A Natural User Interface for Gestural Expression and Emotional Elicitation to access the Musical Intangible Cultural Heritage

Christina Volioti, Sotiris Manitsaris, Edgar Hemery, Vasileios Charisis, Leontios Hadjileontiadis, Stelios Hadjidimitriou, Eleni Katsouli, Fabien Moutarde, Athanasios Manitsaris

A Natural User Interface for Gestural Expression and Emotional Elicitation to access the Musical Intangible Cultural Heritage

CHRISTINA VOLIOTI[1], University of Macedonia
SOTIRIS MANITSARIS[2], MINES ParisTech
EDGAR HEMERY[2], MINES ParisTech
STELIOS HADJIDIMITRIOU[3], Aristotle University of Thessaloniki
VASILEIOS CHARISIS[3], Aristotle University of Thessaloniki
LEONTIOS HADJILEONTIADIS[3,4], Aristotle University of Thessaloniki
ELENI KATSOULI[1], University of Macedonia
FABIEN MOUTARDE[2], MINES ParisTech
ATHANASIOS MANITSARIS[1], University of Macedonia

This paper describes a prototype natural user interface, named the Intangible Musical Instrument, which aims to facilitate access to the knowledge of the performers that constitutes musical Intangible Cultural Heritage, using off-the-shelf motion capturing that is easily accessed by the public at large. This prototype is able to capture, model and recognize musical gestures (upper body including fingers) as well as to sonify them. The emotional status of the performer affects the sound parameters at the synthesis level. Intangible Musical Instrument is able to support both learning and performing/composing by providing to the user not only intuitive gesture control but also a unique user experience. In addition, the first evaluation of the Intangible Musical Instrument is presented, in which all the functionalities of the system are assessed. Overall, the results with respect to this evaluation were very promising.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning – Knowledge acquisition; H.m [**Information Interfaces and presentation (e.g., HCI)**]: Miscallaneous.

General Terms: Intangible Cultural Heritage, Musical gestures, Expert knowledge, Natural User Interface

Additional Key Words and Phrases: Gesture Recognition, Emotional status, Sonification, Evaluation

1. INTRODUCTION

Cultural expression is not limited to architecture, monuments or collections of artifacts. It also includes fragile intangible live expressions, which involve knowledge and skills. Such expressions include music, human skills, etc.. These manifestations of human intelligence and creativeness constitute the Intangible Cultural Heritage (ICH)[1]. ICH is at the same time traditional, contemporary

This work is supported by the Widget Corporation Grant #312-001.
Author's address: C.Volioti, E-mail: christina.volioti@uom.edu.gr. S.Manitsaris, E-mail: sotiris.manitsaris@mines-paristech.fr. E.Hemery, E-mail: edgar.hemery@mines-paristech.fr. S.Hadjidimitriou, E-mail: stelios@psyche.ee.auth.gr. V.Charisis, E-mail: vcharisis@ee.auth.gr. L.Hadjileontiadis, E-mail: leontios@auth.gr. E.Katsouli, E-mail: katsouli@uom.edu.gr. F.Moutarde, E-mail: fabien.moutarde@mines-paristech.fr. A.Manitsaris, E-mail: amanitsaris@uom.edu.gr. [1]Multimedia, Security and Networking Lab, Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece. [2]Centre for Robotics, MINES ParisTech, PSL Research University, Paris, France. [3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece. [4]Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE.

https://doi.org/10.1145/3127324
[1] http://i-treasures.eu/

and living, because it does not only refer to inherited knowledge but also to the renewal of contemporary cultural expressions. It refers to the past, to the present, and, certainly to the future and is the mainspring of humanity's cultural diversity. According to UNESCO, music is the most universal form in the performing arts, since it can be found in every society, usually as an integral part of other performing art forms and other domains of the ICH. Music of different types such as classical, contemporary or popular, sacred etc., can be found in a large variety of contexts. Instruments, artefacts and objects, in general, are closely linked with musical expressions and they are all included in the Convention's definition of the ICH [UNESCO 2003]. Music that fits with the Western form of musical notation is better protected, while those that do not fit are usually threatened with disappearance when their holders die. Thus, the crucial point for all music forms is to develop the motor skills of playing a musical instrument by strengthening the bond between the expert performer and the learner. Motivated by this need, in recent years, researchers focused on the study of embodiment and enactive concepts. These concepts reflect the contribution of body movement to the action/perception and the mind/environment interaction [Noë 2004]. In the performing arts, and, more precisely, in music, body movement is semantically connected with gesture in most activities, such as performing and composing.

Composers bring together knowledge and skills in sound coloring and organization, in terms of structure and form. These skills are depicted on the music score of their pieces, which constitute the Tangible Cultural Heritage (TCH). Nevertheless, the music score usually contains only a few abstract annotations about idiomatic gestures that should be incorporated by the performer during his/her musical and physical playing. Such information leads to the organization of the musical material, which is culminated in a compositional structure. The analysis of the musical material always brings to the surface the question "how does this work?". Music theory explains the musical structure and/or defines the way the material functions, according to various viewpoints, such as those of Allen Forte, Arnold Whittall, Rosemary Killiam and Patrick McCreless [Hadjileontiadis 2014]. Therefore, music theory can explain how a piece of music functions, but it does not provide information about the method and more precisely, the way the musician should interpret the musical score or/and how to perform.

The performance is the result of the symbiosis between the musician and his/her instrument. This symbiosis takes the form of an interactional relationship, where the musician is both a trigger and a transmitter, connecting the *perception* (mediated instrumental mechanisms and physical environment), the *knowledge* (theoretical understanding of the inherited music score) and the *gesture* (semantic motor skills). Consequently, the expert musical gesture can be considered as a fully embodied notion that encapsulates the motor skills of the performer to interpret musical pieces, following the musical notation defined by the composer. Moreover, the musical instrument is a physical interface that can be considered as a means of musical expression and performance. Nevertheless, the learning curve of playing musical instruments requires years of training, practice, and apprenticeship before being able to perform. Furthermore, the learning of expert musical gestures is still viewed as a communicative act of social interaction, rather than "my own" personal experience. Consequently, "learning" musical gestures and "performing" music are usually perceived as separate concepts and experiences. This means that accessing knowledge is a long-term procedure, since there is no quick transition from novice to expert.

Based on the above need, the purpose of this paper is to present a Natural User Interface (NUI) named the "Intangible Musical Instrument" (IMI) for capturing, modeling and recognition of musical gestures of expert performers which will be able to support both "learning" and "performing/ composing" as a unified user experience.

2. STATE OF THE ART

2.1 Musical gestures as referential patterns of composers

Gesture is the core activity of music creation; a dynamic organism, similar to the human organism; an experience that combines structural properties of music together with cultural and historical contexts [Truslit 1938; Coker 1972; Broeckx 1981; Hatten 1994; Cadoz and Wanderley 2000; Cumming 2000]. In talking about musical gestures and cultural heritage, there is an endless list of composers and knowledge that constitute an ICH. For example, short musical patterns, which can easily be imitated through body gestures, constitute, for Beethoven, the palette of his compositions. These short patterns, and their variations, constitute an ongoing unfolding process throughout his musical pieces. Many analysts consider this practice as a self-referential context where musical gestures, similar to other variations of the same gesture, are recognized within the same piece. Another example, which can be given is the musical collage of gestures in the Sinfonia of Berio. Gestural patterns of Mahler, Ravel and Debussy are integrated into the new musical piece so that in the Sinfonia they remain as representative musical idioms that transmit music-related cultural meanings [Godøy and Leman 2009].

Consequently, "musical patterns" and "gestural patterns in music" are closely linked notions, since sonic forms are understood through embodiment. These patterns constitute elements of social interaction and differentiation since their imitation entails the acquisition of cultural models for emulation. According to McNeill [1992], these patterns can be considered throughout history as playing an important role in creating and sustaining human communities and can be understood as a mirror system between composer and listener or even master and learner [Clayton 2000; Keller 2008]. Nevertheless, musical pieces documented through musical scores, which constitute a TCH, encapsulate only abstract information about energy and expressivity of gestures, which are finally incarnated through the interpretation of performers on musical instruments.

