are relevant for emotion expression through music. Palmers and Krumhansl (1987),anderley (2002) report that performers gestures, which are not directly connected to sound generation or to optimizing feedback, highlight important musical properties of the piece being performed and contribute to achieving expressing goals.

In accord with Michotte (1950) theory and the research of Johansson (1973) into the biological motion, the perception of the kinetic qualities of movement and their configuration seems to be crucial for how performers movement are connected to emotion expression. Dahl and Friberg (2007) isolated some movement cues picked from the movements of performers body parts (face, arms, trunk) and their changes according to the parameters of their direction, duration, tempo. Then they displayed these cues to subjects through video clips with and without the audio signal and the subjects had to rate their meaning in terms of expressed emotions. A consistent coupling of acoustic and movement cues with emotions was found. Canazza, De Poli, Rodà, Vidolin, and Zanon (2001) asked subjects to match some terms denoting perceptual qualities and the emotions perceived in excerpts of Mozarts Clarinet Concerto K 622, each played with different intended expression. From the analysis of responses they extracted the kinematics (tempo variation) and energy (intensity variation) as the two dimensions of the cognitive space of subjects experience of musical expression.

On such theoretical and experimental grounds, it can be said that hearing music means extracting the grammatical and the emotional constituents of musical meaning and that music experience is the dimension at which the connection between musical grammar and affective experience is realised. As early as Meyer (1956) grammar and affect in music have been considered the components of the meaning that composers try to convey and listeners try to understand. As sounds unfold over time, the grammar sets bottom up rules of admitted combinations and transformations and those rules, in turn, induce a top down expectancy on how music should be continuing at each given moment. The intended violation, delay or confirmation of such expectancy causes listeners to perceive or feel the tensions and relaxations, which are connected to the specific musical features and to the direction and amount of their alteration. For instance scale passages induce the expectancy of a pitch change in the same direction or with the least deviation as possible, while hearing a large interval gap leads to expect a change of direction. Musical experience is thus a robust dimension that can be modelled. On the one hand, there is agreement between the properties of repose, tension, relaxation of musical pieces, given specific tonal and dynamic features, and subjects rating of maxima or minima of felt intensity in music. On the other hand, the salient features, which induce the tension and relaxation interplay, can be specified and tested by models.

Narmour (1992) provided a theory to translate in principles of melodic shapes the insights of Meyer (1956) on the bottom up conditions of expectancy.Lerdahl (2001) put forth the theory of tonal attraction to account for expectancy. The attraction between two successive notes of definite pitch in a melodic context is related to their harmonic stability and inversely related to the square of their distance. Lerdahl and Krumhansl (2007) has found significant correlation between the tensions predicted by this model and those reported by subjects.

The proposed architecture

The principled approach to the architecture design

The models and the experimental results discussed in the preceding sections point out that there is a cognitive capacity whereby subjects understand music. The conceptual constituents of this capacity allow deriving the melodic, harmonic, rhythmic regularities from the linear streaming of sounds, i.e. the structure embedded in the musical surface. The structure consists of preference rules for the grouping of notes and for the assignment to them of functional roles according to the selected sound features. In analogy with formal theories of language, some models have been proposed to account for the derivation of the rules from sounds and for the reduction of note sequences to those which have a dominant melodic, harmonic or rhythmic role according to a specifiable distance function. It is noteworthy that this cognition amounts to the ability of extracting order from the musical surface, which is likely to be unconscious and stable across genres and individuals with or without musical education, at least at a fundamental level.

