

A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures

Dimitropoulos K.; Manitsaris S.; Tsalakanidou F.; Denby B.; Buchman L.; Dupont S.; Nikolopoulos S.; Kompatsiaris Y.; Charisis V.; Hadjileontiadis L.; Pozzi F.; Cotescu M.; Ciftci S.; Katos A.; Manitsaris A.; Grammalidis N.

ABSTRACT

Intangible Cultural Heritage (ICH) creations include, amongst other, music, dance, singing, theatre, human skills, and craftsmanship. These cultural expressions are usually transmitted orally and/or using gestures and are modified over a period of time, through a process of collective recreation. As the world becomes more interconnected and many different cultures come into contact, local communities run the risk of losing important elements of their ICH, while young people find it difficult to maintain the connection with the cultural heritage treasured by their elders. In this paper, we present a novel holistic approach for the safeguarding and transmission of ICH that goes beyond the mere digitization of ICH content. Based on multisensory technology for the capturing of ICH, the proposed approach enables the generation of completely novel cultural content. High-level semantics are extracted from the acquired data, enabling researchers to identify possible implicit or hidden correlations between different ICH expressions or interpretation styles and study the evolution of a specific ICH. These data, coupled with other cultural resources, are accessible through the i-Treasures Web-platform, which provides the means for supporting knowledge exchange between researchers as well as know-how transmission from ICH bearers to apprentices.

KEYWORDS

Intangible cultural heritage, multimodal analysis, multi-sensor technology, semantic analysis

INTRODUCTION

Cultural expression is not only limited to tangible objects but also includes intangible expressions, such as practices, representations, knowledge, and skills. Their importance lies not only in their value as cultural manifestations, but also in the associated wealth of knowledge, which is transmitted through them from one generation to the next. For this reason, UNESCO introduced the term "Living Human Treasures" for persons who possess to a high degree the knowledge and skills required for performing or recreating specific elements of intangible cultural heritage (ICH).

ICH creations are transmitted orally and/or by gestures and usually undergo modifications over a period of time, resulting to a process of collective recreation. As the world becomes more interconnected, many communities start losing important elements of their ICH, while younger generations find it more difficult to maintain their connection with the cultural heritage treasured by their elders. In the 2003 Convention for the Safeguarding of ICH,[1] UNESCO defined safeguarding as measures aimed at ensuring the viability of intangible heritage. Although their role in safeguarding ICH is not directly addressed in the Convention, new technologies can significantly contribute towards this goal. For instance, modern Information and Communication Technologies (ICTs) provide powerful tools for capturing cultural information from digital multimedia resources, augmenting it with additional metadata and preserving it into digital repositories.

Towards this end, EU released Europeana (<http://www.europeana.eu/>), a digital platform for exploring the digital resources of Europe's museums, libraries, archives, and audio-visual collections. Moreover, a growing number of ICH inventories are digital and are accessible via data management platforms, while many ICH safeguarding projects are driven and managed by local communities.[2] Most approaches primarily focus on archival and their potential for transmission and education relies on how they can be further exploited by local stakeholders, such as museums and cultural centers. Moreover, a number of serious games have been developed in the past, most for history education or for enhancing museum visits, including a few for ICH transmission.[3]

In this paper, an alternative approach is presented that goes beyond the mere digitization of ICH content, offering new knowledge extracted from analyzing ICH in an innovative way: the i-Treasures system supports ICH safeguarding and transmission through a novel ICT-based methodology for ICH capture and analysis in domains where human motion is extremely important, such as performing arts and handicrafts. Several ICH expressions from different European regions were studied broadly classified into four categories: a) traditional singing styles, b) traditional dances, c) craftsmanship, and d) contemporary music composition. In the following, we present the different technologies employed by the i-Treasures system and evaluate the system's efficiency in terms of user satisfaction and technical performance. Hands-on experiments with a large number of users, both ICH practitioners and ICT experts, have shown the great potential of the proposed system.