2.1.1 Typology of musical gestures

The gesture vocabulary described by Delalande, which has been extensively used in the literature [Delalande 1988; Cadoz and Wanderley 2000; Zhao 2001], divides musical gestures into three classes. The first class is named "*effective gesture*" and it concerns movements that are necessary to mechanically produce the sound (e.g. press a key). The second class is named "*accompanist gesture*" and it refers to sound-facilitating movements (e.g. specific postures that permit expressivity). Finally, the "*figurative gesture*" conveys symbolic messages to the audience as a communication act.

2.1.2 Transmission of gestural know-how in music

The examples documented in the previous two sections show that the musical meaning of gestural know-how involves different levels of information, which are: a) first-person, b) second-person and c) third-person perspectives on gesture [Leman 2010].

The *first-person perspective* on gesture defines the meaning of the gesture for the person that actually implements it. Within the ICH context, the expert performers are holders of ICH that have perfected their know-how to include high-level specific characteristics. Additionally, the learner can also have a first-person perspective when playing a musical instrument. The difference between the two is that the expert has developed, at a greater level than the learner (it really depends on the level of the learner), his/her action-based approach to gesture, because s/he knows all the gestural patterns in music. S/he has mental access to how the action, described on the musical score, is deployed over time and s/he has the capacity to control his/her sensorimotor system that produces the corresponding sonic form.

The *second-person perspective* on gesture refers to how other people perceive the musical gesture in a social interaction context. This approach is the most typical one that is used in music schools, conservatories, etc.. The learner observes the experts, which in most cases are his/her teacher,

following the concept of "my" perception of "your" gesture [Leman 2010]. According to this "me-to-you" relationship, a mirroring system is established between expert and learner, where the body movements of the learner are deployed, so that the movement of the expert, incorporating the knowledge of the composer derived from the musical score, is understood as an action by the learner.

The *third-person perspective* on gesture focuses on the measurement and capturing of moving objects. This task can be done by a computer using audio recording, video recording, motion capture technologies and brain scans, as well as physiological body changes [Pratt 1931/1968; Friberg and Sundberg 1999; Camurri et al. 2005]. In this way, the knowledge of the performer is captured, based on techniques of feature extraction and pattern matching.

2.2 Emotional expression in music

There is no need for scientific evidence to support the fact that music expresses emotions, as personal accounts of affective experiences during listening to music are more than sufficient. However, a vast amount of research has been conducted in order to reveal further insights into this phenomenon, ranging from philosophical to biological approaches [Juslin and Sloboda 2010]. It has been suggested that such music-induced emotions are governed by universality in terms of musical culture, meaning etc. and that listeners with different cultural backgrounds can infer emotions in culture-specific music to a certain extent. Such evidence has led to the assumption that neurobiological functions underlying such emotional experiences do not differ across members of different cultures, as the responsible neural networks may be fixed. In general, the processing of musical stimuli involves the gradual analysis of music structural elements from basic acoustic features to musical syntax that leads to the perception of emotions and semantic meanings underlying the stimuli [Koelsch and Siebel 2005]. It is becoming evident that the structure of music defines what it expresses. To be more accurate, music does not literally express emotion, but it is its structural elements and production performance shaping the acoustic outcome that foster the induction of emotional states in the listener. In the following descriptions, affective states will often be characterized based on the *valence-arousal model* [Russell 1980]. Valence denotes whether an emotion is positive or negative, while arousal refers to the level of excitation that the emotion encapsulates.

2.2.1 Emotions in musical performance

Written music can be performed in different ways just as a piece of text can be read with various tones. In an important sense, it can be argued that music and performances of the same work can differ significantly. The latter form the concept of performance expression that refers to both a) the correlation between the performer's interpretation of a musical excerpt and the small-scale variations in timing, dynamics, vibrato, and articulation that shape the microstructure of the performance and b) the relationship between such variations and the listener's perception of the performance. It has been proposed that performance expression emerges from five different sources, i.e. Generative rules, Emotional expression, Random fluctuations, Motion principles, and Stylistic unexpectedness, referred to as the GERMS model [Juslin 2003]. Here, the focus is placed on emotional expression that allows the performer to convey emotions to listeners by manipulating features such as tempo and loudness in order to render the performance with the emotional characteristics that seem suitable for the particular musical piece. Table I reports the primary acoustic cues of emotional expression in music performance [Gabrielsson and Lindström 2010; Juslin and Timmers 2010]; these are mainly empirical relationships, rather than absolute, and constitute an appealing research topic.

Table I. Empirical relationships between sound parameters and emotions [Gabrielsson and Lindström 2010; Juslin and Timmers 2010]

PARAMETERS	DEFINITION	ASSOCIATED EMOTIONS
Tempo	The speed or pace of a musical piece	Fast tempo: happiness, excitement, anger Slow tempo: sadness, serenity

Mode	The type of scale in which the piece is written	Major tonality: happiness, joy Minor tonality: sadness
Loudness/ Volume	The physical strength and amplitude of a sound	Loud sound: happiness or power, anger Soft sound: relaxation, tenderness or sadness
Melody	The linear succession of musical tones that the listener perceives as a single entity	Complementing harmonies: happiness, relaxation Clashing harmonies: excitement, anger
Tonality	Musical key or the relations between the notes of a scale or key of a musical piece.	Tonal: joyful, dull Atonal: angry
Rhythm	The regularly recurring pattern or beat of a song	Smooth/consistent rhythm: happiness, peace Rough/irregular rhythm: amusement, uneasiness

It is clearly conceivable that emotions play a significant role in musical artistic expression. Consequently, the analysis and manipulation of users' affective states should be taken into serious consideration within Intangible Musical Instrument (IMI) design, development and practice that aims to support music performance.

2.3 Gesture control of sound

In order for a musical interface, or instrument, which draws gestural data from sensors and cameras, to feel natural from the point of view of user experience, it should provide intuitive gesture control of sound. With the term "mapping gesture to sound" or "gesture sonification" is meant the procedure, in which the gestural data is being associated with the sound parameters; therefore, the gesture characteristics and features, as well as the sound synthesis variables that are going to be used, have to be defined. Then, a decision about the strategy of mapping, explicit or implicit mapping, also has to be made. In *explicit* mapping, also called direct mapping, the input is directly associated to the output while, *implicit* or indirect mapping refers mostly to the use of machine learning techniques, which imply a training phase to set parameters [Bevilacqua et al. 2011].

2.4 Conclusions from the state-of-the-art and motivation

Leveraging the above, there is a growing interest in the analysis of the gestural knowledge. A large amount of studies conducted in the last years on embodied music cognition have investigate that not only effective, but also accompanist and figurative gestures are very important, since they are related to the expressivity and to the social interaction of the performer with the audience [Jensenius 2007; Maes et al. 2010]. However, "learning" musical gestures and "performing" music are usually perceived as separate concepts and experiences that pass through intermediate physical mechanisms. Usually, for learners, the "challenges" related to the physical aspects, such as the instrument being more important than the "skills" needed in music playing, can cause frustration. Consequently, the achievement of good motor skills is a long-term procedure. Additionally, the learning of the motor skills that are self-referential for a specific performer still constitute a "black box" for the learner, since it can be only approached as a second-person experience; therefore, when the learner observes the expert, s/he perceives as expert motor skills, the limited abstract sonic movements, which are visually derived from the expert gestures. Finally, the skills required of the performer, whether gestural or emotional, are not documented on the music score in much detail. Hence, in cases where a musical piece does not follow the Western form of music notation, it is extremely difficult to transmit it to the next generations.

Motivated by the above, the main objective of the present work is to create a natural user interface of the gestural expression and emotion elicitation in music. This natural user interface refers to Intangible Musical Instrument (IMI), which provides a holistic approach to gesture capturing, recognition and sonification, taking into consideration the emotional status of the performer at the synthesis level. Moreover, IMI can support learning, performing and composing with gestures as a first-person experience, by putting the user at the core of musical activities, such as performing and composing with gestures.

3. METHODOLOGICAL APPROACH

3.1 Methodology of gestures and emotions sonification

IMI supports the continuous and real-time gesture control of sound, taking into consideration the emotional status of the performer. Therefore, the fundamental elements of the proposed methodology (Figure 1) are the modalities that are involved in the musical performance, which are the gestures, the emotions and the sound. The scientific challenge of this methodology is to propose a coherent way of interconnecting these modalities, and thus answering "what to map, where and how?".

