HAL
open science

INTANGIBLE CULTURAL HERITAGE AND NEW TECHNOLOGIES: CHALLENGES AND OPPORTUNITIES FOR CULTURAL PRESERVATION AND DEVELOPMENT

Marilena Alivizatou-Barakou, Alexandros Kitsikidis, Filareti Tsalakanidou, Kosmas Dimitropoulos, Chantas Giannis, Spiros Nikolopoulos, Samer Al Kork, Bruce Denby, Lise Buchman, Martine Adda-Decker, et al.

To cite this version:

Marilena Alivizatou-Barakou, Alexandros Kitsikidis, Filareti Tsalakanidou, Kosmas Dimitropoulos, Chantas Giannis, et al.. INTANGIBLE CULTURAL HERITAGE AND NEW TECHNOLOGIES: CHALLENGES AND OPPORTUNITIES FOR CULTURAL PRESERVATION AND DEVELOPMENT. Mixed Reality and Gamification for Cultural Heritage, Springer International Publishing; Springer International Publishing; Springer International Publishing, pp.129-158, 2017, 10.1007/978-3-319-49607-8_5 . hal-01693634

HAL Id: hal-01693634

https://minesparis-psl.hal.science/hal-01693634v1

Submitted on 21 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter 4
INTANGIBLE CULTURAL HERITAGE AND NEW TECHNOLOGIES: CHALLENGES AND OPPORTUNITIES FOR CULTURAL PRESERVATION AND DEVELOPMENT

Alivizatou-Barakou M.[1], Kitsikidis A.[2], Tsalakanidou F.[2], Dimitropoulos K.[2], Chantas, G.[2], Nikolopoulos S.[2], Al Kork S., Denby B., Buchman, L., Adda-Decker M., Pillot-Loiseau C., Tillmane J., Dupont S., Picart B., Pozzi F., Ott M., Yilmaz E., Charisis V., Hadjidimitriou S., Hadjileontiadis L., Cotescu M., Volioti C., Manitsaris, A., Manitsaris S. and Grammalidis N.[2]

Abstract. Intangible Cultural Heritage (ICH) is a relatively recent term coined to represent living cultural expressions and practices, which are recognised by communities as distinct aspects of identity. The safeguarding of ICH has become a topic of international concern primarily through the work of UNESCO (United Nations Educational, Scientific and Cultural Organisation). However, little research has been done on the role of new technologies in the preservation and transmission of intangible heritage. The chapter examines resources, projects and technologies providing access to ICH and identifies gaps and constraints. It draws on research conducted within the scope of the collaborative research project, i-Treasures. In so doing, it covers the state of the art in technologies that could be employed for access, capture and analysis of ICH in order to highlight how specific new technologies can contribute to the transmission and safeguarding of ICH.

Keywords: Intangible Cultural Heritage, ICT, safeguarding, transmission, semantic analysis, 3D Visualization, game-like educational applications

4.1 Introduction

In the last decades the protection and promotion of cultural heritage (primarily in the form of monuments, historic sites, artefacts and more recently cultural expressions) has become a central topic of European and international cultural policy. Since the end of World War II, UNESCO has been a key organization in defin-

[1] UCL Institute of Archaeology, London WC1H 0PY UK, m.alivizatou@ucl.ac.uk

[2] Information Technologies Institute, CERTH, 57001 Thessaloniki, Greece, ngramm@iti.gr

ing cultural heritage and ensuring its protection through the adoption of a series of conventions, financial and administrative measures. Parallel to the work of UNESCO, governmental and non-governmental organizations, professional associations and academic institutions around Europe have been involved with documenting and providing access to different forms of cultural heritage (ranging from archaeological sites and natural parks to museum collections and folk traditions). In this process, a significant body of resources dealing with the documentation and promotion of cultural heritage through different technologies has been developed. There is little doubt that digital technologies have revolutionized scientific and public access to cultural heritage [1,2].

Following the adoption of the UNESCO Convention for the Safeguarding of Intangible Heritage in 2003, the protection of cultural traditions has become prominent on an international level. One of the key arguments in this area is that humanity's intangible heritage is threatened by processes of globalization. Modern technologies and mass culture are often regarded as a threat to the survival of traditional expressions. According to the Convention, it falls upon national governments, cultural organizations and practicing communities to transmit these vulnerable cultural expressions to the next generations. Safeguarding activities vary according to local and national contexts. Interestingly, although modern technologies are often identified as a threat to traditional expressions, it is these very technological innovations that frequently play a key part in the preservation and dissemination of ICH.

Drawing on the existing literature and body of research, this chapter will provide an overview of current safeguarding programmes with a particular focus on specific technological methods and how they contribute to the documentation and transmission of intangible heritage. What we argue is that new technologies can provide innovative approaches to the transmission and dissemination of intangible heritage by supporting human interaction and communication.

More precisely, this chapter offers an overview of the literature, resources, projects and technologies providing access to ICH. It aims to identify gaps and constraints of existing projects in the area. It covers the state of the art in technologies that could be employed for access, capture and analysis of ICH in order to highlight how specific new technologies can contribute to the transmission and safeguarding of ICH.

Section 4.2 looks at previous work, including the broad scope of safeguarding activities supported by UNESCO, which mainly consist of national and international inventories and rely mostly on archival approaches. Furthermore, projects run by museums, cultural organizations and grassroots initiatives, which are driven by community participation, are examined. Section 4.3 takes a closer look at specific technological methods (facial expression analysis and modelling, vocal tract sensing and modelling, body motion and gesture recognition, semantic multimedia analysis, 3D visualisation and text to song) that relate to the documentation and transmission of ICH. Finally, Section 4.4 draws conclusions and discusses future work.

4.2 Previous work

In the 2003 Convention for the Safeguarding of Intangible Heritage, UNESCO defines safeguarding as 'measures aimed at ensuring the viability of intangible heritage, including the identification, documentation, research, preservation, protection, promotion, enhancement, transmission, particularly through formal and informal education and revitalisation of the various aspects of such heritage. Although their role in safeguarding intangible heritage is not directly addressed in the Convention, new technologies can play an important part in areas of identification, documentation, preservation, promotion and education. Audio-visual documentation, digital and multimedia resources from the areas of information and communication technologies can provide useful tools for recording and collecting information about expressions of intangible heritage.

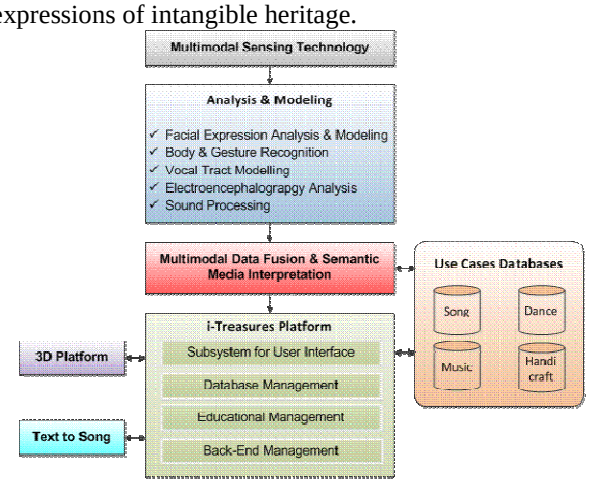

Fig. 4.1: Overview of i-Treasures system (© 2015 EC FP7 i-Treasures project, Reprinted with Permission)

Taking forward this idea, i-Treasures project [87] attempts to explore the challenges and opportunities that emerge when considering the safeguarding of intangible heritage from a technological perspective. More specifically, its overall goal is to develop an open and extendable platform to provide access to intangible cultural heritage resources for research and education. The core of the system lies in the identification of specific features or patterns (e.g. postures, audio patterns, etc.) using multi-sensor technology (e.g. cameras, microphones, EEG etc.) from different ICH forms. Subsequently, data fusion analysis is applied to exploit information across different modalities, while context and content are integrated for mapping the set of low or medium-level multimedia features to high-level concepts using probabilistic inference, i.e. transforming the extracted data into a level of interpretation that is understandable by humans. This information, coupled with other cultural resources, is accessible via the i-Treasures platform (an open-source CMS), in order to enable the widest possible participation of communities, groups

and individuals in the safeguarding of ICH. The platform gives access to different types of content (e.g. text, audio, images, video, 3D graphics) from different types of heritage or educational institutions. Furthermore, using the latest advances in web-based game engines, a learning environment is developed to enhance training and evaluation of the learner's performance by means of sensorimotor learning. Finally, a Text-to-Song system can also be used allowing the user to enter text and/or notes and produce the equivalent singing voice. An overview of the system is presented in Fig. 4.1.

As shown in the cases discussed below, the first attempts at using documentation technologies for safeguarding expressions of intangible heritage have had a primarily archival and encyclopaedic orientation.

Even before the adoption of the 2003 Convention, UNESCO supported projects aimed at the safeguarding of intangible heritage. For example, the Red Book of Endangered Languages (subsequently known as Atlas of Endangered Languages) is a publication and online resource that provides basic information on more than two thousand languages. It has taken the form of an online map and archival **resource** and provides an encyclopaedic list of world languages ranging from vulnerable to extinct. However, the information available online is limited and there are limited learning possibilities available.

Another example has been the Traditional Music of the World, a project that includes recordings of traditional music [3]. The recordings have been made by ethnomusicologists in situ and then copied on vinyl and CD format. Relevant photographs accompany the audio recordings. The project has made available these recordings to an international audience and raising awareness about traditional music. However, it seems to act primarily as an archival resource and has limited educational application. Moreover, there is no online access to the recordings.