The affective constituents of the musical capacity allow recognizing the emotions conveyed by the changes of sound features admitted by the musical rules, be they felt or merely perceived. Although the relation between emotions and musical sounds is not a one-to-one mapping, it is so robust and stable that it can be predicted with confidence. Models have been proposed for the correlational reliability of the emotional attributes of music and for the conceptual space onto which the continuous dimensions of emotions are mapped. The conceptual and the affective constituents of the capacity for music are the cognitive mechanisms dedicated to extract meaning from music. They account for what it is hearing and understanding music. In light of the theoretical and the experimental literature, the constituents are likely to work in parallel in the sense that listeners recognize musical changes at the macro-level (melodic and harmonic relations, dynamics, rhythm, tempo) and at micro-level (attacks, notes onsets) as affective possibilities, that is to say as bearers of emotions with the associated degree of arousal. On the one hand, the changes usually occur in patterns, rather then isolation.
On the other hand, the changes satisfy or frustrate the expectations that listeners build as they derive the preference rules governing the ordered unfolding of the sounds in the musical surface. The construct of tension has been variously introduced in the literature to account for the connection between conceptual and affective understanding of music. According to the preference rules sounds are attracted towards or pushed away from one another or from a dominant reference point. For example the leading tone requires to be followed by the tonic. Since the preference rules work at many levels, the notes always display a tension value according to the melodic, harmonic, rhythmic role assigned to them, to the character of being relatively dominant, superordinate or subordinate, to the distance from the reference point, and finally to their expected or unexpected changes. Dominant tones and expected changes are perceptually associated to low tension or relaxation, whether they are heard as the starting or the end point of a sequence, while subordinate tones, tones that are distant from the dominant one in the relevant musical dimension, rules violations and unexpected changes are perceptually associated to high tension. Emotions are in turn a change of state in listeners that can be also described in terms of tensions, which accounts for the correlation between some patterns of musical changes and some physiological indicators of emotions. The connection between the conceptual and the affective constituents of the capacity for music is found also in the movements that performers make to amplify the expressivity of playing. In this case, the motion cues are describable in terms of forms and degrees of tensions with which the listeners are acquainted.

We hypothesize that such kinematic and dynamic properties of tension and relaxation can be found in patterns of musical sounds, musically perceived and aroused emotions, perceived and realized movements with the restriction that this matching is afforded in the musical experience and, most importantly, that music be considered as a cognitive frame Goffman (1974). In experiencing music the meaning is implicitly and immediately understood as different from ordinary perception and emotional response in everyday life. What is perceived, or felt from composers and listeners affords the cognitive frame of reference in which the matching between perceptual, cognitive and affective constituents of meaning can be hold.

To investigate the factors of the cognitive capacity for music we have designed a robotic integrated architecture. Many computational and robotic models of music cognition have been developed, for which Burger and Schmidt (2009) suggest methodological considerations. Friberg (2008) overviews the machine learning techniques that have been applied to large musical data sets in order to predict emotional expression. Kapur (2005) emphasizes that musical robotics has been focusing since long time on building automated mechanical sound generators. Furthermore, machine musicianship has addressed the design and the development of artificial agents that analyse, perform, compose music according to the principles of music theory and even interact with human performers (for instance see Weinberg (2007), Giardina et al. (2017)). Our architecture aims instead to map the music experience and to specify the type and the degree of correlation between the conceptual and the affective constituents of musical understanding in this cognitive frame. The architecture is a system that consists of distinct modules, each of which can be used to generate tests for the combination of musical features, for their correlation to perceived or felt emotions, and to simulate the associated tensions and relaxations through the kinematics of the movements of the humanoid robot, which are operated by the BCI-signals Sorbello et al. (2017) of subjects listening to music.

In accord with Jackendoff Jackendoff and Lerdahl (2006) we assume that, in the frame of musical experience and understanding of subjects, the pattern of tension and relaxation that unfolds in time according to rules is sufficient to derive the musical structure and in turn recognize the emotion by means of a shared kinematics. We expect that the Geminoid humanoid robots can display such derivation and recognition of the perceived or felt emotions by means of movements, which are controlled by the signals extracted by BCI modules from listeners.

The architecture design

The system is based on UnipaBCI framework Tramonte, Sorbello, Giardina, and Chella (2017) and the architecture, as shown in Fig. 1, is a modular system composed by five modules: A musical stimuli interface, A BCI acquisition and Pre-Process module, A Biological BCI Parameters Extraction and Geminoid Controller.

The musical stimuli interface provides the musical input selected randomly from a list of musical pieces, which were collected according to the sound feature or to the combination thereof that may be of interest: pitch, intervals, dynamics, melody, harmony, rhythm. It
synchronizes brain signal acquisition, tagging the EEG signal, acquired in real time from the Brain acquisition module, with labels related to the events (beginning and ending of a musical piece).