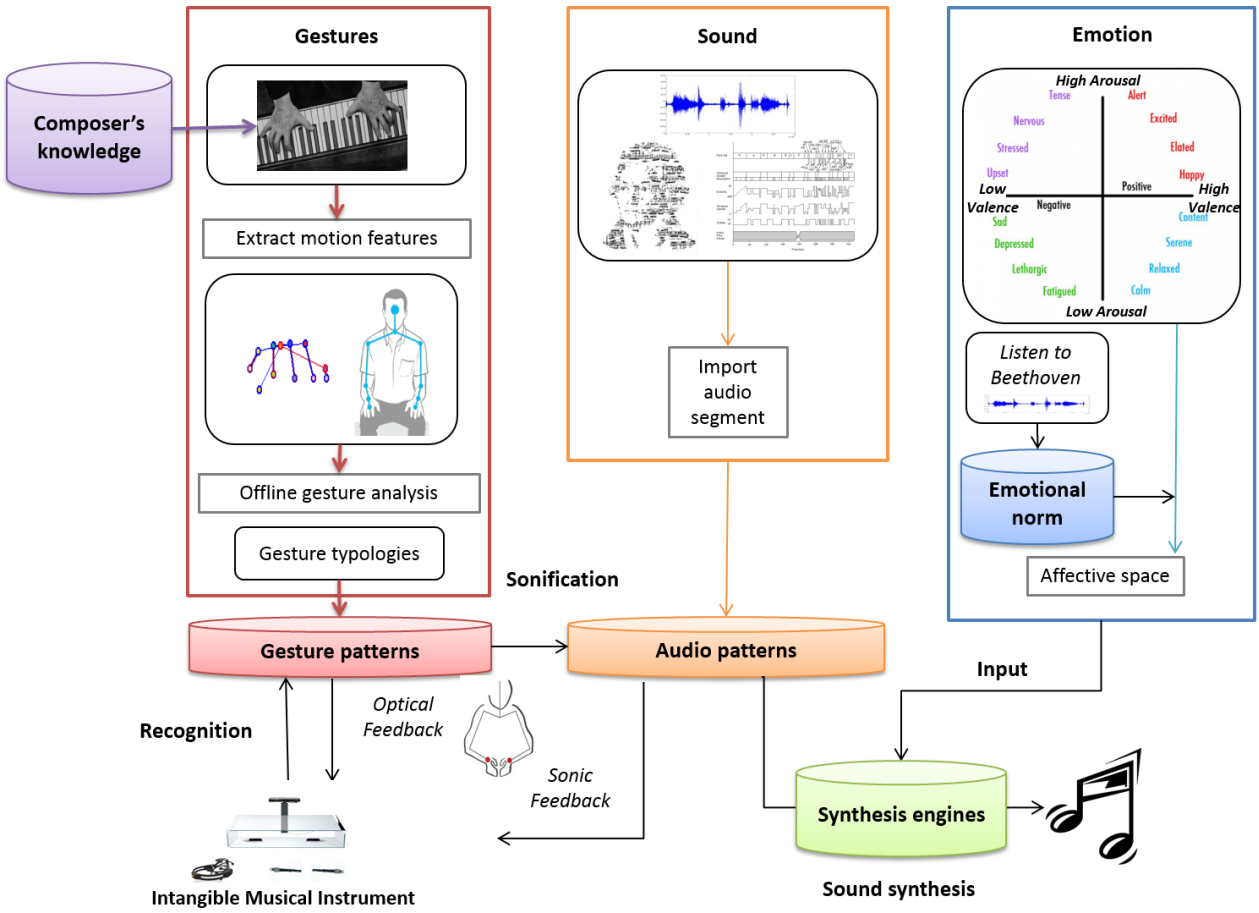

Fig. 1. Methodology of gesture sonification taking into consideration the emotional status of the performer

The motor skills of the user are put at the core of the sound creation, by using the gestural modality as a trigger of sound processes. These motor skills are captured by motion capture sensors, which are exactly the same for all the user profiles (e.g. expert, learner) for both learning and performing/composing. Taking into consideration the need for modeling the expert gestures, robust gestural information is captured in order to apply on it machine learning and pattern recognition methods. All the movements of the upper body part, including finger and head motions, are captured and represented through measurable physical descriptors; therefore, the physiological analysis of gestures focuses on an analytical description based on cinematic, spatial and frequential characteristics. More precisely, a hybrid rotational and Cartesian representation of the motion is applied, using inertial sensors and depth cameras. The descriptors of the expert gestures are used to

create deterministic models (on a frame-by-frame basis) as well as to train stochastic models based on time series. In a learning context, when the learner performs a gesture, his/her gestural descriptors are compared online with the expert models of all the gestures of the vocabulary and a gesture is recognized according to which model outputs the highest probability. In a musical performance context, the performer can train the models with his/her own gestural data and recognize them online. The gesture recognition engine that uses time series as input is based on a hybrid approach of Hidden Markov Models (HMMs) and Dynamic Time Warping (DTW) [Bevilacqua et al. 2007], where HMMs are used to recognize the gesture and DTW to temporarily align the modeled gesture with the input gesture.

As long as the gesture is recognized, different mapping strategies are proposed. The first strategy refers to the connection of gesture perceptual parameters to some set of sound perceptual parameters, which are translated into concepts that can be perceived visually (gestures) or sonically (sounds). This strategy is also known as explicit mapping and it is used for associating fingerings to pitches as a one-to-one relationship. Similarly, with bijective functions, there is a mapping between 3D positions of fingertips to the creation of specific notes. This function takes gesture as input and outputs sounds with a MIDI piano synthesizer. When the fingertips come into contact with the surface of the IMI or hover less than a centimeter from it, a note is produced, the sound of which is determined by a set of parameters such as speed and the fingers' trajectory before contact. Moreover, the musically interactive surface is articulated in three zones, similar to the octaves of acoustic pianos, and each of them is associated with the hand's centroid.

The second mapping strategy, called implicit mapping, is based on a temporal mapping method [Bevilacqua et al. 2011]. The basic advantage of this approach is the time warping of the sound that is produced, depending on the speed of the performed gesture in real-time. It replays sound samples at various speeds, according to the gesture performed in real-time. Audio time stretching and compressing, as well as re-synthesis of audio can be accomplished by using the granular sound synthesis engine. In particular, the temporal mapping method associates a sound with a template gesture and links temporal states of a sound with the temporal states of the template gesture. Implicit mapping is based on information that is given from head, arms and the vertebral axis, meaning the upper body, without including the fingers.

Finally, music is well-known for affecting human emotional status, but the relationship between specific musical parameters and emotional responses is still not clear. Taking into account Table I, the sound parameters that are proposed in this research and are directly associated to the emotional status (Valence-Arousal model) are the loudness and the pitch. More specifically, the values of Valence modify the pitch of the sound, while the values of Arousal change the loudness. In both learning and performing/composing contexts, the Valence and Arousal parameters of the user are used as input to the sound synthesis engine, thus, mostly affecting the intensity and the timbre of the sound.

3.1.1 Learning the expert musical gestures

As has already been mentioned, the key point for all music forms is to have access to the gestural knowledge of playing a musical instrument and the strengthening of the bond between the expert holder of the ICH (which is the composer or performer) and the learner. As a result, in a learning scenario, the learner performs pre-defined expert gestures, taken from the vocabulary. Therefore, s/he imitates these expert gestures. S/he attempts to get close enough to the expert gesture model, so that the sound can be re-synthesized at its original speed. The re-synthesis of the sound is based on the granular sound synthesis engine.

3.1.2 Composing with gestures

In a composing scenario, the composer has the ability to create his/her own vocabulary of musical gestures, to describe the expressiveness by defining the appropriate emotions that the performer

should imitate and to sonify his/her gestures and emotions by defining the sonic spaces and parameters. As a result, the composer is able to experiment with his/her own gesture-sound mappings and audio synthesis, as well as to compose contemporary music by performing gestures one after the other, by using fingers, body gestures and emotions. The goal of the composing scenario is to provide a generic system, which can be adapted according to the needs of each performer and composer. The variety of sounds the IMI can produce is equivalent to most synthesizers, but the way the musician interacts with it is totally unique, making the interface a powerful tool for both performing and composing music. However, it is important to highlight that the IMI is not a virtual replacement for the piano (or any other keyboard instrument), but an adaptation of the existing techniques for this instrument to computer music, including electronic and electroacoustic music.