The 2003 Convention has put in place two lists that act as mechanisms for identifying intangible heritage on a global level: The International List of the Intangible Cultural Heritage of Humanity and the List of Intangible Cultural Heritage in Need of Urgent Safeguarding. These are international inventories of traditional expressions which are accessible online and include photographs and audio-visual recordings of cultural expressions. Their primary function is that of an archival resource that raises awareness about the listed expressions and their communities. The drawback of the lists is that they seem to be serving primarily promotional objectives rather than activities that have a direct impact on local communities [4]. Moreover, the amount of documentation available online is relatively limited. Prior to these lists, UNESCO supported the programme for the Proclamation of Masterpieces of the Oral and Intangible Heritage of Humanity [5,6]. This was the first international project to raise major awareness at a governmental level and influence the adoption of the 2003 Convention. The database of selected masterpieces of intangible heritage is available online and includes audio-visual and photo material, but has a relatively limited educational scope.

UNESCO has also supported national inventories of intangible heritage in several countries. Some examples with information available on the Internet are the

national inventories and/ or registers of Japan, Brazil, Portugal, Bulgaria and Venezuela. Different methodologies have been used for the creation of these inventories. In Bulgaria, catalogues were set up drawing on information collected through questionnaires distributed to local communities via cultural centres. In Japan and Brazil, inventories were drawn mostly though ethnographic research. The national inventories have raised awareness about the importance of intangible heritage among local communities. However, the amount of documentation available seems to be relatively limited. An innovative methodology has been used for the documentation of intangible heritage in Scotland. The UK Commission of UNESCO in collaboration with the Scottish Arts Council funds this project. It consists of an online archive of Scottish intangible heritage that is open to the public in the form of a wiki. It contains photographic and audio-visual documentation and uses participatory approaches.

In terms of the transmission and revitalisation of cultural expressions of intangible heritage, UNESCO collaborates with relevant organisations for their promotion and enhancement. In this process, UNESCO has proposed a group of accredited NGOs and INGOs to provide advice and support in the nomination of elements of intangible heritage for the International List. For the case of tradition-al dance, The International Council for the Organisation of Folklore Festivals (http://www.cioff.org) and the European Association of Folklore Festivals (http://www.eaff.eu/en) are relevant cases to consider. Their work has allowed for a more direct contact between UNESCO and cultural practitioners and the enhanced visibility of traditional expressions.

Further uses of new technologies in the safeguarding of intangible heritage have been made by local community centres, museums and EU-funded research networks.

The EU has supported projects relating directly and indirectly to the transmission of intangible heritage. For example, the I-maestro project aims to build a multimedia environment for technology enhanced music education. This employs self-learning environments, gestural interfaces and augmented instruments promoting new methods for music training. The question that is raised, however, (and is also relevant for i-Treasures) is whether technology risks replacing human interaction as a process of transmission.

More directly related to intangible heritage and local development is the EU-funded project entitled Cultural Capital Counts. The project aims to enable a positive development of six regions in Central Europe by focusing on intangible heritage resources like living traditions, knowledge and talents. By developing a strategy that is based on intangible cultural resources, the project aims to enable a sustainable regional development in order to increase the region's attractiveness for enterprises. The project appears to develop around a website that contains a list of various traditions and expressions of intangible heritage in the six regions. It proposes strategies for local, sustainable development and collaborative research. But the focus seems to be more on the commercialization of intangible heritage

than on how these practices and traditions can be transmitted to the next generations.

Europeana (http://www.europeana.eu/) is a well-known EU portal for exploring the digital resources of Europe's museums, libraries, archives and audio-visual collections, thus offering direct access to millions of books, manuscripts, paintings, films, museum objects and archival records that have been digitised throughout Europe. Europeana's collection includes more than 53 million items and has released all metadata under a CC0 waiver, making it freely available for re-use. However, until now, Europeana includes just a small number of 3-D tangible objects and no 3-D representations of ICH.

Indigenous groups and communities have made innovative uses of new technologies in local attempts at safeguarding intangible heritage. The Oral Traditions Project of the Vanuatu Cultural Centre is a useful case to consider. The national museum and cultural centre runs a project [7] through a network of volunteers, called fieldworkers. These are representatives of different communities who conduct research on traditional customs and cultural expressions. The fieldworkers have been trained in ethnographic research methods and photographic and audio-visual documentation. The material collected during their research is kept in a specifically designated room of the national museum. The project has been instrumental in raising awareness about the importance of traditional culture. Its primary function is to create a 'memory-bank' of traditional culture and languages, but the collected material are not only kept for posterity but used in educational programmes for schools, the museum and the radio and community development. A key theme is the idea of 'heritage for development' translated in eco- and cultural tourism projects. The project is active since the early 1970s and some of the issues raised relate to the limited budget, the engagement of fieldworkers and how best to protect traditional culture from commercialization [8].

Online learning resources constitute another area of intangible heritage preservation. For instance, many indigenous groups in partnership with museums have created online heritage resources with a pedagogical focus. One example is the online resource created by the U'mista Cultural Centre for the interpretation of ceremonial masks. The resource, entitled 'The Story of the Masks' presents stories related to the masks, which are narrated by tribe members. The specific project involves the digitization of relevant content, the use of audio-visual technology and the creation of a website. Through the project community members are empowered to share their stories and memories.

Interestingly, the above projects are driven and managed by local communities. As such they are not only about documenting and archiving intangible heritage but also making the information available to cultural practitioners and the new generations and using it in educational programmes and activities. However, the technologies discussed so far are primarily archival and their potential for transmission and education relies primarily on how they can be used further by local agents and documentation institutions, like museums and cultural centres.

Taking the idea of participatory methods forward, i-Treasures project aims to make a contribution to collaborations between researchers and local communities by employing technologies that allow for a more direct interaction in terms of learning and transmission. The idea is to empower local actors to use new technologies as a new learning tool by bringing into the discussion a different range of technologies that are discussed in more detail below. One of the central arguments of the project is that although modern technologies cannot replace human interaction in the transmission of intangible heritage, they can contribute significantly to processes of dissemination, especially among younger generations. For this reason, the project proposes the development of territorial schools acting as local hubs for the transmission of local intangible heritage expressions.

4.3 Modern Technologies in the Transmission and Documentation of Intangible Heritage

This Section looks more closely at the technological methods that could be employed for safeguarding and transmission of intangible heritage. New technology can be used not only for the digitization and archiving of cultural expressions but also in terms of cultural transmission, education and community development. Although technology cannot replace human interaction, it can nevertheless support cultural transmission in new and innovative ways.

To this end, this chapter provides a detailed analysis of the state-of-the-art in some of the main technologies for capture, preservation and transmission of the ICH (Fig. 4.2), namely: i) Facial expression analysis and modelling, ii) Vocal tract sensing and modelling, iii) Body motion and gesture recognition, iv) Encephalography analysis, v) Semantic multimedia analysis, vi) 3D Visualization vii) Text-to-Song synthesis.

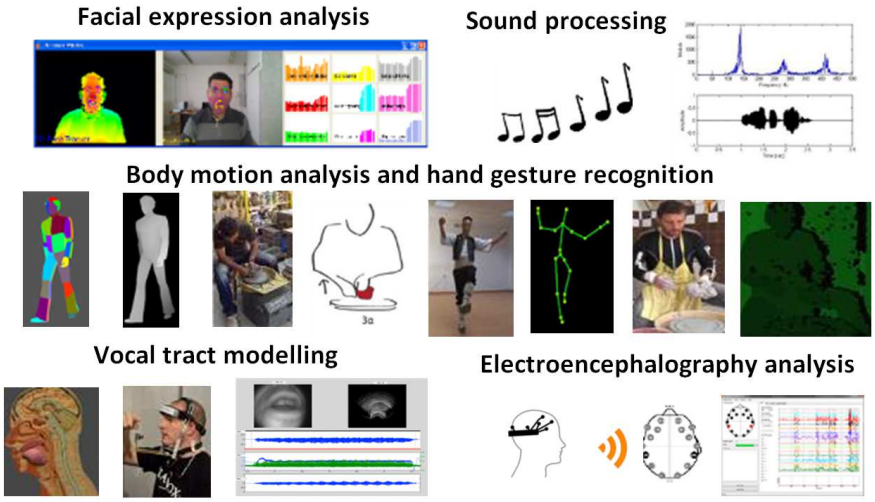

Fig. 4.2: Overview of main capture technologies for ICH (© 2015 EC FP7 i-Treasures project, Reprinted with Permission)

For each of the above, some key developments in each specific field are presented and the potential use and contribution of each technology in the preservation of intangible heritage are discussed.

4.3.1 Facial expression analysis and modelling

Facial expressions are one of the most cogent, naturally pre-eminent means for human beings to communicate emotions and affective states, to clarify and stress what is said, to signal comprehension, disagreement, and intentions and in brief regulate interactions with the environment and other persons in the vicinity [9]. Facial expressions are generated by facial muscle contractions, which result in temporary facial deformations of facial geometry and texture. Human faces are estimated to be capable of more than ten thousand different expressions. This versatility makes non-verbal expressions of the face extremely efficient and honest, unless deliberately manipulated. Many of these expressions are directly associated with emotions and affective states such as happiness, sadness, anger, fear, surprise, disgust, shame, anguish and interest, which are universally recognized [10].

This natural means of communication becomes even more important in the case of artistic expressions like singing or acting where the face and body are the main tools used by the performer to communicate the emotional aspects of their role. A great singing performance is not only the result of a great voice but also reflects the emotional involvement of the performer who expresses what he feels through his voice and body.

The performer's facial expressions could be analysed in terms of facial actions, which can either be used as a means of extracting facial muscle movements useful for describing the performer's technique or as a means for decoding the performer's emotional state.

The preservation and transmission of this expression involves more than analysing voice and music patterns and decoding voice articulation. It should also involve analysing and preserving the expressive and emotional aspects revealed by the performer's face, since the performance is more than correct voice articulation: it is also emotion revealed through voice and face. This is also very important for educational purposes. New singers will not only be taught how to use their vocal tract to sing different types of songs but can also learn how to give a complete performance.

Besides the emotional aspect, facial expressions can also be used to reveal details of the performer's technique, e.g. how much he/she opens the mouth while singing.