The BCI acquisition and pre-processing module consists of the tools to record the EEG data from subjects listening to music and to process the signal. Users wear a standard 10-20 EEG cap, with wet electrodes located in FP1, FP2, F3, F4 location of scalp. We have connected the ground at Oz position and the mass on the right earlobe. Users are connected to a bio signal amplifier, which convert the electrical activity of the brain from analogical to digital and ensure synchronization. The bio signal amplifier used for experiment is g.usb.a.m, produced by gtech.

BCI pre-processing module acquires signals in real time and aggregated in windows of 15 s. The signal was notched at 50 Hz, band passed between 1 and 60 Hz and digitized a 256 Hz. The signals were filtered using a winisor filter to limit extreme values to reduce the effect of EOG artifacts.

The Biological BCI Parameters Extraction module carries out the analysis of the data to extract the parameters of the music experience correlated with the conceptual and the affective constituents of music understanding, which are likely due to the features of the musical input. We decided to extract those parameters from the alpha and theta EEG frequency bands on the grounds of the works that employ them as proxy signals for cognitive measures in affective contexts because of the reduction of error due to noise in such frequency range (see for instance Rached & Perkusich (2013)). The parameters have been extracted from the signals aggregated into one signal as mean of the original signals using the “Grand mean” technique to obtain a single channel signal:

\[
X = \frac{X_1 + X_2 + X_3 + X_4}{4}
\]

(1)

where \(X_i\) is the filtered \(i\)-th channel.

The signal obtained from the mean process is therefore decomposed into its constituent frequencies using the Daubechies D4 Wavelet to extract the bands of interest to associate the emotions. We decomposed the signals in the constituent rhythms theta, alpha, beta, gamma and delta. Only alpha and theta bands were selected for our study.

The parameters extracted from the signal are construed in terms of energy and entropy. Energy is considered as the distribution of energy features of EEG signal at different resolutions accordingly to the Parseval’s theorem Omerhodzic, Avdakovic, Nuhanovic, and Dizdarevic (2013):

\[
ED_l = \sum_{j=1}^{N} |D_j|^2 , \quad i = 1,...,l
\]

(2)

\[
EA_l = \sum_{j=1}^{N} |A_j|^2
\]

(3)

where \(i = 1,...,l\) is the wavelet decomposition level, \(N\) is the number of coefficients, \(ED_l\) is the energy of the detail at decomposition level \(l\) and \(EA_l\) is the energy of the approximate at decomposition level \(l\). Entropy is considered as the measure of the level of chaos in the emotional activities of the brain Aminoff (2013), Jie, Cao, and Li (2014) and is defined in Eq. (4):

\[
H(X) = -c \sum_{i=1}^{m} p(x_i) \ln(p(x_i))
\]

(4)

\[
\sum_{i=1}^{m} p(x_i) = 1
\]

(5)

where \(c\) is a positive constant, \(x_i\) is a 256 Hz sample of brain signal, \(p(x_i)\) is the probability of \(x_i \in X\) with \(X\) defined as a set of finite random variables. \(x_i\), moreover, must satisfy in (5). We assumed that the values and the relationship of energy and entropy correspond functionally to the experience of the affective and the conceptual constituents of music cognition.

The parameter of energy represents the affective response of the subject, which varies continuously along the perpendicular axes of arousal and valence in the model of Russell (1980). The parameter of entropy represents the conceptual constituent of the response of the subject, intended in terms of the perceived complexity of the musical input due to the selected sound features, their combination, their development in the course of the musical sequence and the satisfaction or the violation of the preference rules.

Therefore that the combination of these two values represents the engagement of subjects in the musical experience, that is to say the investment of the cognitive resources in understanding music weighted against the particular sound features involved and the correlated affective attribution. The robotic controller module is the humanoid Geminoid.

The Geminoid Controller integrates the connection with the Geminoid Robot. Geminoid is a robot that will work as a duplicate of an existing person. It appears and behaves as a person and is connected to the person by a computer network Nishio, Ishiguro, and Hagita (2007).

The Geminoid is controlled by subjects on the basis of the BCI signals recorded during music listening. The function of this module is displaying the subjects engagement in music experience through the movements of the robot, Spataro et al. (2017) which match the emotions and the tensions perceived or felt by them as represented by the entropy and the energy parameters extracted by the EEG.