4. TECHNICAL IMPLEMENTATION AND SOFTWARE DEVELOPMENT

The aforementioned methodology, which refers to capturing, analyzing, recognizing data, mapping gesture to sound, as well as sound synthesis, is implemented with Max/MSP programming language[2]. The setup prototype is a construction made of Plexiglas, shaped so as to look like a table on which user can put his/her hands (Figure 2). The dimensions of the table are 70 cm long, 40 cm wide and 13 cm high. The setup lies on a table so that the hands on the Plexiglas are placed at a comfortable height.

Fig. 2. Intangible Musical Instrument (IMI)

The successive steps, which were achieved in order to first capture the gesture and then to model it, are described below. For the capturing part, two types of depth camera and two inertial sensors are used. The first type of depth camera is the Kinect[3], originally created for video gaming purposes. Equipped with a structured light projector, it can track the movement of the whole body of individuals in 3D using a Random Decision Forest algorithm [Shotton et al. 2013]. However, the proposed methodology focuses on the upper part of the body and the current algorithm delivers a fairly accurate tracking of the head, shoulders, elbows and the hands, but not the fingers. The second type of camera used is the Leap Motion[4], which works with two monochromatic cameras and three infrared LEDs. The Leap Motion provides an accurate description of the hand skeleton, with more than 20 joints positions and velocities, both in 3D (x, y, z coordinates). Two Leap Motion are used, one for each hand. Each Leap Motion has a field of view of 150° and tracks the hand from below efficiently up to 30 cm above the camera center (the camera is oriented upwards). Once placed on their slots on the IMI, they cover the whole surface of the table and a volume above it. Additionally, two inertial sensors are taped to the user's wrists (Animazoo motion capture suit[5]), which deliver rotation angles (Euler angles).

[2] https://www.cycling74.com/
[3] https://www.microsoft.com/en-us/kinectforwindows/
[4] https://www.leapmotion.com/
[5] http://synertial.com/

Finally, an electroencephalogram is mounted on the head to record brain electrical patterns via the Emotiv sensor[6]. These patterns are then translated to the emotional status of the user.

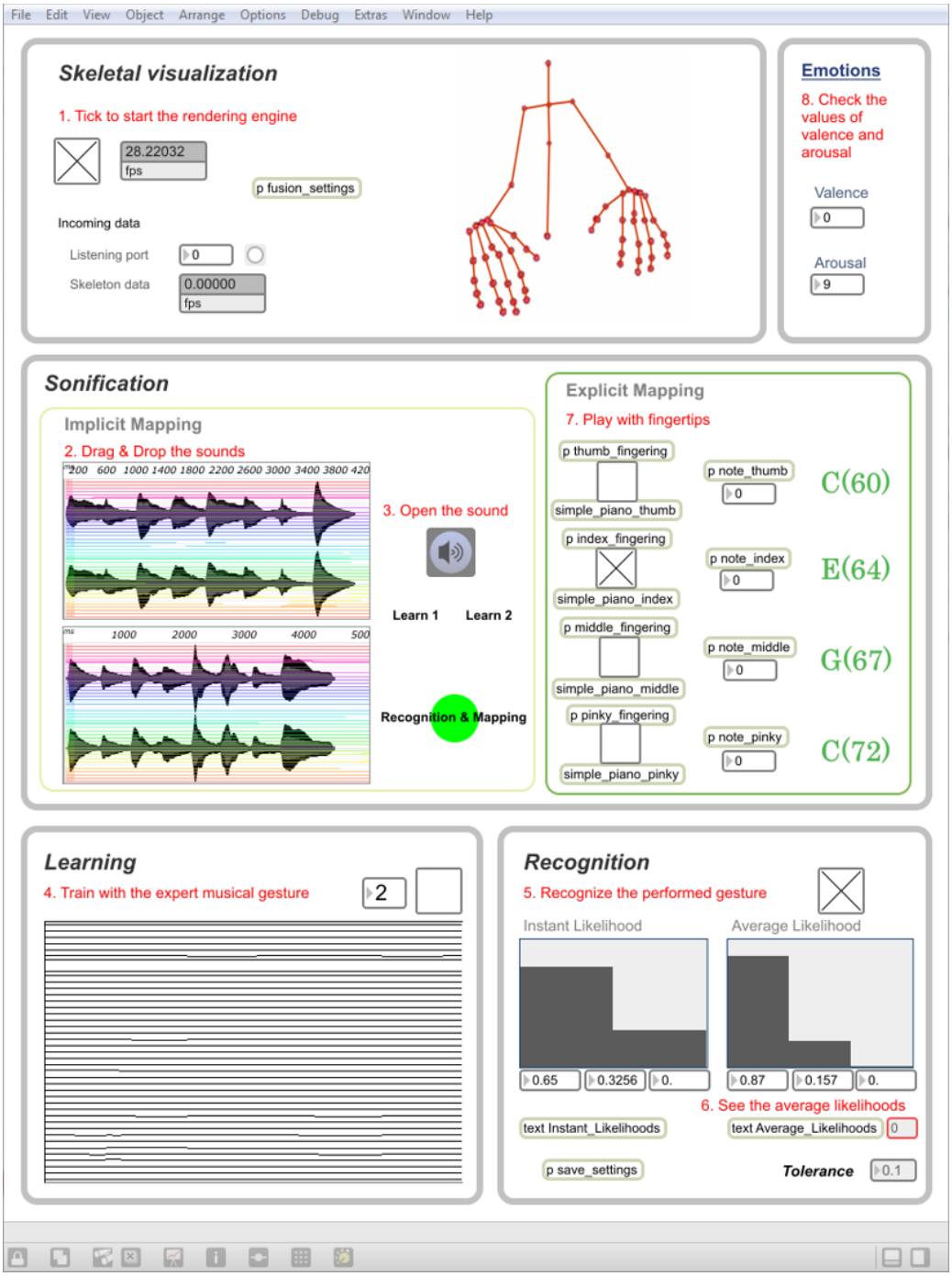

Fig. 3. Unified interface for gesture and emotion recognition and sonification

[6] https://emotiv.com/

In a live learning scenario of the proposed methodology, which is a "learning by doing" approach, the learner has to stand in front of the capture system (Kinect and Leap Motions) and wear the two inertial sensors on his/her wrists and the Emotiv sensor on his/her head, so as to see his/her skeletal representation on the IMI's interface (Skeletal visualization in Figure 3) and attempt to perform the musical expert gestures as well as to embrace the respective emotional status. The expert vocabulary contains some basic musical gestures such as ascending and descending scales, ascending and descending arpeggios, as well as some basic musical excerpts of Beethoven.

In order to perform i.e. ascending and descending scales, the learner's gestural data (positions and Euler angles) are analyzed and used for the machine-learning phase, which is based on HMMs and DTW technique [Bevilacqua et al. 2007; Bevilacqua et al. 2010]. The main advantage of this hybrid approach, instead of using other algorithms, is that it permits a time alignment between the model and the data used as input for the recognition. The two phases are Training (or Learning) and Recognition. In the Training Phase, the expert trains the system with his/her musical gesture, and a pre-recorded sound is associated with the template gesture and links the sound with temporal states of the template gesture (Learning in Figure 3). In the Recognition Phase, the learner tries to imitate in real-time the expert's musical gesture. The meaning of real-time performance and recognition is that the technique does not recognize the gesture once it is completed, but it estimates the gesture in real-time, moment by moment over time. As a result, it is designed to continuously output information about the gesture, by providing the learner's probabilistic estimations (Recognition in Figure 3). Simultaneously, the sonification is taking place based on granular sound synthesis engine, in which the system predicts the sound according to the performed gesture (Implicit Mapping in Figure 3). Moreover, the sound can be modified according to the values of valence and arousal, meaning the emotional status of the user (Emotions in Figure 3). The values of arousal modify the loudness and the values of valence the pitch of the sound. Finally, the learner has the ability to play a musical sequence (i.e. ascending and descending arpeggios), in which each fingertip is associated to each specific note (Explicit Mapping in Figure 3).

5. EVALUATION

This section deals with the evaluation of the IMI and its functionalities in terms of complying with the user's requirements, expectations and experiences. The survey instrument is a structured questionnaire that has been distributed during three scheduled demos in Greece, in which 105 respondents took part. The demographics of the respondents are presented in Table II.