4.3.2 Vocal tract sensing and modelling

Since the dawn of human communication, man has been curious about the speech production mechanism, and has sought to model and exploit it in a variety of useful applications. The first vocal tract models were physical models, constructed of tubes and valves and resonators, which sought to duplicate the intricate process by which speech is produced in the human vocal tract via the articulators: the larynx (vocal folds), tongue, lips, teeth, jaw, and nasal cavity. With the advent of powerful digital computers, it became possible to produce 2D and 3D vocal tract models of surprising realism in software, often referred to as 'talking heads' [11,12,13] which, when coupled with an appropriate acoustic simulation, allow to synthesize speech in a way totally analogous to actual human speech production. Although, such so-called articulatory synthesis systems have been claimed by some researchers for having poorer performance than the codebook-style vocoder synthesizers, articulatory synthesis remains an active area of research, as many researchers believe it will ultimately lead to the most effective means of communication between man and machines.

To model the vocal tract effectively, it is necessary to study and understand its physical characteristics. Early studies on cadavers were the first sources of such information, followed by various types of endoscopic investigations, many of which are still in use today, but in the 20th century, the non-invasive nature of real time medical imaging techniques led to significant breakthroughs in vocal tract sensing. Very high-resolution real-time imaging of the entire vocal tract is possible using cineradiography (X-rays) and magnetic resonance imaging (MRI). As the use of X-rays on living subjects is regulated by strict radiation exposure limits, its use as a research tool is rather limited, although a number of studies have been

carried out [14,15]. Concerning MRI, although real time studies of the vocal tract have been carried out [14,16,17], the procedure requires the subject to recline within the confines of a very constrained volume containing a strong magnetic field. Time on an MRI machine is also very expensive, and, finally, the repetition rate of MRI, at best several Hertz, is insufficient for a delicate, real time physioacoustic study of speech production.

The routine availability of inexpensive, powerful computing resources, are today beginning to make unexpected inroads into a variety of new fields.

In addition, the combination of instruments might enhance the level of knowledge. For example, in the Sardinian Canto a Tenore some singers use a traditional laryngeal phonation, others use a method that pitch-doubles the fundamental frequency [18,19]. It is not clear if this is done by vibrating both the vocal folds and the ventricular folds as is found in diplophonia, or whether this is done by amplifying an overtone as is done in Tuva throat singing. The combination of ultrasound and EGG could allow the recording of the tongue, anterior pharyngeal wall and vocal folds during the methods. Between them, all the structures and behaviours of interest could be recorded, and allow visual and auditory documentation of the technique for purposes of archiving and future teaching. Like [20], it's possible to assess changing true vocal fold length with ultrasonography, and to observe vowel tongue shapes in untrained and trained singers [21].

Sensors that can be considered for data collection include: microphone, external photoglottography (ePGG), electroglottography (EGG) (for vocal cords, ultrasound (for tongue contour detection), RGB or RGB-D cameras (for lip/mouth movement), piezoelectric accelerometer, breathing belt.

Data from these sensors may be used for studies such as: a) pharyngeal or labial embellishment (soloists), b) nature of tiling, c) position of tongue and lips d) vocal quality tessitura of voice alone and ornamentations, e) comparison of voice alone/accompanied and e) correlation between body gestures and laryngeal gestures.

4.3.3 Body motion and gesture recognition

The study of human body motion is central in different scientific fields and applications. In the last decade, 3D motion capture systems have known a rapid evolution and substantial improvements, which have attracted the attention of many application fields, such as medicine, sports, entertainment, etc.

The applications of motion capture are numerous in different application fields and the related research directions can be categorized as follows:

- Motion capture system design: motion capture technologies, developing new approaches for motion capture, or improving the current motion capture tools.

- Motion capture for motion analysis: the use of existing motion capture systems for understanding the motion, gesture recognition, extracting information from motion capture sequences, analysing similarities and differences between motions, characterize the motion and recognize specific information (identity, style, activity, etc.) from the motion capture sequence, etc.
- Motion capture for animation: the use of motion capture, performed either in real-time or offline, to animate virtual characters using motions recorded from human subjects.

Motion capture (or *mocap*) systems can be divided into two main categories: marker-based and marker-less technologies. Even if some very important improvements have been made in the last years, no perfect system exists, each one having its own advantages and drawbacks.

Marker-based systems include optical systems and inertial systems (accelerometers, gyroscopes, etc.). The optical motion capture systems are based on a set of cameras around the capture scene and on markers, reflecting or emitting light, placed on the body of the performer. Various types of sensors [23] and [24] or commercial interfaces (e.g. Wii joystick, MotionPod or the IGS-190 inertial motion capture suits from Animazoo) can easily provide real-time access to motion information. On the contrary, markerless technologies do not require subjects to wear specific equipment for tracking and are usually based on computer vision approaches. Even if the accuracy and sensitivity of the tracking results do not yet meet the needs of the industry for the usual use of motion capture for animation, marker-less systems are the future of the field. Nonetheless, markerless systems still suffer from a lack of precision and cannot compete with marker-based technologies that now reach sub millimetre precision in real time. On the other hand, marker based systems are often very expensive and need a more complicated setup.

Markerless motion capture technologies based on real-time depth sensing systems have taken a huge step ahead with the release of Microsoft Kinect and its accompanying skeleton tracking software (Kinect for Windows) and other affordable depth cameras (ASUS Xtion , PMD nano). These sensors are relatively cheap and offer a balance in usability and cost compared to optical and inertial motion capture systems. Kinect produces a depth-map stream at 30 frames per second with subsequent real-time human skeleton tracking. Estimation of the positions of 20 predefined joints that constitute a skeleton of a person is provided by software SDKs (Microsoft Kinect SDK, OpenNI) together with the rotational data of bones. Subsequent algorithmic processing can then be applied in order to detect the actions of the tracked person. The estimated 3D joint positions are noisy and may have significant errors when there are occlusions, which pose an additional challenge to action detection problem. Multi Kinect setups with subsequent skeleton fusion techniques have been employed to combat the occlusion problems [22].

In conclusion, we can say that no perfect motion capture system exists. All systems have their advantages and drawbacks, and must be carefully chosen accord-

ing to the use case scenarios in which they are to be used. A compromise must be found between motion capture precision, the need for burdensome sensors and other external constraints like the motion capture area, the lighting environments, the portability of the system, etc.

4.3.3.1 Motion capture technologies for dance applications

As the interdisciplinary artist Marc Boucher says in [25] “Motion-capture is the most objective form of dance notation insofar as it does not rely on subjective appreciation and verbal descriptions of individuals but rather on predetermined mathematical means of specifying spatial coordinates along x, y and z axes at given moments for each marker. These data can be interpreted (inscribed, 'read,' and 'performed') cybernetically (human-machine communication) while previous dance notation methods are based on symbolic representations, written and read by humans alone.” However, as discussed above, all motion capture solutions have advantages and drawback, and even though motion capture is the most informative tool for recording dance, issues like obtrusiveness of markers, need to wear specific costumes and motion recoding precision are different subjects that require further investigation and appropriate solutions. Furthermore, motion capture is not yet widely known, and its costs and complexity have also prevented this technology to reach most artists and dancers. A widely adoption of these technologies needs adapted and usable tools and convincing system demonstrations.

Although motion capture technologies are most often designed and developed in generic application purposes, we have identified several studies where new sensors were designed or adapted to be used in the specific use case of dance motion capture. The SENSEMBLE project [26] designed a system of compact, wireless sensor modules worn at the wrist or ankles of dancers and meant to capture expressive motions in dance ensembles. The collected data enabled them to study if the dancers of the ensemble were moving together, if some were leading or lagging, or responding to one another with complementary movements. However, this sensor is aimed to be worn at the wrists and ankles of a dancer, not at every body segment and thus does not consist of a true motion capture system since the whole body is not captured, and the dance motion cannot be reconstructed based on the recorded information. The sensor captures some information about the motion, but not the 3D motion itself.

Saltate! [27] is a wireless force sensors system mounted under the dancers' feet which is used to detect synchronisation mistakes, and emphasize the beats in the music when mistakes are detected in order to help the dancer stay in synchronisation with the music. Once again, the sensor records some information about the dance moves, and more especially about the feet interactions with the ground, but the whole body motion is not captured at all.

Other approaches consist in capturing the dancer's motion through motion capture in order to control the soundtrack through the gesture to music mapping. This

is for instance the approach followed by [28, 29], whose goal is mainly to explore possible relationships between gesture and music using the optical motion capture Vicon 8 system.

Detection, classification and evaluation of dance gestures and performances are research fields, in which existing commercial products have been often employed [30]. Experiences such as Harmonix' Dance Central video game series , where a player repeats the motion posed by an animated character are becoming commonplace. Research is being conducted on automatic evaluation of dance performance against the performance of a professional, within 3D virtual environments or virtual classes for dance learning [31,32].

Numerous research studies have addressed the issue of synthesizing new dance motion sequences. They often base their synthesis model on existing dance motion capture databases [33]. Although their aim is not to preserve cultural heritage of the dance content, these studies have developed interesting approaches and tools, which can be used in order to analyse dance motions and the synchronized music track. For instance, [33] have developed a dance move detection algorithm based on the curvature of the limbs' path, while [88] have developed and unsupervised dance-modelling approach based on Hidden Markov Models.

Laban movement analysis (LMA) is a method developed originally by Rudolf Laban, which aims at building a language capable of describing and documenting precisely all varieties of human movements. The Laban movement analysis describes movements through six main characteristics of the motion: body, effort, shape, space, relationship, and phrasing. Even though this method has its drawbacks and requires a long training, it is one of the very few attempts at building a vocabulary or dictionary of motions that have been adopted quite widely. [34] use Laban movement analysis (LMA) to extract movement qualities, which are used to automatically segment motion capture data of any kind. They hence use concepts initially developed for dance and apply them to general motions. Kahol et al. [35] implement an automated gesture segmentation dedicated to dance sequences.

Dance motion capture has also been attracting great interest recently in the performing arts for its use in interactive dance performances [36].