I-TREASURES SYSTEM

The i-Treasures system is an attempt to explore the challenges and opportunities that emerge when we consider the safeguarding of ICH from a technology perspective. Our main goal is to: a) extract novel information from ICH using a multisensory approach and b) provide easy access to various ICH resources for research work and education to local communities, groups, and individuals interested in the safeguarding of ICH. The core of i-Treasures lies in the identification of specific features or patterns (e.g., postures, audio patterns, etc.) in different ICH expressions using multisensor technology (e.g., cameras, mocap systems, microphones, acquisition interfaces for electroencephalography, etc.). Subsequently, data fusion analysis is applied to exploit information across the different modalities. Moreover, context and content are integrated for mapping the set of low- or medium-level 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 Web platform, which offers different types of content (e.g., text, audio, images, video) and supports different roles: expert, learner, and researcher. Furthermore, based on the latest advances in Web-based game engines, a learning environment has also been developed to support training and evaluation by means of sensorimotor learning (see Figure 1).

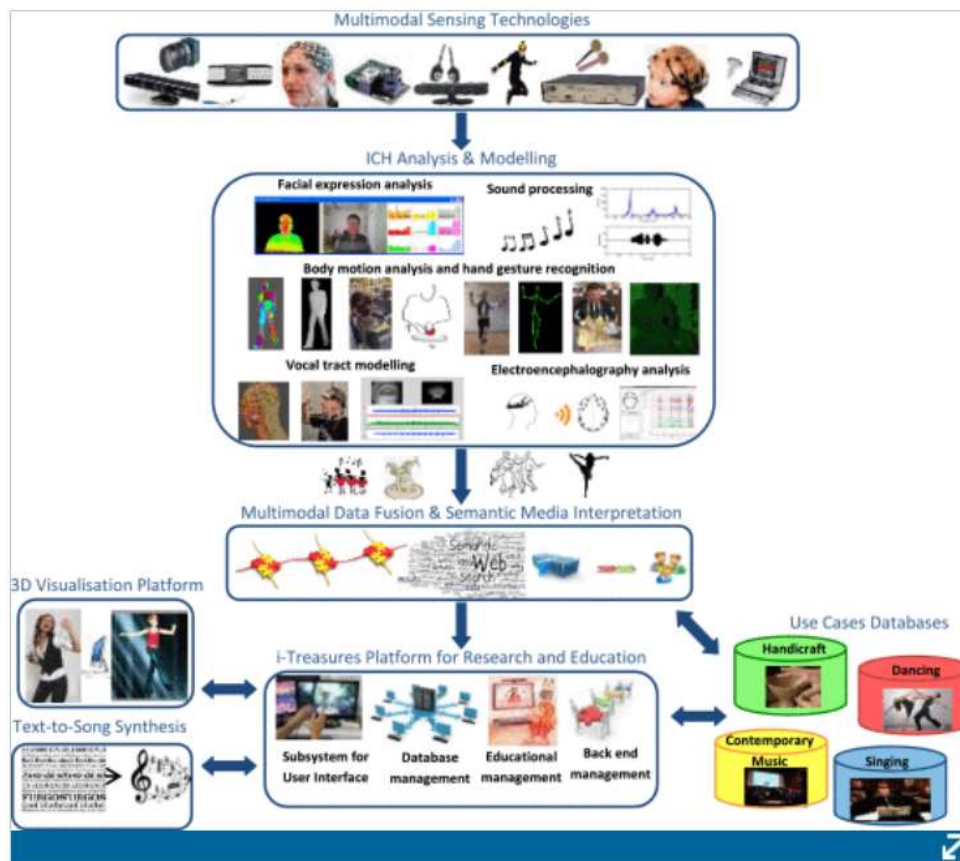


Figure 1. i-Treasures system architecture.