Table II. Demographics of respondents (n=105)

	Number	Percentage (%)
Sex:		
Male	31	29.5
Female	74	70.5
Age:		
Up to 20	21	20.0
21 – 30	36	34.3
31 – 40	20	19.0
41 - 50	28	26.7
Perceived familiarity in using computers:		
Not much	8	7.6
Adequate	66	62.9
Good	31	29.5
Music Literacy:		
Not at all	50	47.6
Not much	39	37.1
Adequate	12	11.4

Good	4	3.8
Familiarity with classical music:		
Not at all	48	45.7
Not much	28	26.7
Adequate	21	20.0
Good	8	7.6

A common rule for considering whether a sample size is acceptable is the ratio of sample size to the number of the latent variables parameters to be equal to 5 to 1 [Bentler and Chou 1987]. Taking into consideration that a sample size that follows this rule is equal to at least 90, it is concluded that the sample size of this study is acceptable. Figure 4 shows the workshops in progress, in which users experiment with IMI while learning and performing musical gestures.

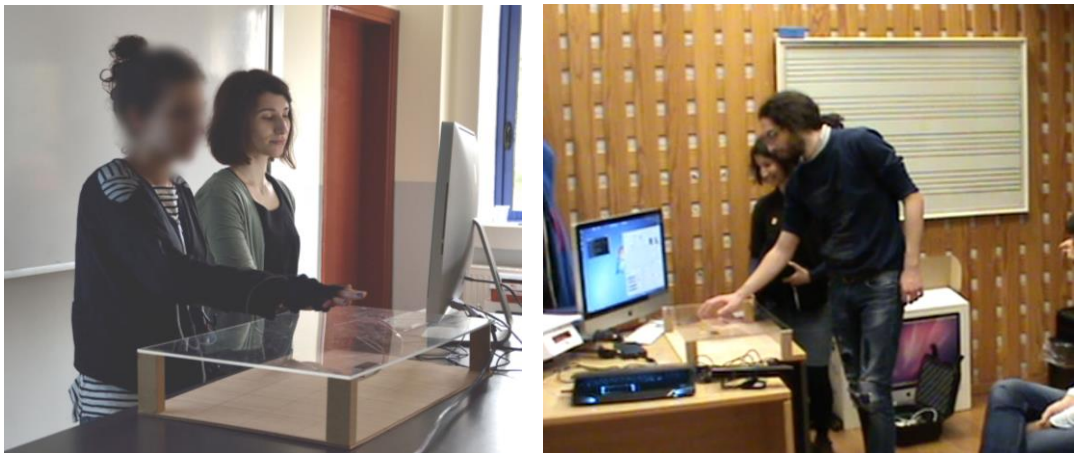

Fig. 4. IMI workshops where the users experiment with the IMI in both learning and performing/composing contexts

During the workshops, the researchers presented the IMI to the participants and a number of expert musicians performed the musical expert gestures that are described in Section 4. Each of those experts used his/her personal musical style to perform on the IMI. Then, participants (users) were asked to identify basic referential elements, meaning musical gestures of the experts and each participant tried to imitate the gestures of his/her favorite expert on the IMI in order to control the expert sound. Figure 5 presents the operational model that was used. Operational model is an abstract and visual representation of how an activity is working, or in other words, it is the blueprint of this activity. Each "box" refers to a general construct, which is constituted by items (questions). Constructs were proposed by expert musicians, teachers and engineers, thus verifying content validity. Special attention was given to basic users, without any specific knowledge of music, in order to verify whether the IMI facilitates the learning of the expert gestures.

The key concept of this operational model argues that the skills of the user on recognizing referential stylistic elements or even specific movement patterns of a given expert musician, mediate the relationship between the quality of interaction with gesture sonification and the performance of the user when playing on the IMI. Based on this operational model the specific goals of this evaluation may be summarized as follows:

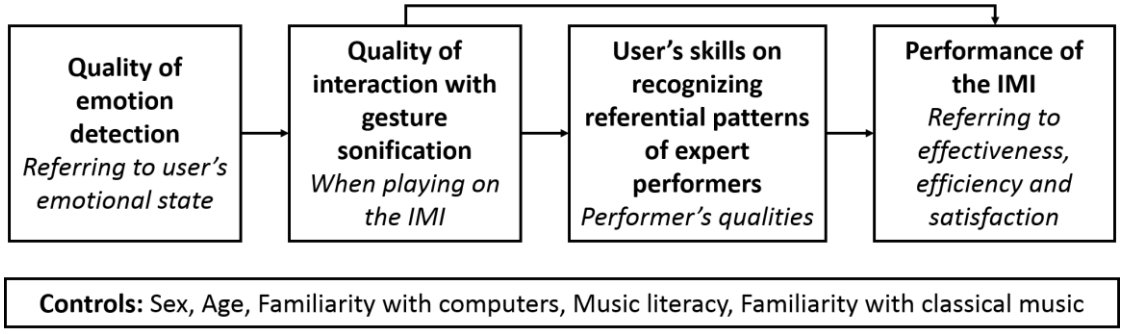

Fig. 5. The IMI operational model indicating the relationship between quality of emotion detection and IMI performance

G1. To evaluate the overall perceived performance of IMI, with respect to effectiveness (i.e. if IMI meets its objectives), efficiency (i.e. if IMI responses satisfactorily and in a short time in gestures, emotions and sound production), and satisfaction (i.e. if IMI provides satisfaction to the user).

G2. To evaluate whether the personal style (in terms of gesture, music style, and emotions) of a performer (while s/he interprets classical or contemporary composers) can be recognized.

G3. To evaluate the usability and the user-friendliness of the IMI in terms of "outer interactions", such as hands, playing, and setup, and "inner interactions", such as freedom, expression, feedback, motivation, and learning (see Section 5.1).

G4. To evaluate whether the user's experience on emotion detection, such as images, timing, and colors, influences the overall perceived performance of IMI, through the quality of interaction with gesture sonification and the users' perception in recognizing the performer's personal style.

G5. To estimate the total influence of each entity on the overall perceived performance of IMI.

5.1 Validating the IMI operational model

Firstly, exploratory factor analysis (EFA) was performed in order to investigate the dimensions of the constructs proposed. All constructs were uni-dimensional, except for the construct with respect to the interaction with gesture sonification, which produced two dimensions (factors). These dimensions have been labelled "outer interactions", including items such as placing hands (loading = 0.901), playing comfortability (0.847), and setup environment to perform (0.624), and "inner interactions", including items such as freedom (0.656), expression (0.764), audio feedback (0.578), visual feedback (0.690), motivation (0.618), and learning (0.718). Furthermore, the Kaiser-Meyer-Olkin (KMO), measuring the sampling adequacy, and the Bartlett's test of sphericity, measuring the appropriateness of factor analysis, was used [Field 2005]. The KMO value found to be equal to 0.776 (i.e. above the critical value of 0.50) and Bartlett's test exact significance equal to 0.000 (i.e. bellow the critical value of 0.05). These findings, taking also into consideration the corresponding scree plot presented in Figure 6, indicated that factor analysis is appropriate for these data [Kaiser 1974]. In addition, the values of the estimated Cronbach Alphas (above 0.70) and the percentage of the total variance (above 50%), explained from factor analysis for each construct, verified the consistency of the survey instrument and the instrument content validity.

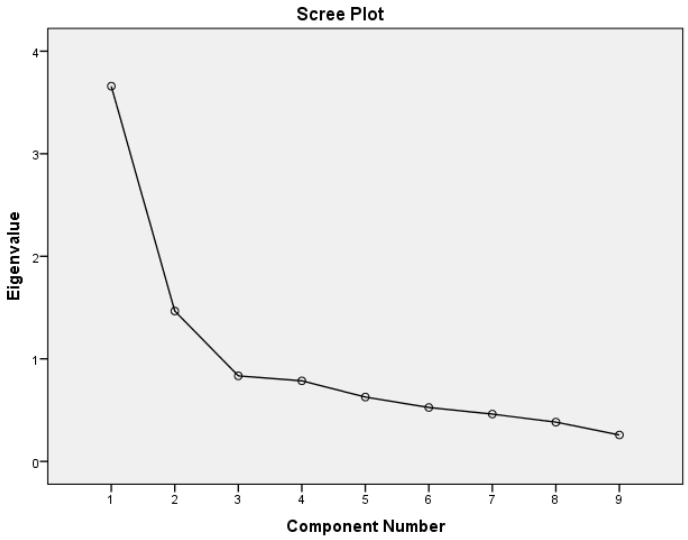

Fig. 6. Scree plot of factor analysis

Furthermore, a joint confirmatory factor analysis (CFA) of all constructs as well as the Kolmogorov – Smirnov normality test [Smirnov 1948] for each construct individually were performed and then the operational model was estimated. The CFA test indicated that the constructs could be used in the estimation of the operational model, in the form presented in Figure 5. The normality tests indicated that all the constructs followed normal distribution patterns and thus the maximum likelihood test could be used in the estimation of the operational model.