4.3.3.2 Hand and finger motion recognition

Hand motion recognition and more especially finger motion recognition is very different from the usual motion capture approaches, which are generally designed for full body motion capture. Although special gloves for capturing finger motion are commercially available, the above motion capture methods are usually not suitable for finger gesture recognition. In [80], recognition of the musical effect of the guitarist’s finger motions on discrete time events is proposed, using static finger gesture recognition based on a specific Computer Vision web-platform. The approach does not take into consideration the stochastic nature of the gestures and this method cannot be applied in the human-robot collaboration. Recently, a new

method for dynamic finger gesture recognition in human computer interaction has been introduced by [81]. This method, based on a low-cost webcam, recognizes the entire finger gesture individually and it is non-obtrusive since it doesn't put any limit on finger motions.

When considering gesture analysis and more specifically fingering analysis in music interaction, there are four main approaches. These are: (a) the pre-processing using score analysis based on an acyclic graph. This approach does not take into consideration all the factors influencing the choice of specific fingering, such as physical and biomechanical constraints [82]; (b) the real-time using midi technology. This approach doesn't concern classical musical instruments [83]; (c) the post-processing using sound analysis that works only when one note is played at a time [84] and (d) the computer vision methods for the guitarist fingering retrieval [80]. The existing Computer Vision (CV) methods are of a low cost but they presuppose painted fingers with a full-extended palm in order to identify the guitarist fingers in the image, and specific recognition platforms, such as EyesWeb. Another great example of fingering recognition is the system of Yoshinari Takegawa, who used colour markers on the fingertips in order to develop a real-time Fingering Detection System for Piano Performance [85]. This system is restricted in electronic keyboards, such as synthesizers and it cannot be applied for classical music instruments neither for the finger gesture recognition and mapping with sounds in space. Moreover, MacRitchie used Vicon System and Vicon Markers modelling in order to visualize musical structures. His method requires the music score in advance [86]. None of the above methods can be extended towards a dynamic gesture recognition taking into consideration the stochastic nature of gestures. They all recognize the musical effect of finger motions on discrete time events.

The study of the above categories for gesture analysis in music interaction can lead to the conclusions that: (a) the gesture measurement approaches are based on rather expensive commercial systems, they are suitable for offline analysis and not for live performances and they cannot be applied on finger gestures; (b) gesture recognition via WSBN or CV does not cost a lot and has many important paradigms of live performance applications. On the other hand, sensors cannot be applied for finger gestures performed on the piano keyboard or on woodwind musical instruments; (c) fingerings can be retrieved with low cost technologies but the information acquired is related to discrete time events without taking into consideration the stochastic nature of the gestures and (d) new paradigms for the recognition of the musician gestures performed on surface or keyboards, with a semi-extended palm, can only be based on CV.

4.3.3.3 Intangible heritage preservation and transmission

Very few attempts at using body and gesture recognition for intangible heritage preservation can be found in the literature. To our knowledge, past attempts for

preserving the ICH of the traditional dances were mainly based on informal interviews of the people practising these dances. The results of these interviews were then summarized in books such as [37]. According to [38], dance has probably been the slowest art form to adopt technology, partially because useful tools have been slow to develop because of the limited commercial opportunities brought by dance applications. In their article they describe applications such as animate and visualize dance, plan choreography, edit and animate notation, enhance performance, but they do not cover intangible performance preservation. However they interestingly underline a recurring issue of such applications, i.e. the need for a unique, unambiguous way to represent human movement, and dance in particular.

In [39], the concept of using motion capture technology is introduced for protecting national dances in China. However their report lacks basic details and information. In [40], the creation of a motion capture database of 183 Jamaican dancers is reported. Their study aimed at evaluating if dance revealed something about the phenotypic or genotypic quality of the dancers, and showed that there are strong positive associations between symmetry (one measure of quality in evolutionary studies) and dancing abilities. However, the aim of this research is not to preserve the dance, but rather to study it, here at a very fundamental level.

For contemporary dance, the DANCERS! project [41] aimed at collecting a database of dancers. The recording setup consisted of a formatted space, videos recorded from the front and top of the scene and metadata describing the dancer. No motion capture was performed, and no precise motion information is hence available, the only possible views of the scene are the ones originally recorded by the videos since the scene was not captured in 3D.

Some research projects have shown that dance-training systems based on motion capture technologies could successfully guide students to improve their dance skills [42] and have evaluated different kinds of augmented feedback modalities (tactile, video, sound) for learning basic dance choreographies.

4.3.4 Encephalography analysis and Emotion Recognition

Emotion Recognition (ER) is the first and one of the most important issues affecting computing (AC) brings forward and plays a dominant role in the effort to incorporate computers, and generally machines, with the ability to interact with humans by expressing cues that postulate and demonstrate emotional intelligence-related attitude. Successful ER enables machines to recognize the affective state of the user and collect emotional data for processing in order to proceed toward the terminus of emotion-based Human Machine Interface, the emotional-like response. Toward effective ER, a large variety of methods and devices have been implemented, mostly concerning ER from face [43,44], speech [45,46] and signals from the autonomous nervous system (ANS), i.e., heart rate and galvanic skin response (GSR) [47,48,49].

A relatively new field in the ER area is the EEG-based ER (EEG-ER), which overcomes some of the fundamental reliability issues that arise with ER from face, voice, or ANS-related signals. For instance, a facial expression recognition approach would be useless for people with the inability to express emotions via face, even if they really feel them, such as patients within the autism spectrum [50], or for situations of human social masking; for example, when smiling though feeling angry. Moreover, voice and ANS signals are vulnerable to "noise" related to activity that does not derive from emotional experience, i.e., GSR signals are highly influenced by inspiration, which may be caused from physical and not emotional activity. On the other hand, signals from the Central Nervous System (CNS), such as EEG, Magneto-encephalogram (MEG), Positron Emission Tomography (PET), or functional Magnetic Resonance Imaging (fMRI), are not influenced by the aforementioned factors as they capture the expression of emotional experience from its origin. Toward such a more reliable ER procedure, EEG appears to be the less intrusive and the one with the best time resolution than the other three (MEG, PET, and fMRI). Motivated by the latter, a number of EEG-ER research efforts have been proposed in the literature.

There are important cultural differences in emotions that can be predicted, understood and connected to each other in the light of cultural expressions. The main cultural differences reflected at the affective space are expressed through initial response tendencies of appraisal, action readiness, expression and instrumental behaviour, but also in regulation strategies. Moreover, the ecologies of emotion and contexts, as well as their mutual reinforcement are different across cultures. By capturing the emotions, and even better their dynamic character using EEG signals during cultural activities, the response selection at the levels of different emotional components, the relative priorities of initial response selection and effortful regulation, the sensitivity to certain context, the plans that are entailed by the emotions, as well as, the likely means to achieve them, could be identified and used as dominant source of information to acquire ICH elements. Consequently, the ways in which the potential of emotions is realized could reveal cultural facets that are intangible in character but form tangible measures at the affective space, contributing to their categorization and preservation, as knowledge-based cultural/emotional models.

Moreover, most folklore/popular culture is shaped by a logic of emotional intensification. It is less interested in making people think than it is in making people feel. Yet that distinction is too simple: folklore/popular culture, at its best, makes people think by making them feel. In this context, the emotions generated by folklore/popular culture are rarely personal; rather, to be traditional or popular, it has to evoke broadly shared feelings. The most emotional moments are often the ones that hit on conflicts, anxieties, fantasies and fears that are central to the culture. In this perspective, folklore/cultural expressions try to use every device their medium offers in order to maximize the emotional response of their audience. Insofar as these folklore/popular artists and performers think about their craft, they are also thinking about how to achieve an emotional impact. By using EEG-based

emotion acquisition of the performers of rare singing, and the corresponding audience, the difference in contexts within these works are produced and consumed could be identified at the affective space, contributing to the exploration of the ways intangible cultural hierarchies respect or dismiss the affective dimensions, operating differently within different folklore cultures.

4.3.5 Semantic multimedia analysis

Semantic multimedia analysis is essentially the process of mapping low-level features to high level concepts, an issue addressed as bridging the "semantic gap" and extracting a set of metadata that can be used to index the multimedia content in a manner coherent with human perception. The challenging aspect of this process derives from the high number of different instantiations exhibited by the vast majority of semantic concepts, which is difficult to capture using a finite number of patterns. If we consider concept detection as the result of a continuous process where the learner interacts with a set of examples and his teacher to gradually develop his system of visual perception, we may identify the following interrelations. The grounding of concepts is primarily achieved through indicative examples that are followed by the description of the teacher (i.e. annotations). Based on these samples the learner uses his senses to build models that are able to ground the annotated concepts, either by relying on the discriminative power of the received stimuli (i.e. discriminative models), or by shaping a model that could potentially generate these stimuli (i.e. generative models). However, these models are typically weak in generalization, at least at their early stages of development. This fact prevents them from successfully recognizing new, un-seen instantiations of the modelled concepts that are likely to differ in form and appearance (i.e. semantic gap). This is where the teacher once again comes into play to provide the learner with a set of logic based rules or probabilistic dependencies that will offer him an additional path to visual perception through inference. These rules and dependencies are essentially filters that can be applied to reduce the uncertainty of the stimuli-based models, or to generate higher forms of knowledge through reasoning. Finally, when this knowledge accumulates over time it takes the form of experience, which is a kind of information that can be sometimes transferred directly from the teacher to the learner and help him to make rough approximations of the required models.

In the cultural heritage domain, multimedia analysis has been extensively used in the past decades as a form of automatic indexing the multimedia cultural content. This necessity grows even more these days considering the popularity of digitizing cultural content for purposes such as safeguarding, capturing, visualizing and presenting both tangible and intangible resources that broadly define that heritage. When it comes to ICH, the task of semantic analysis becomes even more challenging, since the significance of heritage artefacts is implied in their context

and the scope of the preservation extends also to the preservation of the background knowledge that puts these artefacts in proper perspective. These intangible assets may for instance derive from performing arts (e.g. singing, dancing, etc.) and semantic multimedia analysis is essential for mapping the low level features originating from the signal of the utilized sensors (e.g. sound, image, EEG) to important aspects that define the examined art (e.g. singing or dancing style). In the typical case, semantic multimedia analysis consists of the following four components: 1) Pattern recognition, 2) Data fusion, 3) Knowledge-assisted semantic analysis, and 4) Schema alignment. Next, further details are provided for each of the above four cases.