In a typical application scenario, such as the case of a traditional dance, an expert can use the system to a) capture a dance performance and store medium-level features (e.g., dance steps) and high-level concepts (e.g., dance style) to the database of the system and b) create educational courses for the learning of the recorded dance. More specifically, different mocap technologies are supported, e.g., Kinect sensors, optical motion capture systems or inertial sensors. The recorded motion is then analyzed, through probabilistic inference, for the extraction of a number of medium-level features, related to the motion of the dancer (e.g., in the case of the traditional dance of Tsamiko the identification of the “single” or “double” step is important for the recognition of the variation of the dance) or high level concepts, e.g., the style or the variation of the dance. Furthermore, the expert can also design educational courses using the pedagogical planner module or create new three-dimensional (3-D) games for dance learning and practicing, using a novel framework that enables the rapid design and development of customizable body-motion-based game-like applications.

On the other hand, a learner or researcher can a) have access to the content of the platform and search for a specific dance in the repository of the system and b) use the interactive learning applications of the i-Treasure platform. In the former, the user can retrieve general information about a specific dance (text, images, and videos) as well as its medium-level features and high-level concepts. Moreover, the user can search for specific medium-level features (e.g., dance figures) to find possible similarities of the dance with other dances. In the latter, the user is able to participate

in LMS courses and practice by playing a 3-D educational game using a Kinect sensor. Each game is designed following a specific learning scenario, defined by an expert, and supports a mechanism for the online evaluation of the player's performance.

System Architecture

An international group of experts and performers, related to the selected ICH expressions, was actively involved in the definition of user requirements, which drove the design of the system. The system architecture is modular and supports different applications and tools. It comprises of five main components: i) the data capture and analysis module, which supports a wide variety of sensors and state-of-the-art technologies in motion, emotion and sound capture, analysis, and recognition, ii) the data fusion and semantic analysis module, which extracts medium and high-level features, iii) the 3-D visualization platform for sensorimotor learning, iv) the Text-to-Song synthesis module, which provides an interface that enables the user to enter text and/or notes and the system to produce the equivalent singing voice supplementing for the absence of a teacher, and v) the integrated Web-platform for ICH research and education.

CAPTURING THE INTANGIBLE

i-Treasures focuses on the capture and analysis of ICH performances with the use of multisensing technologies including optical, depth, inertial, electroglottography, electroencephalography, and ultrasound sensors. Based on the in-depth analysis of the various ICH expressions, as well as discussions with ICH experts, several techniques have been developed for the analysis of ICH performance elements. These include modeling and recognition of facial muscle movements, body movements, hand and finger gestures, brain activity, vocal tract operation (including tongue and lips movement), acoustic speech, and music sounds.

Means of Acquisition

Multisensor calibration and synchronization protocols have also been designed, to allow synchronous recording of different ICH elements (e.g., body movements and singing voice) and novel sensor fusion techniques have been implemented, which combine data captured by different sensors in a common augmented structure (e.g., fusion of data from multiple depth and inertial sensors into a single full-body skeleton), as shown in Figure 2(a) and (b). A prototype light-weight hyper-helmet embedding several sensors for vocal tract capture (i.e., video camera, ultrasound probe, microphone, EGG sensor, piezoelectric accelerometer, and respiratory belt) has been designed [Figure 2(c)].