Table III presents the means and the standard deviations of all the constructs used in the study, and displays their bivariate correlation coefficients. Strong, positive and significant correlations between the variables involved are observed, supporting the hypotheses of the study. However, results based on correlations, although interesting, may be misleading due to the interactions between several variables [Katou et al. 2014].

Table III. Means, standard deviations and bivariate correlation coefficients of all the constructs

Constructs	Mean (standard deviation)	Correlation Coefficients				
		Emotion detection	Outer Interaction	Inner Interaction	Skills on recognizing	Performance of IMI
Emotion detection	3.025 (0.753)	1				
Outer Interaction	3.483 (0.795)	0.366**	1			
Inner Interaction	2.981 (0.711)	0.475**	0.439**	1		
Skills on Recognizing	3.127 (0.686)	0.421**	0.431**	0.561**	1	
Performance of IMI	3.406 (0.794)	0.465**	0.382**	0.539**	0.455**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Therefore, in order to isolate the possible links between the variables involved in the operational model presented in Figure 5, the estimated path diagrams for this proposed framework are presented in Figure 7. The boxes represent exogenous or endogenous observed variables and the circles represent the related latent variables. The light arrows indicate the observed variables that constitute

the related latent variables and the bold arrows indicate the structural relationships between the corresponding variables. For comparison purposes, the numbers that are assigned to each arrow show the estimated standardized coefficients. However, under the structural estimated standardized coefficients, the numbers in brackets present the actual estimated coefficients and their standard errors, indicating that the confidence intervals for the estimated coefficients are very narrow.

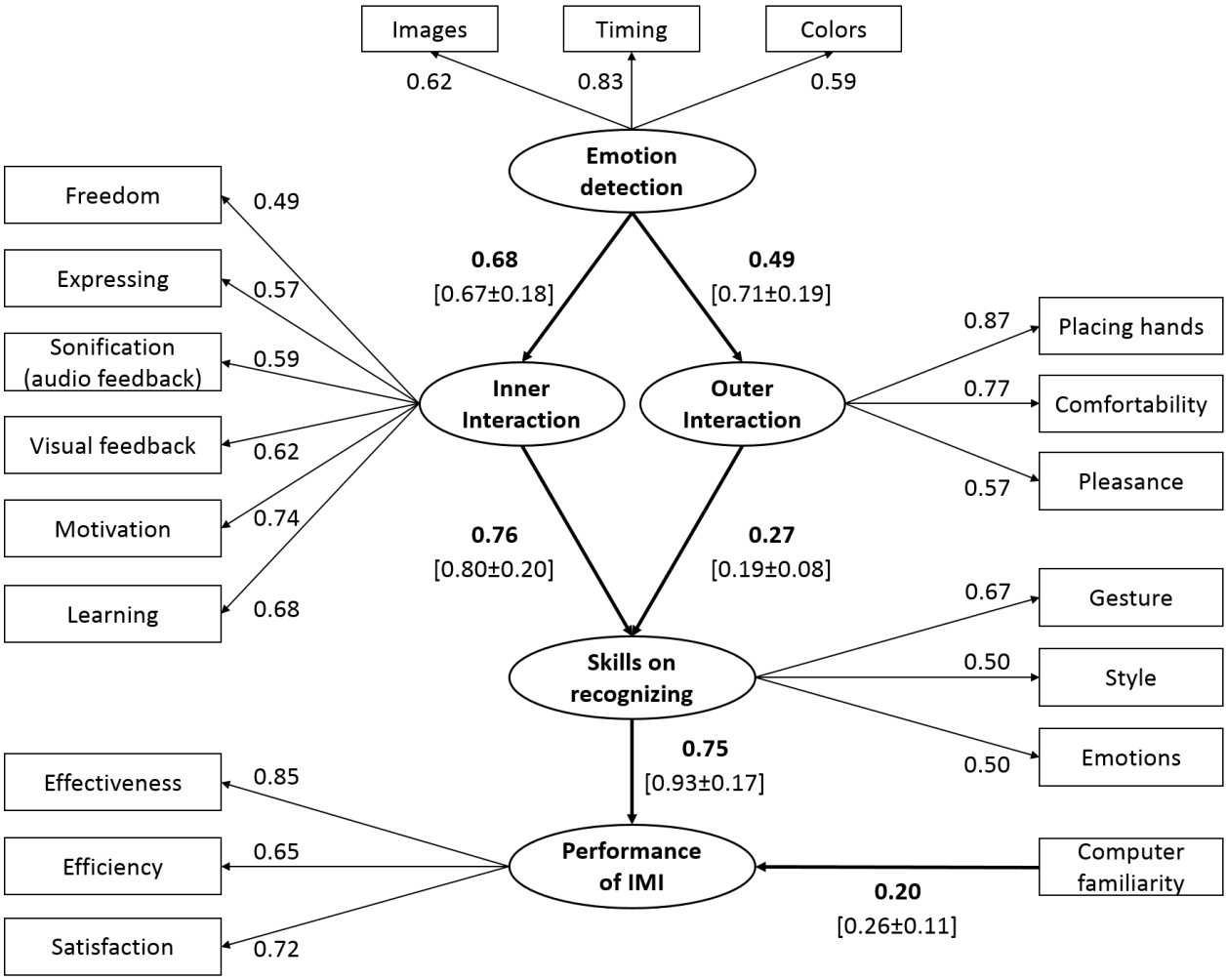

Fig. 7. Estimation results of the IMI operational model

Table IV presents the fit indices that are attached to the results presented in Figure 7. The performance of the IMI operational model is very satisfactory as it can predict the IMI processes with 93.7% overall accuracy. Taking into account that chi-square statistics may be inflated by significant correlations between constructs, the value of the normed-chi-square was used instead. In our case this value is very small (1.412), confirming the validity of our model and indicating that the proposed model is an adequate presentation of the entire set of relationships [Pedhazur and Pedhazur-Schelkin 1991]. In addition, the values of the GFI and the CFI are above the corresponding critical values, verifying that the structure of the model fits the empirical data satisfactorily. However, the value of the NFI (0.738) is much less than its critical value, indicating the usual tendency to underestimate fit in relatively small samples [Bentler 1990].

Table IV. Assessment indices of the IMI operational model

Assessment category	Fit Indices			
	ID	Description	Pass criteria	Value
Overall Performance Evaluation Indices	Chi-square	Exact significance of Chi-square statistic	>0.05	0.001
	Normed Chi-square	Chi-square / Degrees of freedom	<5	1.412
	RMSEA	Root Mean Squared Error Approximation (numerical value [0,1])	<0.10	0.063
Individual Fit Indices	GFI	Goodness of Fit Index (numerical value [0,1])	>0.70	0.837
	NFI	Normed Fit Index (numerical value [0,1])	>0.90	0.738
	CFI	Comparative Fit Index (numerical value [0,1])	>0.90	0.903

5.2 Evaluation results

The above results highlight that the model satisfactorily predicts performance (all standardized coefficients are significant and positive, and fit indices are acceptable overall). Moreover, the skills on recognizing gestures, styles and the emotions of a performer fully mediate [Baron and Kenny 1986] the relationship between interaction and performance. The ease for placing of hands, degree of comfort, motivation, and learning are the most important interactional factors with IMI. In terms of emotional elicitation, the results highlighted the timing of the user's emotional state as the most important factor. Furthermore, gesture recognition is the most important factor in determining performance effectiveness. Performance is also influenced by familiarity in using computers (all the other controls, such as sex, age, educational level, music literacy, used in the study were not significant). This means that IMI belongs to the so-called "Contingency Systems", which support the view that the system maximizes its performance according to the specific context in which it is operating [Delery and Doty 1996]. However, it should be taken into consideration that the model and its estimation is based on perceived subjective data. Perceived data do not undermine the usefulness of the model and the whole IMI evaluation exercise. In any case, a further technical assessment exercise could also be employed to measure performance of the system in a more objective manner.