1) Pattern Recognition

In an effort to simulate the human learning techniques, researchers have developed algorithms to teach the machine how to recognize patterns, hence, the name Pattern Recognition, by using annotated examples that relate to the pattern (positive examples) and examples that are not (negative examples). The aim of this procedure is to create a general model that maps the input signals/features to the desired annotations, and, in parallel, generalize from the presented data to future, unseen data.

Pattern recognition techniques have been used in the cultural domain for various cultural heritage categories. In [51], a method that processes historical documents and transforms them to metadata is proposed. In [52], SVM based classification of traditional Indian dance actions using multimedia data is performed. In [53] and [54] computer vision techniques are employed in order to automatically classify archaeological pottery sherds. Lastly, a Computer Vision technique is also used in [55] where the authors present a search engine for retrieving cultural heritage multimedia content.

2) Data Fusion

Fusion [56] is the process of combining the information of multiple sources in order to produce a single outcome. In general, fusion is formulated as the problem of deducing the unknown but common information existing in all sources that lead to the observed data by using all the observations coming from the multiple sources. Thus, fusion can be seen as an inverse problem that can be naturally be formulated in a Bayesian framework [57]. For example, in [58], heterogeneous media sources are combined in the context of Bayesian inference, in order to analyse the semantic meaning. Also, in [59], semantic analysis of audio-visual content is performed, by employing multimodal fusion based on Bayesian models. In [60], Naive Bayesian fusion was used for ancient coin identification. In [61], a Dynamic Bayesian Network (DBN) is employed in order to fuse the audio and visual information of audio-visual content and provide an emotion recognition algorithm.

3) Knowledge-assisted semantic analysis

Research has shown that, in general, expert knowledge can augment the efficiency of the semantic analysis task when applied to a domain. Particularly, in [62], it shown that the accuracy of retrieving cultural ob-jects is increased when

the data are appropriately structured, using knowledge about the objects. In [63], a video semantic content analysis frame-work is proposed, where an ontology is used in combination with the MPEG-7 multimedia metadata standard. In [64], an approach to knowledge-assisted semantic video object detection is presented where Semantic Web technologies are used for knowledge representation. Another example of an ontology framework used in order to facilitate ontology-based mapping of cultural heritage content to corresponding concepts is proposed in [52]. Towards the same direction, the authors of [65] perform ontology-based semantic analysis with a view to link media, contexts, objects, events and people.

An interesting work is that presented in [66], developed in the framework of the DECIPHER project, which proposes a methodology for the description of museum narratives (i.e., the structure of the exhibits). Narratives automate the presentation of the exhibits to the public in a coherent manner and by including the context of the exhibit in which the latter was created and being used.

4) Schema alignment

A vast number of Europe's cultural heritage objects are digitised by a wide range of data providers from the library, museum, archive and audio-visual sectors, and they all use different metadata standards. This heterogeneous data needs to appear in a common context. Thus, given the large variety of existing metadata schemas, ensuring the interoperability across diverse cultural collections is another challenge that has received a lot of research attention.

Europeana data model (EDM), which was developed for the implementation of the Europeana digital library, was designed with the purpose to enforce interoperability between various content providers and the library. EDM transcends metadata standards, without compromising the range and richness of the standards. Also, it facilitates Europeana’s participation in the Semantic Web. Finally, the EDM semantic approach is expected to promote richer resource discovery and improved display of more complex data. It is worth to note that the work in [67] provides a methodology to map semantic analysis results to the EMD metadata schema. In this way, metadata are made available and reusable by end-users and heterogeneous applications.

The PREMIS Data Dictionary for Preservation Metadata is an international standard for metadata that was developed to support the preservation of digital objects/assets and ensure their long-term usability. PREMIS metadata standard has been adopted globally in various projects related to digital preservation. It supports numerous digital preservation software tools and systems. The CIDOC Conceptual Reference Model (CRM), an official standard since 9/12/2006, provides the ability to describe the implicit and explicit relationships of cultural heritage concepts in a formalized manner. Thus, CIDOC CRM is intended to promote a common understanding of cultural heritage information by providing a common and extensible semantic framework that can represent any cultural heritage information. It is intended to be a common language for cultural knowledge domain experts to formulate user requirements for information systems, and thus, facilitat-

ing in this way the interoperability between different sources of cultural heritage information in a semantic level.

Due to the multimodal nature of the content that is to be analysed semantically in i-Treasures, a common metadata schema was designed and implemented for the interoperability between the elementary concept detection and the semantic analysis tasks. More specifically, the results of both of the above tasks are stored in an XML file, with a structure a priori specified. The XML file of the first task is, first, embedded with metadata containing general info (similarly to the EMD metadata schema) and, after the basic concepts are also stored in the file, it is deposited in a central repository (i.e., the i-Treasures web-platform). Next, the file is given as input to the semantic analysis task. The results of this task to be also deposited in the repository by storing an XML with a structure a priori defined. Finally, a user, by using the repository access providing facilities, can conveniently obtain and access the above information.

4.3.6 3D Visualization of Intangible Heritage

Intangible culture is quite different from tangible culture, since intangible culture such as skills, crafts, music, song, drama, and the other recordable culture cannot be simply touched and interacted with or without use of other means. In real life, tangible cultural heritage can be demonstrated in an environment like museums and related exhibitions. A cultural heritage structure which is totally destroyed such as a temple can be even reproduced as a replica, so that audience can personally wander inside. On the other hand, due to its non-physical nature, ICH is more restricted and hard to demonstrate in real life which is a real challenge to prevent it from disappearing. This is where 3D visualization and interaction technology comes into play.

Thanks to the recent advances in computer graphics, it is now possible to visualize almost anything, either tangible [68] or intangible. What can be done is limited only by imagination and allows reaching new larger audiences via the Internet. It is certain that 3D visualization and interaction will hardly be on par with the real thing and instruction system in a computer application and simulation cannot possibly match a real life master's tutoring. Obviously, the degree of reality in interaction, visualization and physics simulation becomes a very important concern for users to become well accustomed to the culture and encourage others to do so.

ICT technologies are increasingly becoming one of the pillars of Cultural Heritage Education [69][70]. Virtual worlds are often used in the field of Cultural Heritage education in order to broaden the opportunity to appreciate cultural contents that are remote in space and/or time. Even though they should be considered very helpful for widening access to cultural contents, these applications, for example Virtual Museums, often are not intrinsically engaging and, sometimes, fail

in supporting active learning, just giving the opportunity to access information [71].

Digital games support learning in a more active and engaging way and, from the pedagogical viewpoint, they offer advanced interaction, such as the possibility of customizing the learning paths and of keeping track of the learners' behaviour and successes/failures and are more adaptive to meet the specific users' learning needs.

As to the digital games available in the Cultural Heritage (CH) area, Anderson et al. [72] and afterwards Mortara et al. [71] carried out interesting state-of-the-art reviews. While the first focuses more on technical aspects, the second sketches a panorama of the actual use of SGs in CH education. According to [71] in the field of CH, SGs of different kind are adopted: from trivia, puzzle and mini-games to mobile applications for museums or touristic visits, (e.g. Muse-US[3], Tidy City[4]), to simulations (e.g. the battle of Waterloo[5]), to adventures and role playing games (the Priory Undercroft[6], Revolution[7]).

As it could be expected, games are more widespread in the Tangible Cultural Heritage (TCH) area, where several different examples can be found [73]. An example is Thiatro[8], a 3D virtual environment, where the player acts as a museum curator, or a curator of other digital artefacts, My Culture Quest[9], which aims at advertising real collections or even the History of a Place[10], which is integral part of a museum experience at the Archaeological Museum of Messenia in Greece.

---

[3] Coenen T.(2013). MuseUs: case study of a pervasive cultural heritage serious game. Journal on Computing and Cultural Heritage (JOCCH), 6(2), 8:2-8:19

[4] http://totem.fit.fraunhofer.de/tidycity The game consists in solving riddles about a specific city, which might require the player to explore places never seen before while learning about the city's cultural heritage

[5] http://www.bbc.co.uk/history/british/empire_seapower/launch_gms_battle_waterloo.shtml, a strategy game reconstructing the famous battle.

[6] A. Doulamis et al. (2011). Serious games for cultural applications. In D. Plemenos, G. Miaoulis (Eds.), Artificial Intelligence Techniques for Computer Graphics, Springer (2011). The game is a reconstruction of the Benedictine monastery in Coventry, dissolved by Henry VIII.

[7] Francis R. (2006). Revolution, learning about history through situated role play in a virtual environment. Proc. of the American educational research association conference. The game is a role-playing game in the town of colonial Williamsburg during the American Revolution;

[8] http://www.thiatro.info/

[9] http://www.mylearning.org/interactive.asp?journeyid=238&resourceid=587

[10] http://www.makebelieve.gr/mb/www/en/portfolio/museums-culture/54-amm.html

A number of games for smartphones also exist, like Tate Trumps[11] and YouTell[12], which, for instance, allow museum visitors to create and share through smart phones their own media and stories.

Many games also exist in the area of historical reconstruction; for instance, the Battle of Thermopylae[13] or the Playing History[14], which are mainly based on 3D technology, to closely recreate the environment in which each event happened.

Although to a lesser extent, a number of promising games have been developed in the field of ICH [29]. Some examples are:

- Icura[15], a 3D realistic environment to teach about Japanese culture and etiquette, which can raise cultural interest and support a real pre-trip planning.
- Discover Babylon[16], Roma Nova[17] and Remembering 7th Street[18], that are aimed at raising awareness about ancient Mesopotamia's contribution to modern culture, ancient Rome and West Oakland in the time period post-World-War-II.
- Africa Trail[19] and Real Lives 2010[20] simulate a 12,000 mile travel by bicycle through Africa or a different life in any country of the world (e.g., a peasant farmer in Bangladesh, or a computer operator in Poland), respectively.
- Papakwaqa[21], a serious game about the Atayal minority in Taiwan, particularly focused on ICH assets like tribal beliefs, customs, and ceremonies.