The matching is obtained by allowing the controller to develop a kind of movement language by which to impart motor commands to the robot on condition that they are significantly correlated to music listening. This language is developed during the training and the calibration sessions run in parallel. In the training session subjects listen to a series of musical stimuli, each lasting three minutes, which are interleaved by 10 s interval of silence every 10 s. At the same time, subjects are asked to impart motor commands to the robot, which are consistent with the tension or relaxation and the affective arousal felt while listening to music.

In connection with the musical stimuli interface, the BCI acquisition module labels the brain signals recorded in the two conditions respectively as target and silence. In the calibration session the controller sends target signals as motor commands to the robot every 650 ms to keep the synchronization between her brain activation and the joints movements that realize the correlated movements by means of the robot actuators (all values were normalized between 0 and 255 ms). The potential movements of the robot are mapped to the joints and the actuators according to a four state configuration Fig. 2.

At the end of the training phase for each subject, the variance of the BCI signals is computed and a threshold value is set to select the share of changes in entropy and/or energy for the alpha and theta frequency bands, which is significantly correlated with music. The threshold \(r^2\) establishes in real time the baseline condition with respect to the silence signals, above which the target signals sent to the robot are executed. According to the proportion at which the parameters value of the target signals is above the real time threshold \(r^2\) the robot executes the movements, whose joints localization and actuation are so modulated to display the tensions and the emotions of music experience. The joints movements and the movement configuration states result to obtain a self-assigned association with the emotional values of the response of subject to music represented as variables according to the dimensions of arousal and valence. In this case the entropy values are collapsed to the dimension of entropy valence. The resulting tensile and affective value of the movement configurations can be thus summarized: B1 is a motor behaviour displaying high arousal and valence; B2 is a motor behaviour displaying high arousal and low valence; B3 is a motor behaviour displaying low arousal and high valence; B4 is a motor behaviour displaying low arousal and low valence as shown in Fig. 2.
Movement evaluation scheme and the classification pilot experiment

We hypothesize that the movements of the Geminoid display the subject tensions and emotions that characterize music experience. The factors that produce them at each module of the architecture can be manipulated across experimental conditions in order to test either the contribution of a particular type of conceptual and affective constituent of the cognitive capacity for music or their correlation. Furthermore, models of movement configurations can be predicted and fitted to or inversely built from experimental data to simulate the correspondence between the derivation of preferential rules, the affective attribution, the perception of tension according to the fulfillment or the violation of the expectancy of the regularities on which the understanding of music is based. All the relevant factors can be operationalised. The tension and relaxation values, which can be used as descriptor of what subjects impart as motor commands equivalent to how they experience whether their expectancy is fulfilled or not, may be qualified in advance during the training phase. The qualification may regard for instance: the degree of (high, medium, low), the direction (up, down) and the spatial reference (bottom, above, left, right) of tension or features like: fast vs. slow, sudden vs. gradual, smooth vs. resistant; compressed vs. diluted. These qualifications may be explicitly provided to subject as characterizing the conditions of tasks they are required to perform while listening to music with the aim of manipulating the motor commands, which are to be sent to the robot. The conditions may vary, for instance, across listening passively to music and imaging to be moved or to move ones body or a thing according to or in contrast with the tension experienced while hearing music. The BCI signals can be entered as independent variables in the training and calibration stages before the experimental manipulation and the movements of the Geminoid can be taken as dependent variables. The number of the configuration states of the potential movements of the robot may be increased allowing the joints to move according to the following parameters: the number of displacement of a body part in the time unit; the number of movements allowed by a particular reference system centred on the robot body in the time unit; the position held by a body part repeated in multiples of the time unit that delimit the time window of interest. The properties of movement can be varied too: for instance, their width, average or initial speed, direction. These descriptors can be associated with features that distinguish kind of musical stimuli like tempo, dynamics, articulation, attack. For the time being we have conducted a pilot experiment restricted to the emotion classification of the affective state correlated on average with different genres of music. Twelve subjects were asked to impart motor commands to the Geminoid that would let it display the emotions perceived or felt while they were hearing randomly selected instances of electronic, rock, pop and chill out music. The parameters values of energy and entropy were used as input for a SVM trained over DEAP, a database built by Koelstra et al. (2012) for emotional analysis with reference to a six items list of emotional states. The average distribution of the emotional states perceived or felt for each kind of music was computed (see Fig. 3 and Table 1). Finally, the changes regarding the affective constituent of the engagement of subjects for each genre of the musical inputs have been plotted Fig. 3. The changes correlated with the music experience as a function of the particular music genres takes the values along the arousal and the valence axes, which are predicted by Russell (1980) model and are consistent with the combination of arousal and valence that characterize the motor behaviours of the Geminoid given the actual four states movement configurations of the robot (see Fig. 4).