Overall, Table V presents the total effect of each column variable on each row variable after standardizing all variables. For example, the standardized total (direct and indirect) effect of the "emotion detection" on satisfaction is 0.347. That is, due to both direct (unmediated) and indirect (mediated) effects of the "emotion detection" on satisfaction, when it goes up by 1 standard deviation, satisfaction goes up by 0.347 standard deviations (for further discussion of direct, indirect and total effects, see Kline [Kline 1998]).

Table V. Total standardized effects for IMI operational model

ID of Question (refers to the items in the Questionnaire in Appendix)	Variable Abbreviation	Constructs				
		Emotion detection	Inner Interaction	Outer Interaction	Skills on recognizing	Performance of IMI
	Inner Interaction	0.675	0.000	0.000	0.000	0.000
	Outer Interaction	0.489	0.000	0.000	0.000	0.000
	Skills on recognizing	0.644	0.761	0.267	0.000	0.000
	Performance of IMI	0.482	0.569	0.200	0.748	0.000
Q1.1.1	Hands	0.423	0.000	0.865	0.000	0.000

Q1.1.2	Comfort	0.374	0.000	0.766	0.000	0.000
Q1.1.3	Pleasant	0.277	0.000	0.566	0.000	0.000
Q1.1.4	Freedom	0.333	0.493	0.000	0.000	0.000
Q1.1.5	Expression	0.387	0.573	0.000	0.000	0.000
Q1.1.6	Sonification	0.396	0.586	0.000	0.000	0.000
Q1.1.7	Visual	0.418	0.619	0.000	0.000	0.000
Q1.1.8	Motivation	0.500	0.740	0.000	0.000	0.000
Q1.1.9	Learning	0.459	0.680	0.000	0.000	0.000
Q1.2.1	Colors	0.588	0.000	0.000	0.000	0.000
Q1.2.2	Time	0.835	0.000	0.000	0.000	0.000
Q1.2.3	Images	0.616	0.000	0.000	0.000	0.000
Q2.1	Gesture	0.429	0.506	0.178	0.666	0.000
Q2.2	Style	0.319	0.377	0.132	0.496	0.000
Q2.3	Emotions	0.322	0.381	0.134	0.500	0.000
Q3.1	Effectiveness	0.408	0.482	0.169	0.634	0.848
Q3.2	Efficiency	0.313	0.369	0.130	0.485	0.649
Q3.3	Satisfaction	0.347	0.410	0.144	0.539	0.721

Table V could guide the designers of the IMI to put effort into improving entities in the system in order to get better results, as perceived by the users of the system; however, any amendment should take into consideration the cost-benefit analysis of each change.

A summary of Table V is presented in Figure 8, where the relationship between the components (i.e. explanatory variables) and performance (i.e. dependent variable) of IMI is presented. This summary indicates both the mean values of each component and their contribution (loadings) to the performance of IMI. Thus, any amendment/update of the components will influence performance.

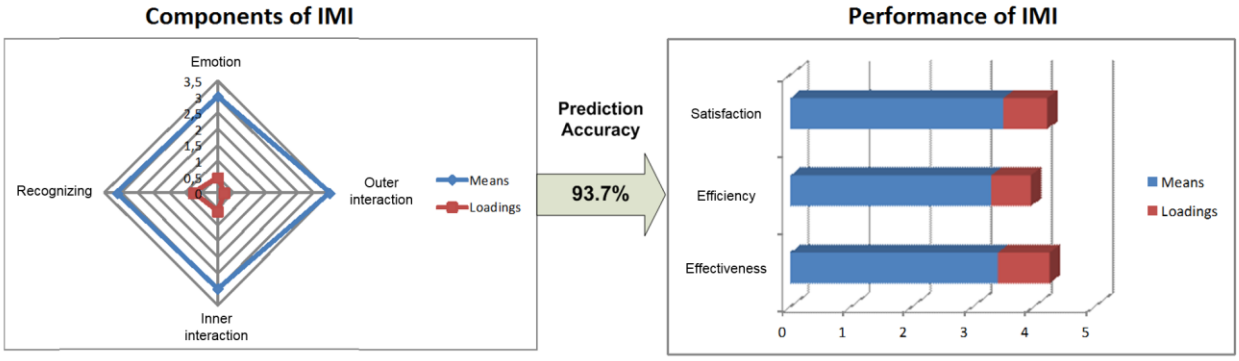

Fig. 8. The relationship between components of IMI and its performance

For amending/updating the components of IMI, the following rules should be considered:

(1) the prediction accuracy index, explaining the relationship between the explanatory variables (components of IMI) and the dependent variable (dimensions of performance) should be as close as possible to one (perfect prediction);
(2) the means of the variables (explanatory and/or dependent) should be as close to level five (perfect perceived user's evaluation response); and
(3) the loadings of the variables (explanatory and/or dependent) should be as close to one (perfect contribution to performance).

Additionally, the following information are also used:

a) a low mean of a variable means that there is "room" for the corresponding component to be improved; however, a cost-benefit analysis should also be used to investigate how easy it is to improve a component;
b) the high loading of a variable means that the corresponding component is important in determining performance.

This means that in ranking the amendments/updates of the components of IMI, it should be taken into consideration whether there is "room for improvement" and whether there is a "high contribution". This can be achieved by considering the following general rule, which accordingly determines a "component ranking index": "The lower the ratio of the mean of a component is, the higher the priority for amendments/updates for this component".

Following this rule, the major conclusions and recommendations with respect to IMI can now be summarized. First of all, the components of IMI predict performance with very high accuracy (93.7%). The perception of the dimensions of IMI performance range between 3.30 and 3.50 (on a five-level Likert scale), indicating that the performance of IMI is above average. According to the component ranking index, the components of the IMI, in a descending order, are arranged as follows: outer interaction with gesture sonification (17.40), emotion detection (6.27), inner interaction with gesture sonification (5.24), and skills on recognizing performer's qualities (4.17). Considering the arrangement in (3), it is recommended that the priority for improvements to be taken should be in the areas of "skills on recognizing performer's qualities" and "inner interaction with gesture sonification". Applying the same rule within the components of "skills on recognizing performer's qualities", priority for improvements should be taken in the areas of gestures (4.58), style (6.26), and emotions (6.36). Applying further the same rule within the components of "inner interaction with gesture sonification", the priority for improvements should be taken in the areas of motivation (3.96), visual feedback (4.68), audio feedback/sonification (4.95), expressing (4.84), learning (5.00), and freedom (6.04).

6. CONCLUSIONS

A prototype natural user interface, named the Intangible Musical Instrument (IMI), was presented in this paper. From a technical point of view, this first prototype is able to capture, model and recognize musical gestures and emotions as well as to sonify gestures and emotions. The IMI is conceived to transmit the multi-layer musical ICH to the public, by developing a unified interface framework that supports learning, performing and composing with gestures. This means that the learning of gestural knowledge of expert performers becomes a first-person experience with the IMI. Most importantly, a significant effort has been made to put the user at the core of the musical performance and of composition with gestures in both combined and autonomous ways.

Summarizing, the first evaluation of the IMI shows very satisfactory results in performance prediction. However, what emerges from the study described here is that future work should focus more on improvements in terms of research into expert musical gesture recognition as well as visual and audio feedback (sonification). Our future research should also focus on augmenting ordinary musical scores by providing gestural and emotional annotations together with the musical notation, in order to further facilitate the learning experience.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7-ICT-2011-9) under grant agreement n° 600676. We would also like to thank Frédéric Bevilacqua for giving us his permission to use the 'Gesture Follower' in testing our hypothesis.

REFERENCES

Reuben M. Baron and David A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–82.

Peter M. Bentler. 1990. Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.

Peter M. Bentler and Chih-Ping Chou. 1987. Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78-117.

Frédéric Bevilacqua, Fabrice Guédy, Nobert Schnell, Emmanuel Fléty and Nicolas Leroy. 2007. Wireless sensor interface and gesture-follower for music pedagogy. In *Proceedings of the International Conference of New Interfaces for Musical Expression*, New York, USA, 124-129.

Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, Fabrice Guédy and Nicolas Rasamimanana. 2010. Continuous realtime gesture following and recognition, *LNAI* 5934, 73–84.

Frédéric Bevilacqua, Norbert Schnell, Nicolas Rasamimanana, Bruno Zamborlin and Fabrice Guédy. 2011. *Online gesture analysis and control of audio processing*. In: J. Solis & K.C. Ng (Eds.). Musical Robots and Interactive Multimodal Systems (Springer Tracts in Advanced Robotics, 74, 127–142). Berlin, Heidelberg: Springer.