Game-like applications are also a powerful tool for ICH transmission, following a well consolidated trend in the Technology Enhanced Learning field, which promotes the adoption of digital games to sustain learning and training in a variety of educational fields aiming to empower constructive, experiential, self-regulated learning and increase the user's engagement and motivation.

---

[11] http://www.hideandseek.net/tate-trumps/

[12] Cao, Y.et al (2011). The Hero's Journey – template-based storytelling for ubiquitous multimedia management. Journal Multimedia, 6 (2) 156–169.

[13] Christopoulos, D. et al (2011). Using virtual environments to tell the story: The battle of Thermopylae. Proceedings of VS-Games 2011.

[14] http://www.playinghistory.eu

[15] Froschauer, J., et. Al, (2010). "Design and evaluation of a serious game for immersive cultural training". Proceedings of the 16th International Conference on Virtual Systems and Multimedia (VSMM) 253–260.

[16] http://www.fas.org/babylon/

[17] http://www.seriousgamesinstitute.co.uk/applied-research/Roma-Nova.aspx

[18] http://7thstreet.org/

[19] http://www.mobygames.com/game/africa-trail

[20] http://www.educationalsimulations.com/products.html

[21] Huang, C. & Huang, Y. (2013). Annales school-based serious game creation framework for Taiwan indigenous cultural heritage. Journal of Computing in Cultural Heritage, 6 (2)

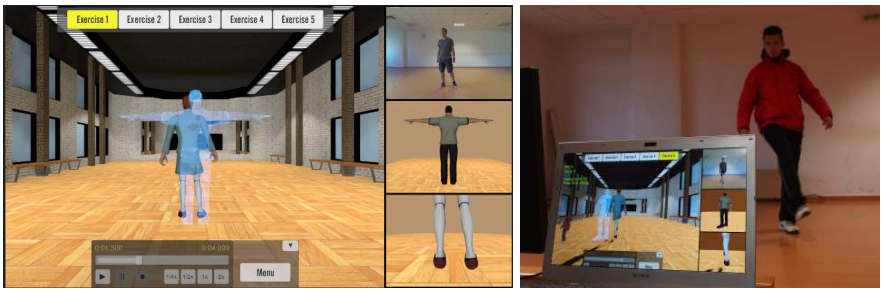

Fig. 4.3: Screenshot from the game-like application for Tsamiko dance. It presents the ghost avatar implementation, which allows the user to see his/her own character 'superposed' with the expert's one so as to be able to visually detect where s/he is making a mistake (© 2016 EC FP7 i-Treasures project, Reprinted with Permission)

For instance, within i-Treasures project, seven prototype educational game-like applications for sensorimotor learning have been implemented for selected ICH sub-use cases (Tsamiko dance, Walloon dance, Calus dance, Human Beat Box singing, Byzantine music, pottery, and contemporary music composition). These games (Fig. 4.3) are designed to get input from various sensors and game devices, such as the prototype hyper-helmet or off-the-shelf commercial sensors like Kinect. The system allows the user to a) observe the expert performance and b) practice by trying to reproduce the expert performance and then getting an evaluation and additional feedback about his/her performance.

4.3.7 Text-to-song synthesis

Singing Voice Synthesis (SVS) is a branch of the text-to-speech (TTS) technology that deals with generating a synthetic interpretation of a song, given its text and musical score. Synthesis of singing voice has been a research area for a long time and for each decade a major technology improvement has been achieved.

The beginnings of the synthetic voice in musical representations in the artistic domain date back to the beginning of the 1980s, with the 'Chant' project developed at IRCAM, and used by composers of contemporary classical music[74]. 'Chant' is based on the synthesis by formants (by rules), like other well-known systems, as those developed at KTH in Stockholm [75], or CCRM Stanford [76]. A state of the art description can be found in [77] or in [78]. Such systems were capable of synthesizing realistically vocal vowels (vocals), at the price of a big studio work to analyse and adjust the settings of the systems.

For singing, as for speech, the 1990's are marked by the generalization of concatenative synthesis, driven by the impressive increase in size of the speech corpora. An alternative approach to singing synthesis that found its way in professional

quality productions is based on voice conversion. The first example was the merger of two voices, a male alto and a soprano voice from a woman to ’create’ the voice of a castrate (male soprano) in the film ”Farinelli”, examples can be found on youtube and in [89]. The voice was capable of high quality articulate speech; however it was not synthesis but just voice transformation.

The 2000's were marked by the appearance of Vocaloid [79]. It is the first software for singing voice synthesis with articulate lyrics which had a very important mainstream development. Vocaloid is marketed by Yamaha since 2003. During this period, concatenative synthesis is being generalized, and statistical parametric synthesis systems appeared for speech and more marginally for singing.

Recently, there is a revival of interest for singing voice synthesis. The request is coming from both the composers, the audio-visual and public games industries. Recent innovations include real-time control speech synthesis, both in the sphere of mobile orchestras (project Chorus Digitalis, Vox Tactum, meta-orchestra, laptop orchestra), the use of various sensors in conjunction with a synthesizer.

4.4 Discussion and future challenges

From the above, it becomes obvious that the advanced technological methods discussed earlier have not been thoroughly applied in the preservation or transmission of intangible heritage. Yet, previous projects and studies would suggest that there are several opportunities to employ the specific technologies in cultural heritage preservation with a particular focus on transmission and education. Although technology cannot replace human interaction, there is significant scope to develop activities that on the one hand document and preserve the knowledge of rare songs, dances, composition and craftsmanship, but also ensure the transmission of this knowledge to younger generations. In this sense, technology can help sustain the knowledge of the past and enable its transmission to future generations. For this reason, it is important to ensure the active involvement of cultural practitioners in all related research and development efforts. Rather than a threat, new technologies constitute a great opportunity for the documentation and dissemination of intangible heritage. Instead of only focusing on the documentation of intangible heritage, the combination of the technologies discussed above can create a novel approach for the safeguarding of intangible heritage that is primarily focused on education, training and pedagogical interaction.

In the case of rare traditional singing, the combination of these technologies can contribute significantly not only to the documentation of the knowledge of singing, but also to its dissemination to a broader audience. Facial expression analysis technologies could therefore enable the detailed recording of the singers' expression and singing technique and EEG analysis could provide information about the performers' emotional state. Vocal tract sensing technologies could be

used to document the various changes of the vocal tract and motion capture technologies could give an indication of the performers' body movements. Text to Song technology could provide sophisticated software for the creation of an educational tool to be used in educational scenarios and a 'performative tool' that can be used as a traditional musical instrument.

With respect to the case of dance, motion capture technologies can provide a detailed representation of the movement of the human body in performance bringing new insights to motional and gestural aspects whose examination is not always possible due to complex outfits and costumes. Recent studies [90] report promising results on recognizing predefined dance motion patterns from skeletal animation data captured by multiple Kinect sensors. However, the recognition of different dance styles and analysis of more complex dance patterns and variations is still a challenging problem for the future. EEG analysis and facial expression analysis can contribute to the examination of the emotional state of the dancers when performing. Moreover, 3D visualisation can enable the representation of dance movements in a 3D and sensorimotor learning context.

Regarding the case of craftsmanship and pottery, motion capture can be used again for the detailed documentation of hand and finger movement during the creation process. As discussed earlier, 3D visualization can provide educational opportunities for virtual learning scenarios.

Concerning contemporary music composition, technologies such as motion capture and EEG could potentially provide combined information on finger movement and emotional condition. In addition, facial expression analysis could give insight as to how the creative process is mirrored in the composer's face.

Finally, in all use cases, multimodal semantic analysis enables, first, the combination of different levels of information and data for documentation, and second, the detection of high level concepts in the data in an automatic manner. The former is very useful for the system integration, since the problem of combining heterogeneous modalities disappears, while, the latter is useful for the detection of high level concepts automatically and without the need for intervention by an expert. As a subsequent result, using the semantic analysis results, we can have a detailed and with useful information documentation in order to provide a convenient access to the content.

As the world becomes increasingly dependent on digital resources, there is an important opportunity to develop a platform that enhances the transmission of traditional knowledge and skills by using current advances in the field of digital technologies. Platforms such as i-Treasures offer services for knowledge exchange between researchers and for the transmission of rare ICH know-how from LHTs to apprentices, acting as a means for stimulating creative and game industry and education as well as promoting local cultural tourism. A big challenge is to find the optimal technologies to capture, analysis, presentation and re-use of ICH, which typically contains a huge wealth of multimodal information and corresponds to a rich knowledge domain. In the future, further advances in technologies for digitization (i.e. audio, visual and motion capture), e-documentation (3D modelling en-

riched with multimedia metadata and ontologies), e-preservation (standards), visualization (virtual/augmented reality and gamification technologies) and re-use (e.g. applications for research and application) of ICH are expected to exploit the full potential of ICH and offer multiple benefits to the different stakeholders involved. So technology is no longer a threat to the survival of customs and traditions, but a tool for their sustained development in an increasingly global 21st century.

Acknowledgments The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2011-9) under grant agreement no FP7-ICT-600676 "i-Treasures: Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures".