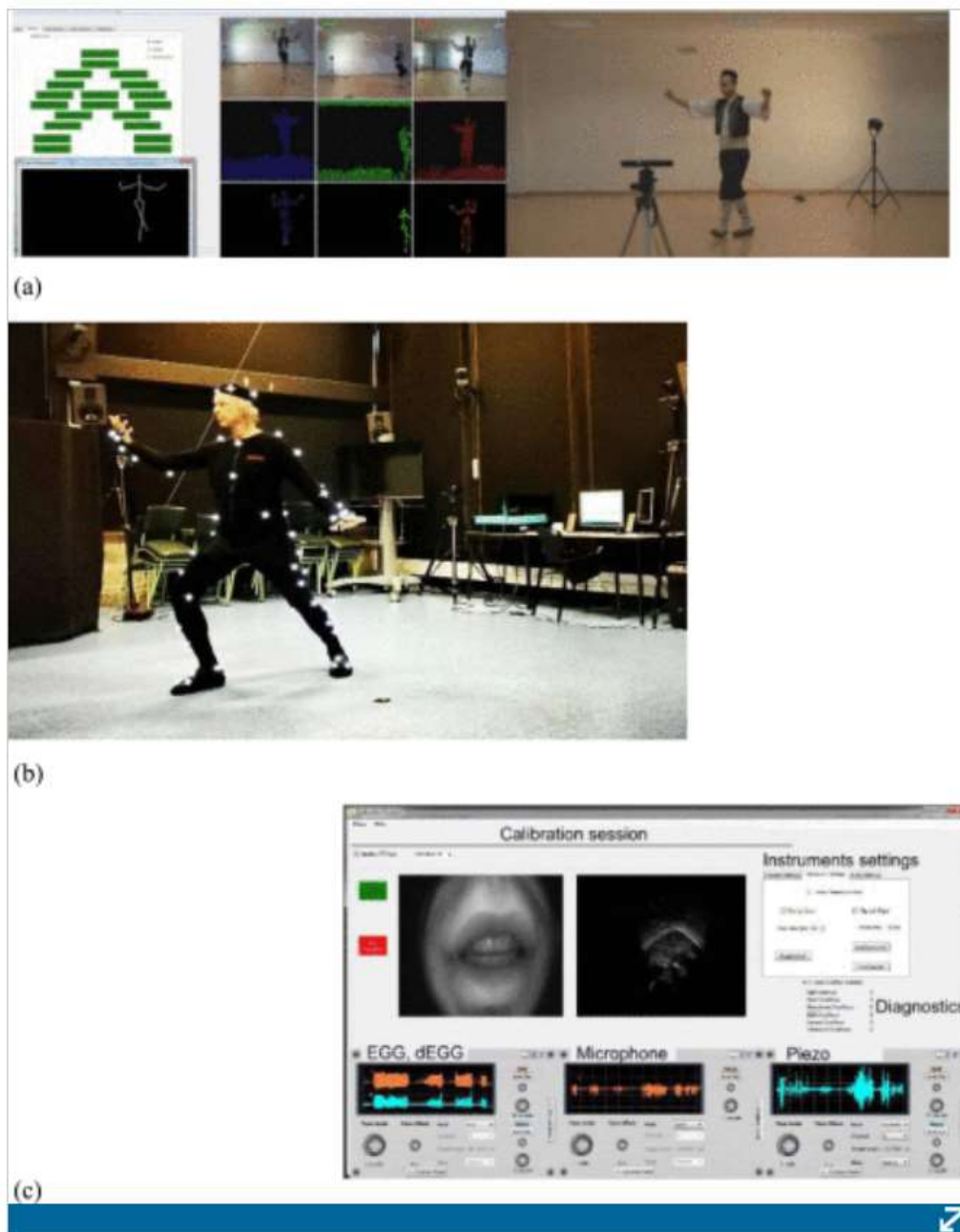


Figure 2. Data acquisition using: (a) multiple depth sensors (a fused human skeleton is produced by combining data from all sensors), (b) a marker-based mocap system, and (c) a prototype light-weight hyper-helmet for vocal tract signal capturing.

Means of Analysis

Several algorithms have been developed in the context of i-Treasures; they are used for the analysis of ICH performances and the extraction of medium-level features (e.g., dance steps/figure types, gesture types, voice/sound features, emotional status), which are subsequently used for the extraction of high-level metadata.⁴⁻⁶ Moreover, a cross-platform API (called

Post-print: K. Dimitropoulos *et al.*, "A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures," in *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 3-16, Nov.-Dec. 2018, doi: 10.1109/MIS.2018.111144858. © IEEE
<https://ieeexplore.ieee.org/document/8255779/authors#authors>

MotionMachine)[7] that enables the rapid prototyping, extraction, and visualization of motion features has been recently publicly released (<https://github.com/numediart/MotionMachine>). Finally, the MyoWebToolkit (<http://augreal.mklab.iti.gr/MyoWebToolkit/>) has been developed for the purpose of recognizing and understanding finger gestures in the pottery use case based on electromyography signals.