Jan L. Broeckx. 1981. Muziek, ratio en affect over de wisselwerking van rationeel denken en affectief beleven bij voortbrengst en ontvangst van muziek. Antwerpen: Metropolis.

Bernd Buxbaum. 2002. Optische Laufzeitmessung und CDMA auf Basis der PMD-Technologie mittels phasenvariabler PN-Modulation, Schaker Verlag, Aachen.

Claude Cadoz and Marcelo M. Wanderley. 2000. *Gesture-music. Trends in gestural control of music*. In: M.M. Wanderley and M. Battier (Eds.), Trends in gestural control of music. Paris, IRCAM/Centre Pompidou, 71–94.

Antonio Camurri, Gualtiero Volpe, Giovanni de Poli and Marc Leman. 2005. Communicating Expressiveness and Affect in Multimodal Interactive Systems. *IEEE MultiMedia*, 12(1), 43-53.

Martin R. L. Clayton. 2000. Time in Indian Music: Rhythm Metre and Form in Indian Rag Performance. Oxford: Oxford University Press.

Wilson Coker. 1972. Music & Meaning: A Theoretical Introduction to Musical Aesthetics. New York: The Free Press.

Naomi Cumming. 2000. *The sonic self: Musical subjectivity and signification*. Bloomington: Indiana University Press.

Francois Delalande. 1988. *La gestique de Gould: éléments pour une sémiologie du geste musical*. In: G. Guertin (Eds.) Glenn Gould pluriel, Louise Courteau éditrice, Montréal.

John Delery and Harold D. Doty. 1996. Modes of theorizing in strategic human resource management: test of universalistic, contingency and configurational performance predictions. *Academy of Management Journal*, 39, 802-835.

Andy P. Field. 2005. *Discovering statistics using SPSS* (2nd edition). London: Sage.

Anders Friberg and Johan Sundberg. 1999. Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, 105(3), 146-148.

Aalf Gabrielsson and Erik Lindström. 2010. *The role of structure in the musical expression of emotions*. In: P.N. Juslin & J.A. Sloboda (Eds.). Handbook of music and emotions theory, research, applications. New York, NY: Oxford University Press, 367-400.

Leontios J. Hadjileontiadis. 2014. Conceptual Blending in Biomusic Composition Space: The "Brainswarm" Paradigm. In *Proceedings of the ICMC/ SMC Conference*. Athens.

Robert S. Hatten. 1994. Musical meaning in Beethoven: markedness, correlation, and interpretation. Bloomington: Indiana University Press.

Rolf Inge Godøy and Mark Leman. 2009. *Musical gestures: Sound, Movement, and Meaning*. In: Rolf Inge Godoy and Marc Leman (Eds.), Routledge, New York.

Alexander R. Jensenius. 2007. *ACTION – SOUND Developing Methods and Tools to Study*. Ph.D. Dissertation, University of Oslo.

Patrik N. Juslin. 2003. Five facets of musical expression: A psychologist's perspective on music performance, *Psychology of Music*, 31(1), pp. 273-302.

Patrik N. Juslin and John Sloboda. 2010. *Handbook of music and emotions theory, research, applications*. New York, NY: Oxford University Press.

Patrik N. Juslin and Renee Timmers. 2010. *Expression and communication of emotion in music performance*. In: P.N. Juslin & J.A. Sloboda (Eds.). Handbook of music and emotions theory, research, applications. New York, NY: Oxford University Press, 453-489.

Henry F. Kaiser. 1974. An index of factorial simplicity. *Psychometrika*, 39, 31-36.

Anastasia A. Katou, Pawan S. Budhwar and Charmi Patel. 2014. Content vs. process in the HRM-performance relationship: An empirical examination. *Human Resource Management*, 53(4), 527-544.

Peter Keller. 2008. *Joint action in music performance*. In: F. Morganti, A. Carassa and G. Riva (Eds.). Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interaction. Amsterdam, 205-221.

Rex B. Kline. 1998. Principles and practice of structural equation modeling. NY: Guilford Press.

Stefan Koelsch and Walter A. Siebel. 2005. Towards a neural basis of music perception, *Trends in Cognitive Sciences*, 9(12), 578 – 584.

Marc Leman. 2010. Music, Gesture, and the Formation of Embodied Meaning. *Musical gestures: Sound, Movement, and Meaning*. 126-153.

Pieter-Jan Maes, Marc Leman, Micheline Lesaffre, Michiel Demey and Dirk Moelants. 2010. From expressive gesture to sound.

The development of an embodied mapping trajectory inside a musical interface. *Journal on Multimodal User Interfaces*, 3(1), 67-78.

David Mcneill. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

Alva Noë. 2004. *Action in Perception*. Cambridge, MA: MIT Press.

Elazar J. Pedhazur and Liora Pedhazur-Schmelkin. 1991. Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Lawrence Erlbaum.

Carroll C. Pratt. 1931/1968. The meaning of music: A study in psychological aesthetics, New York: Johnson.

James A. Russell. 1980. A circumflex model of affect, *Journal of Personality and Social Psychology*, 39, 1161-1178.

Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman and Andrew Blake. 2013. Real-Time Human Pose Recognition in Parts from a Single Depth Image. *Machine Learning for Computer Vision*, SCI, 411, 119-135.

Alexander Truslit. 1938. *Gestaltung und Bewegung in der Musik*. Berlin-Lichterfelde: C.F. Vieweg.

N. Smirnov. 1948. Table for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics, 19, 279–281.

UNESCO 2003. Convention of the Safeguarding of Intangible Cultural Heritage of UNESCO. Available: https://ich.unesco.org/en/convention.

Frank Weichert, Daniel Bachmann, Bartholomäus Rudak and Denis Fisseler. 2013. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13, 6380–6393.

Matthias Wiedemann, Markus Sauer, Frauke Driewer and Klaus Schilling. 2008. Analysis and characterization of the PMD camera for application in mobile robotics. In *Proceedings of the 17th IFAC World Congress*, 6-11.

Liwei Zhao. 2001. Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures. Ph.D. Dissertation. University of Pennsylvania, Philadelphia, PA, USA. AAI3015399.

Appendix:

SECTION 1: QUALITY CHARACTERISTICS OF INTANGIBLE MUSICAL INSTRUMENT

1. How would you rate the interaction with IMI?

No.	Interactive Characteristics	Not at all				Very much
		1	2	3	4	5
1	How easy did you find it to place your hands correctly (correct octave) on the IMI?					
2	How comfortable did you find the playing/performance of the musical gestures on the IMI?					
3	How pleasant did you find the setup environment to perform with in terms of design and aesthetics (Plexiglas, motion sensors, brain activity sensors)?					
4	Did you feel that your gestures were more free compared to a real piano?					
5	To what extent could you express yourself through the IMI?					
6	How much did the IMI help you to improve/correct your gesture during performance, by comparing the sound that you produced in real-time (sonification/audio feedback) with the sound that you listened to while watching the video with the expert's gestures?					
7	Did the virtual avatar (visual feedback) help you improve/correct the performance of your gestures?					
8	How much did the IMI activity of "Final Challenge" motivate you to focus more on the learning of fundamental musical gestures and activities?					
9	How much do you think the IMI would help you to learn musical gestures?					

2. How would you rate the quality of emotion detection?

No.	Emotions	Very Bad				Very good
		1	2	3	4	5
1	To what extent did the images that were shown to you excite you emotionally?					
2	To what extent was the time enough for you to follow the emotional state at every level of the game?					
3	To what extent were the colors representative of emotions?					

SECTION 2: USER'S MUSIC PERCEPTION

How easily can you recognize the personal style (expressiveness, gestures, sound) of a performer while s/he interprets classical or contemporary composers?

No.	Performer's qualities	Very Bad				Very Good
		1	2	3	4	5
1	Gestures					
2	Music Style (*e.g. to understand differences between playing/performance while listening to the same piece of music*)					
3	Emotions					

SECTION 3: PERFORMANCE OF THE IMI

How would you rate the overall performance of the IMI?

No.	Performance Measures	Very Bad				Very Good
		1	2	3	4	5
1	**Effectiveness** (*if the IMI meets its objectives*)					
2	**Efficiency** (*if the IMI responds satisfactorily and in a short time to gestures, emotions and sound production*)					
3	**Satisfaction** (*if the IMI provides satisfaction to the user*)					

Thank you very much for your co-operation.