References

1. Cameron, F. and Kenderdine, S. (2010) Theorizing Digital Cultural Heritage: A Critical Dis-course. Cambridge, Mass, London: MIT Press
2. Ioannides, M., Fellner, D. Georgopoulos, A. and Hadjimitsis, D. (ed.) (2010), Digital Heritage, Third International Conference, Euromed 2010, Lemnos, Cyprus, Proceedings., Berlin, Heidelberg, New York: Springer.
3. Aikawa, N. (2004) An Historical Overview of the Preparation of the UNESCO International Convention for the Safeguarding of Intangible Heritage, Museum International, vol. 56, pp. 137-149.
4. Hafstein, V. (2009) Intangible Heritage as List: From Masterpieces to Representation, in Smith L. and Akagawa N. (ed.) Intangible Heritage, Abingdon: Routledge, pp. 93-111.
5. Nas, P. (2002) Masterpieces of Oral and Intangible Heritage: Reflections on the UNESCO World Heritage List, Current Anthropology, vol. 43, no. 1, pp. 139-143.
6. Alivizatou, M. (2007) The UNESCO Programme for the Proclamation of Masterpieces of the Oral and Intangible Heritage of Humanity: A Critical Examination, Journal of Museum Ethnography, vol. 19, pp. 34-42.
7. Bolton, L. (2003) Unfolding the Moon: Enacting Women's Kastom in Vanuatu, Honolulu, Hawai'i: University of Hawai'i Press.
8. Huffman, K. (1996) The Fieldworkers of the Vanuatu Cultural Centre and their Contribution to the Audiovisual Collections in Bonnemaison, J., Huffman, K. and Tryon, D. (ed.) Arts of Vanuatu. Honolulu: University of Hawai'i Press, pp. 290-293.
9. Zafeiriou, S. and Yin, L. (2012) 3D facial behaviour analysis and understanding, Image and Vision Computing, vol. 30, October, pp. 681-682.
10. Ekman, P., Levenson R.and Friesen W. (1983) Emotions Differ in Autonomic Nervous System Activity, Science, vol. 221, pp. 1208-1210.
11. Engwall, O. (1999) Modeling of the vocal tract in three dimensions, in Proceedings, Eu-rospeech99, Hungary, Budapest, pp. 113116.
12. Fels, S., Lloyd J.E., Van Den Doel K., Vogt F., Stavness I. and Vatikiotis-Bateson E. (1991) Developing physically-based, dynamic vocal tract models using Artisynth , Proceedings of ISSP 6, pp.419-426,
13. Stone, M. (1991) Toward a model of three-dimensional tongue movement, Phonetics, vol. 19, pp. 309320.

14. Badin, P., Bailly, G., Reveret, L., Baciu, M., Segebarth, C. and Savariaux, C. (2002) Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images, Journal of Phonetics, vol. 30, no. 3, pp. 533-553.
15. Stone, M. (1990) A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data, The Journal of the Acoustical Society of America,vol.87, pp. 2207.
16. Engwall, O. (2004) From real-time MRI to 3D tongue movements, in Proceedings, 8th International Conference on Spoken Language Processing (ICSLP), Jeju Island, Korea, vol. 2, pp. 1109-1112.
17. Stone, M. and Lundberg, A. (1996) Three-dimensional tongue surface shapes of English con-sonants and vowels, Journal of the Acoustical Society of America, vol.99, no.6, pp. 37283737.
18. Henrich N., Lortat-Jacob B., Castellengo M., Bailly L. and Pelorson X. (2006), Period-doubling occurences in singing: the ”bassu” case in traditional Sardinian ”A Tenore” singing, Proceedings of the International Conference on Voice Physiology and Biomechanics, Tokyo, Japan, Jul. 2006.
19. Henrich N., Bailly L., Pelorson X., Lortat-Jacob B. (2009) Physiological and physical understanding of singing voice practices: the Sardinian Bassu case, AIRS Start-up meeting, Prince Edward Island, Canada.
20. Cho, W., Hong, J., Park, H. (2012) Real-Time Ultrasonographic Assessment of True Vocal Fold Length in Professional Singers, Journal of Voice, vol. 26, no.6, pp. 1-6.
21. Troup, G., Griffiths, T., Schneider-Kolsky, M. and Finlayson, T. (2006) Ultrasound Observa-tion of Vowel Tongue Shapes in Trained Singers, Proceedings of the 30 th Condensed Matter and Materials Meeting, Wagga, Australia.
22. Caon, M. (2011) Context-Aware 3D Gesture Interaction Based on Multiple Kinects, Proceedings of the First International Conference on Ambient Computing, Applications, Services and Technologies, Barcelona, Spain, pp.7-12.
23. Coduys, T., Henry, C. and Cont, A. (2004) TOASTER and KROONDE: high-resolution and high-speed real-time sensor interfaces, Proceedings of the Conference on New Interfaces for Musical Expression, Singapore, pp. 205-206.
24. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guedy, F. and Rasamimanana, N. (2010). Gesture in embodied communication and human-computer interaction, 8th International Gesture Workshop , pp. 73-84.
25. Boucher, M. (2011) Virtual Dance and Motion-Capture, Contemporary Aesthetics, vol. 9
26. Aylward, R. and Paradiso, J. A. (2006) Sensemble: A Wireless, Compact, Multi-User Sensor System for Interactive Dance, in Proceedings of the International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, Centre Pompidou, pp. 134-139.
27. Drobny, D., Weiss, M.and Borchers, J. (2009) Saltate!: a sensor-based system to support dance beginners, Extended abstracts on Human factors in Computing Systems, Proceedings of the CHI 09 International Conference, New York: ACM, pp. 3943-3948
28. Bevilacqua, F., Naugle, L. and Dobrian, C. (2001) Music control from 3D motion capture of dance. CHI 2001 for the NIME workshop.
29. Dobrian, C. and Bevilacqua, F. (2003) Gestural control of music: using the vicon 8 motion capture system. Proceedings of the Conference on New Interfaces for Musical Expression (NIME), National University of Singapore, pp. 161-163.
30. Raptis, M., Kirovski, D., and Hoppe, H. (2011) Real-time classification of dance gestures from skeleton animation, Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation, New York, NY, USA, pp. 147-156.
31. Alexiadis, D. S., Kelly, P., Daras, P., O’Connor, N. E., Boubekeur, T. and Moussa, M. B. (2011). Evaluating a dancer’s performance using kinect-based skeleton tracking,

Proceedings of the 19th ACM international conference on Multimedia, ACM, New York, USA,pp. 659- 662.

32. Essid, S., Alexiadis, D. S., Tournemenne, R., Gowing, M., Kelly, P., Monaghan, D. S., et al. (2012) An advanced virtual dance performance evaluator, Proceedings of the 37th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, Japan, pp. 2269-2272.
33. Alankus, G., Bayazit, A. A. and Bayazit, O. B. (2005) Automated motion synthesis for dancing characters: Motion Capture and Retrieval, Comput. Animat. Virtual Worlds, vol. 16, no. 3-4, pp. 259-271.
34. Bouchard, D. and Badler, N. (2007) Semantic segmentation of motion capture using laban movement analysis,Intelligent Virtual Agents , Springer, pp. 37-44.
35. Kahol, K., Tripathi, P., and Panchanathan, S. (2004) Automated gesture segmentation from dance sequences, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR04), Seoul, Korea, pp. 883-888
36. James, J., Ingalls, T., Qian, G., Olsen, L., Whiteley, D., Wong, S., et al. (2006) Movement-based interactive dance performance, Proceedings of the 14th annual ACM International Conference on Multimedia, New York: ACM, pp. 470-480
37. Malempre, M. (2010) Pour une poignee de danses, Dapo Hainaut(Ed.).
38. Calvert, T., Wilke, W., Ryman, R. and Fox, I. (2005) Applications of computers to dance. Computer Graphics and Applications, IEEE, vol. 25, no. 2, pp. 6-12.
39. Shen, Y., Wu, X., Lua, C. and Cheng, H. (2012) National Dances Protection Based on Motion Capture Technology, Chengdu, Sichuan, China, Singapore: IACSIT Press, vol. 51, pp. 78-81.
40. Brown, W. M., Cronk, L., Grochow, K., Jacobson, A., Liu, C. K., Popovic, Z., et al. (2005) Dance reveals symmetry especially in young men, Nature , vol.438, no. 7071, pp. 1148-1150.
41. Tardieu, D., Siebert, X., Mazzarino, B., Chessini, R., Dubois, J., Dupont, S., Varni, G. and Visentin, A. (2010) Browsing a dance video collection: dance analysis and interface design, Journal on Multimodal User Interfaces , vol. 4, no.1, pp. 37-46.
42. Chan, J. C., Leung, H., Tang, J. K. and Komura, T. (2011) A virtual reality dance training system using motion capture technology, Learning Technologies, IEEE Transactions, vol 4, no. 2, pp. 187-195.
43. Cohen, I., Garg A. and Huang T. (2000) Emotion Recognition from Facial Expression Using Multilevel HMM, Proceedings of the Neural Information Processing Systems Workshop on Affective Computing, Breckenridge.
44. Bourel, F., Chibelushi, C. and Low, A. (2002) Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localized Facial Dynamics, Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Washington, USA.
45. Schuller, B., Reiter, S. Mueller, R. Hames, A. and Rigoll, G. (2005) Speaker Independent Speech Emotion Recognition by Ensemble Classification, Proceedings of the IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, pp. 864-867.
46. Busso, C., Deng, Z., Yildirim, S, Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U. and Narayanan, S. (2004) Analysis of Emotional Recognition Using Facial Expressions, Speech and Multimodal Information, Proceedings of the International Conference on Multimodal Interfaces, ACM, New York, pp. 205-211
47. Picard, R., Vyzas, E. and Healey, J. (2001) Toward Machine Emotional Intelligence: Analysis of Affective Physiological State, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175-1191.
48. Nasoz, F., Lisetti, C. Alvarez K. and Finkelstein, N. (2003) Emotion Recognition from Physiological Signals for User Modeling of Affect, In Proceedings of the International Conference on User Modeling, Johnstown, PA, USA.