SEMANTIC ANALYSIS

For the representation of the ICH domain knowledge corresponding to different forms of ICH, a set of ontologies was developed. The latter is based on a multistage approach, which builds upon user requirements and discussions between ICH domain experts and technology developers. Novel knowledge representation models and inference algorithms were investigated, in order to model the inherent uncertainty that is prevalent in the domain of ICH. The goal of these models is to encode the multimodal nature of the ICH content and combine the medium-level features that have been extracted (e.g., visual and auditory characteristics of a dance or gestural and emotional features of a musical performance). The ability to combine these features allowed extracting high-level metadata, such as different styles and variations of the same ICH expression, or even assessing the proficiency level of a performer and evaluating the synchronization across different modalities based on the notion of rhythm. Towards this end, the multientity Bayesian networks were employed.[8]

For the stylistic analysis and comparison between different ICH expressions, various medium-level features can be used. For instance, in the case of two traditional dances, Tsamiko and Walloon, we extracted: i) relational features: feet crossed, feet apart sideways, ii) effort features: weight effort and iii) ergonomics: expansion, deviation, and balance. In general, the stylistic analysis showed that i) Walloon dancer crosses the feet more than the Tsamiko dancer, i.e., the “feet apart sideways” is higher for Tsamiko dance, ii) the weight effort during performances in both dances is not high, iii) in both cases the muscles of dancers are crouched and dancers do not adopt an expanded pose, and iii) Walloon dance is less balanced than the Tsamiko dance. Similar studies between other forms of ICH can be easily performed using the extracted medium level features, enabling the researchers to identify possible correlations between different ICH expressions.

Due to the multimodal nature of the data, a common metadata schema was designed and implemented for the interoperability between medium-level features and high-level concepts. The result of medium-level features extraction is an XML file containing: i) general information (similarly to the Europeana data model metadata schema) and, ii) the medium-level features. This file is then deposited to a central repository (i.e., the i-Treasures Web-platform), where semantic analysis that produces high-level concepts (e.g., ICH style) is performed. The final result is a new XML file, also deposited in the central repository, from which a user can conveniently obtain and access all this information.

I-TREASURES WEB PLATFORM

The i-Treasures platform constitutes the gateway through which access to intangible treasures is provided. It is developed based on an open-source CMS enriched with a significant number of functionalities, contributing to the preservation, and widespread diffusion of ICH treasures. The platform covers different aspects of ICH, offering search of metadata, as well as structured and formal courses, game-like applications, and other functionalities (such as the Text-to-Song synthesis module) that can support teaching and learning processes.

© 20XX IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Post-print: K. Dimitropoulos *et al.*, "A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures," in *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 3-16, Nov.-Dec. 2018, doi: 10.1109/MIS.2018.111144858. © IEEE
<https://ieeexplore.ieee.org/document/8255779/authors#authors>

A novel methodology for ICH education has been developed that is based on three essential steps: i) learning the “basics,” i.e., the corpus of theoretical knowledge that will allow the learner to appreciate and understand the ICH in its intrinsic and cultural value; ii) exploring the ICH, i.e., being exposed to a number of expert performances and learning to appreciate/detect the main features of the ICH itself and the different styles/variations; iii) immersing in the ICH, i.e., having the chance to try to perform, having the possibility to get feedback in such a way as to detect one's mistakes and improve the quality of the performance.

Exemplar “educational scenarios,” inspired by the described methodology, were prepared for the different ICH expressions. These scenarios were conceived and authored thanks to an ad-hoc tool (the Pedagogical Planner) and were implemented into courses, which are delivered to users through the learning management system (LMS) of the platform.

ICH TRANSMISSION THROUGH SERIOUS GAMES

A novel 3-D visualization module for sensorimotor learning has also been released, aiming to provide support for practicing different types of ICH expressions. The sensorimotor learning module was endowed with game-like applications, 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. Seven prototype educational games have been implemented for selected ICH expressions (Greek Tsamiko dance, Belgian Walloon dance, Romanian Calus dance, Human Beat Box singing, Byzantine music, pottery, and contemporary music composition). The games (Figure 3) are presented in the framework of a single application and are designed to get input from various sensors, from the prototype hyper-helmet to off-the-shelf commercial sensors like Kinect.[9]



Figure 3. Game-like applications for ICH sensorimotor learning: (a) pottery, (b) contemporary music composition through gestures and emotion, (c) Beat-box singing, and (d) Tsamiko dance.