Conclusion

We described the design of a BCI robotic augmented architecture to build and test models of the cognitive capacity for music. The neurobiological evidence on music specific theories suggests that this capacity is biologically founded, although it is still not clear in which way it is hard-wired in the brain or whether it has emerged as a network of neural circuits that however provide this capacity with domain specificity. On the grounds of the models of the computational theories and of the results of the experimental research in psychology and cognitive sciences, we presented such capacity as consisting of conceptual and affective constituents that work in parallel to establish the understanding of music. We introduced the construct of tensions as the descriptor of the relations that are common to structures embedded in the linear stream of sounds, governed by the preferential rules of the musical grammar, and to the recognition and attribution of emotions as a function of the satisfaction or the violation of the expectancy of the sound changes that is founded on those rules. In the light of the interpretation of music experience as cognitive frame, we have justified the view that the constituents of the cognitive capacity are likely to be domain specific, in the sense that the specificity derives from the specialization of cognitive functions ordinarily serving in everyday life.
Table 1
Percentage of emotions recognized for each category of musical stimuli inputs.

<table>
<thead>
<tr>
<th></th>
<th>Excitement</th>
<th>Pleasure</th>
<th>Tension</th>
<th>Anger</th>
<th>Sadness</th>
<th>Relax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic</td>
<td>30.56%</td>
<td>19.44%</td>
<td>9.72%</td>
<td>18.06%</td>
<td>8.33%</td>
<td>13.89%</td>
</tr>
<tr>
<td>Rock</td>
<td>16.67%</td>
<td>23.08%</td>
<td>12.82%</td>
<td>12.82%</td>
<td>15.38%</td>
<td>19.23%</td>
</tr>
<tr>
<td>Pop</td>
<td>27.78%</td>
<td>9.72%</td>
<td>4.17%</td>
<td>4.17%</td>
<td>8.33%</td>
<td>45.83%</td>
</tr>
<tr>
<td>Chill out</td>
<td>12.82%</td>
<td>15.38%</td>
<td>7.69%</td>
<td>11.54%</td>
<td>37.18%</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Russel model for perceived type of music.

and the robust conceptual-affective correlation reported in the literature, notwithstanding the lack of a one-to-one mapping. The music experience provides the cognitive frame in which configurations of sound musical properties, which satisfy the preferential rules of the grammar of music cognition, are systematically coupled with affects.

We discussed the principal design of the modules of the proposed architecture and their functional connection. The architecture allows controlling and manipulating the musical inputs fed as stimuli to subjects who experience music, recording the physiological correlates of the conceptual and the affective constituents of the capacity for music, testing the correlation between sound features, rules of combination, affective attribution and tensions by means of the observation and measures taken on the movements imparted as BCI-mediated motor commands to the robot. As a preliminary result we have presented a pilot experiment on the classification of the energy dependent affective change perceived or felt by subjects in connection with different music genres and its correlation with the configurations of the motor behaviour of the robot.

The future research will be devoted to extend the potentiality of the architecture modules to yield experimental conditions to study the constituents of the cognitive capacity for music as well as their interaction. The connection between the BCI modules and the robot will be employed as a test bed for interdisciplinary models of music cognition, which preserves the complexity of the constructs and the concepts developed across the disciplines and at the same time requires their translation into implementation and system design terms.

Acknowledgment

Authors wish to thank former master students Dario Di Figlia, Ivan Ficarra, Fabio Galiano and Giuseppe Grassadonia for their support in the development of the architecture.

References


Lerdahl, F. (2001). Tonal pitch space. *Oxford University Press*


Massachusetts: The Massachusetts Institute of Technology Press.