49. Lisetti, C., and Nasoz F. (2004) Using Non-invasive Wearable Computers to Recognize Human Emotions from Physiological Signals, EURASIP Journal on Applied Signal Processing, vol. 11, pp. 1672-1687.
50. McIntosh, D., Reichmann-Decker, A. Winkielman, P. and Wilbarger, J. (2006) When the Social Mirror Breaks: Deficits in Automatic, But Not Voluntary, Mimicry of Emotional Facial Expressions in Autism, Developmental Science, vol. 9, pp. 295-302.
51. Esposito, F., Malerba D., Semeraro G., Altamura O., Ferilli S., Basile T., Berard M. and Ceci M. (2004) 'Machine learning methods for automatically processing historical documents: from paper acquisition to XML transformation', Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL, 04), Palo Alto, CA, USA, pp. 328-335.
52. Mallik, A., Chaudhuri, S. and Ghosh, H. (2011) 'Nrityakosha: Preserving the Intangible Heritage of Indian Classical Dance', ACM Journal on Computing and Cultural Heritage, vol. 4, no. 3, pp. Article 11.
53. Makridis, M., & Daras, P. (2012). Automatic classification of archaeological pottery sherds. Journal on Computing and Cultural Heritage (JOCCH), 5(4), 15.
54. Karasik, A. (2010) 'A complete, automatic procedure for pottery documentation and analysis', Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops (CVPRW),San Francisco, CA, USA, pp. 29-34.
55. Vrochidis, S., Doulaverakis, C. Gounaris, A. Nidelkou, E. Makris, L. and Kompatsiaris, I. (2008) 'A Hybrid Ontology and Visual-based Retrieval Model for Cultural Heritage Multimedia Collections', International Journal of Metadata, Semantics and Ontologies, vol. 3, no. 3, pp. 167-182.
56. Liggins, M., Hall D. L. and Llina J. (2008) Handbook of Multisensor Data Fusion, Theory and Practice (2nd Edition), CRC Press.
57. Punska, O. (1999) 'Bayesian approach to multisensor data fusion', PhD. Dissertation, Department of Engineering, University of Cambridge.
58. Nikolopoulos, S., Lakka, C. Kompatsiaris, I. Varytimidis, C. Rapantzikos K. and Avrithis, Y.(2009) 'Compound document analysis by fusing evidence across media', Proceedings of the International Workshop on Content-Based Multimedia Indexing, Chania, Crete, pp. 175-180.
59. Chang, S., Ellis, D., Jiang, W., Lee, K., Yanagawa, A., Loui, A. C. and Luo, J. (2007) 'Largescale multimodal semantic concept detection for consumer video', Proceedings of the international workshop on Workshop on multimedia information retrieval (MIR '07), September, Germany, pp. 255-264.
60. Huber-Mörk, R., Zambanini, S., Zaharieva M.and Kampel, M. (2011) 'Identification of ancient coins based on fusion of shape and local features', Machine Vision and Applications, vol. 22, no. 6, pp. 983-994.
61. Datcu, D. and Rothkrantz, L. J. M. (2008), Semantic Audio-Visual Data Fusion for Automatic Emotion Recognition, in Euromedia, Porto.
62. Koolen, M. and Kamps J.(2010) 'Searching cultural heritage data: Does structure help expert searchers?', Proceedings of RIAO '10 Adaptivity, Personalization and Fusion of Heterogeneous Information, Paris, France, pp. 152-155
63. Bai, L., Lao,. S., Zhang,W., Jones, G. J. F. and Smeaton, A. F. (2007) 'Video Semantic, Content Analysis Framework Based on Ontology Combined MPEG-7', Adaptive Multimedia Retrieval: Retrieval, User, and Semantics, Lecture Notes in Computer Science, July, pp. 237-250.
64. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V. K. and Strintzis, G. M. (2005), 'Knowledge-Assisted Semantic Video Object Detection, IEEE Transactions on Circuits and Systems for Video Technology', Special Issue on Analysis and Understanding for Video Adaptation, vol. 15, no. 10, pp. 1210-1224.

65. Lien, J., Kanade T., Cohn J.and Li C. (1998) 'Automated Facial Expression Recognition Based on FACS Action Units', Proceedings of the Third IEEE Conference on Automatic Face and Gesture Recognition, Nara, pp.390-395 .
66. Mulholland, P., Wolff, A. Collins T. and Zdrahal, Z. (2011) 'An event-based approach to describing and understanding museum narratives', Proceedings: Detection, Representation, and Exploitation of Events in the Semantic Web Workshop in conjunction with theInternational Semantic Web Conference, Bonn, Germany.
67. Kollia, I., Tzouvaras, V. Drosopoulos N. and . Stamou G. (2012) 'A Systemic Approach for Effective Semantic Access to Cultural Content' Semantic Web – Interoperability, Usability, Applicability, vol. 3, no.1, pp.65-83
68. Gaitatzes, A., Christopoulos, D., & Roussou, M. (2001, November). Reviving the past: cultural heritage meets virtual reality. In Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage (pp. 103-110). ACM.
69. Ott, M., & Pozzi, F. (2011). Towards a new era for Cultural Heritage Education: Discussing the role of ICT. Computers in Human Behavior, 27(4), 1365-1371. J.R. Savery, T.M. Duffy, Problem-based learning: An instructional model and its constructivist framework. Educational Technology 35 (1995) 31–38.
70. Veltman K.H., (2005), Challenges for ICT/UCT Applications in Cultural Heritage. In: Carreras, C. (ed.) ICT and Heritage, online at http://www.uoc.edu/digithum/7/dt/eng/dossier.pdf.
71. Mortara, M., Catalano, C.E., Bellotti, F., Fiucci, G., Houry-Panchetti, M. & Petridis P. (2014). Learning Cultural Heritage by serious games. Journal of Cultural Heritage, 15(3), 318–325.
72. Anderson, E. F. McLoughlin L., Liarokapis F., Peters, C., Petridis, P., de Freitas, S. (2009). Serious Games in Cultural Heritage. M. Ashley and F. Liarokapis (Editors). Proceedings of the The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST - State of the Art Reports.
73. Ott, M., & Pozzi, F. (2008). ICT and Cultural Heritage Education: Which Added Value?. In Lytras et al (Eds). Emerging Technologies and Information Systems for the Knowledge Society. Lecture Notes in Computer Science, 5288 (pp.131-138). Springer Berlin Heidelberg.
74. Rodet, X., Potard, Y. and Barriere, J.-B. (1984) The CHANT project: from the synthesis of the singing voice to synthesis in general, Computer Music Journal, vol. 8, no.3, pp.15-31.
75. Berndtsson, G. (1996) The KTH rule system for singing synthesi, Computer Music Journal, vol. 20, no. 1, pp. 7691.
76. Cook, P. (1992) Physical Models for Music Synthesis, and a Meta-Controller for Real Time Performance, Proceedings of the International Computer Music Conference and Festival, Delphi, Greece
77. Cook, P. (1996) Singing voice synthesis: History, current work, and future directions, Computer Music Journal, vol. 20, no. 3, pp.3846.
78. Bennett, G. and Rodet, X. (1989) Synthesis of the singing voice, in Mathews, M.V. and. Pierce, J.R, (ed.) Current directions in computer music research, Cambridge, MA, USA: MIT Press, pp.19-44.
79. Kenmochi, H. and Ohshita, H. (2007) Vocaloid–commercial singing synthesizer based on sample concatenation. Presented at Interspeech 2007, Antwerp, Belgium, pp.4009-40010
80. Burns, A.-M. and Wanderley, M. M. (2006) Visual methods for the retrieval of guitarist fingering, Proceedings of the Conference on New interfaces for musical expression, Pompidou: IRCAM-Centre, pp. 196-199.
81. Vision par ordinateur pour la reconnaissance des gestes musicaux des doigts, Revue Francophone d'Informatique Musicale [Online] Available at: http://revues.mshparisnord.org/rfim/index.php?id=107 [13 July 2013].

82. Grunberg, D. (n.d.) Gesture Recognition for Conducting Computer Music. [On line] Available at: http://schubert.ece.drexel.edu/research/gestureRecognition [10 January 2009]
83. Verner, J. (1995) MIDI guitar synthesis yesterday, today and tomorrow, an overview of the whole fingerpicking thing, Recording Magazine , vol. 8, no.9, pp. 52-57.
84. Traube, C. (2004) An interdisciplinary study of the timbre of the classical guitar, PhD Thesis. McGill University.
85. Takegawa, Y., Terada, T., and Nishio, S. (2006) Design and Implementation of a Real-time Fingering Detection System for Piano Performances, Proceedings of the International Computer Music Conference, New Orleans, USA, pp. 67-74.
86. MacRitchie, J., Buck, B., and Bailey, N. (2009) Visualising musical structure through performance gesture, Proceedings of the International Society for Music Information Retrieval Conference, Kobe, Japan, pp.237-242.
87. K. Dimitropoulos, S. Manitsaris, F. Tsalakanidou, S. Nikolopoulos, B. Denby, S. Al Kork, L. Crevier-Buchman, C. Pillot-Loiseau, S. Dupont, J. Tilmanne, M. Ott, M. Alivizatou, E. Yilmaz, L. Hadjileontiadis, V. Charisis, O. Deroo, A. Manitsaris, I. Kompatsiaris, and N. Grammalidis, 'Capturing the Intangible: An Introduction to the i-Treasures Project', in Proc. 9th International Conference on Computer Vision Theory and Applications (VISAPP2014), Lisbon, Portugal, 5-8 January 2014.
88. M. Brand and A. Hertzmann. Style machines. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000), pages 183–192. ACM Press, 2000.
89. Bennett, G. and Rodet, X. (1989) 'Synthesis of the singing voice', in Mathews, M.V. and. Pierce, J.R, (ed.) Current directions in computer music research, Cambridge, MA, USA: MIT Press, pp.19-44.
90. Kitsikidis, A., Dimitropoulos, K., Douka, S., and Grammalidis, N. (2014, January). Dance analysis using multiple kinect sensors. In Computer Vision Theory and Applications (VISAPP), 2014 International Conference on (Vol. 2, pp. 789-795). IEEE.

Figure captions

Fig. 4.1 Overview of i-Treasures system
Fig. 4.2 Overview of main capture technologies for ICH
Fig. 4.3 Screenshot from the game-like application for Tsamiko dance. It presents the ghost avatar implementation, which allows the user to see his/her own character 'superposed' with the expert's one so as to be able to visually detect where s/he is making a mistake