Creating games from scratch usually demands significant effort and time. Therefore, we have decided to focus on the development of a novel framework for rapid design of body-motion-based customizable game-like applications. This framework consists of two components: i) a game design module that allows the user to easily define a simple game scenario and activities, capture the required motion data (i.e., the expert performance that will be used for the animation of the expert avatar and against which the learner will be evaluated) and select several game parameters (virtual environment assets, evaluation algorithm, etc.) and ii) a customizable game-like application for motor skills learning and practicing, which is automatically configured based on the xml output of the game design module.

EVALUATING I-TREASURES IN PRACTICE

The i-Treasures system was evaluated both in terms of user satisfaction/system usability and technical performance. The evaluation results are presented below.

Usability Evaluation

For assessing the usability of the proposed system and the acceptance of users, the system was demonstrated to a variety of users, covering a wide range of the population, including students of all ages, experienced ICH practitioners and experts (Figure 4), ICH apprentices, and general public in different European cities and villages. The activities followed different demonstration scenarios

ranging from structured learning paths for school students to online courses for ICH beginners, to applications for museum visitors and to demonstrations for the general public during science or cultural festivals. The demonstration events were organized in a variety of locations, including public schools, dance schools, pottery workshops, museums, festivals, as well as open days and dedicated workshops. In total, 26 different activities were organized in the context of 10 subuse cases, which involved more than 650 users of all ages and with different backgrounds. During these activities, the users (acting as learners or ICH experts/teachers) interacted with the various parts of the i-Treasures system in real conditions, e.g., they used the Web-platform and its components (LMS, games, etc.) to teach/learn a traditional dance in the context of a class taught in school, and provided feedback by means of structured questionnaires but also through interviews/discussions.

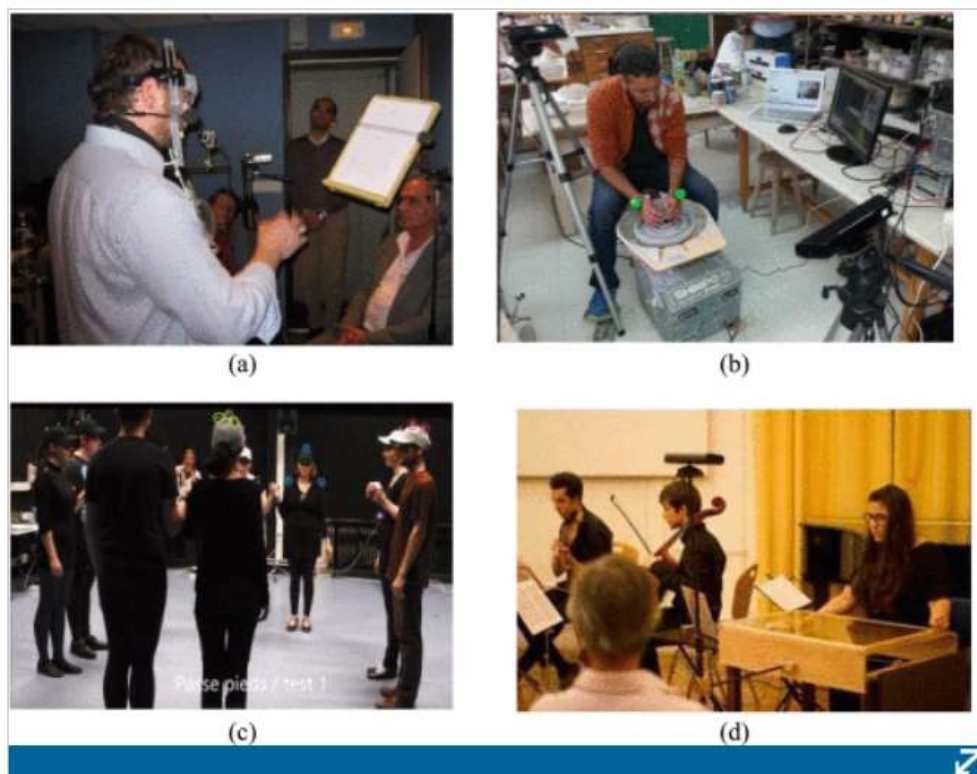


Figure 4. Experts from different ICH domains using the i-Treasures technology: a) singing, b) pottery, c) dancing, and d) music composition.

In total, 127 questionnaires were collected, which allowed us to quantitatively evaluate the usability of the proposed system (in many cases, especially for scenarios involving demonstration to the general public at festivals or workshops, users were happy to share how they felt about this experience in face-to-face discussions or mini-interviews but were reluctant to fill in questionnaires). The questionnaires, which included Likert-scale (1–5) questions, were developed based on operational models proposed by ICH experts and covered the evaluation of informational, interactional, and performance characteristics of the system components. The participants evaluated the effectiveness ($m = 3.93$), efficiency ($m = 3.97$), satisfaction ($m = 3.95$) and innovation ($m = 4.18$) of the proposed technology with an average mean value $mav=4.01/5$.

In addition to the quantitative (questionnaire-based) assessment of system usability and user acceptance, we were also able to perform a more qualitative assessment based on discussions and

mini-interviews with users during the aforementioned activities. The users were, in general, satisfied by the developed technologies and thought that they constitute an innovative and promising set of tools that can help and advance ICH analysis, education, and research. They pointed out that by taking advantage of recent advances in ICT and especially multisensing technology, the i-Treasures system is able to offer a deeper understanding of the intrinsic characteristics of ICH. Moreover, the use of novel educational tools such as the game-like applications can boost ICH education in ways that make it more appealing and attractive and enjoyable, especially to young people.

Technical Assessment

The technical assessment of the i-Treasures system included: i) the overall assessment of the system, based on nonfunctional requirements, and ii) the overall assessment of the main technical indices (main system modules), based on functional requirements. Based on the work in the paper by Dagnino *et al.*, [10] two questionnaires were prepared, which were filled by 108 experts with advanced knowledge in ICT and, also, expertise in research domains related to i-Treasures components (e.g., human motion analysis, ontology engineering, game development, etc.) The questionnaires included Likert-scale (1–5) questions classified in several main categories and their analysis was based on descriptive statistics and confirmatory factor analysis.

Table 1 presents the results for the various categories included in the questionnaire for the overall system assessment. The results are quite satisfactory with an average mean value $mav=3.77/5$ and standard deviation $std = 0.78$. Table 2 summarizes the results of the assessment of the main components/different technologies of the i-Treasures system. As it can be seen, the experts were in general satisfied by the developed technologies ($mav=3.89/5$, $std = 0.73$). Some improvements, however, are still needed, e.g., in the content of the platform or the prototype hyperhelmet for vocal tract capturing (which, in some cases, can cause discomfort to the singers, possibly affecting their performance).

Table 1. Overall assessment.

Categories	Mean	Std	Cronbach's Alpha
Cost optimality	3.38	0.67	0.677
Accessibility / Usability	3.63	0.80	0.922
Documentation / Support	3.92	0.66	0.565
Interoperability / Portability	3.74	0.61	0.612
Extensibility / Scalability	3.68	0.69	0.842
Auditing	3.56	1.08	0.839
Security / Privacy	4.05	0.64	0.708
Fault tolerance / Recoverability	3.58	0.97	0.854
Licensing / Copyright	4.16	0.87	0.814

Table 2. System's technologies.

Categories	Mean	Std	Cronbach's Alpha
Facial expression analysis	3.91	0.73	0.939
Human body motion and gesture recognition	4.01	0.60	0.874
Electroencephalography analysis	4.15	0.69	0.935
Vocal tract sensing – Ultrasound analysis	3.79	0.91	0.862
Sound processing	3.89	0.62	0.927
Text-to-song	3.95	0.70	0.782
Semantic Analysis	3.89	0.73	0.960
Platform	3.52	0.87	0.891

CONCLUSION

In this paper, a novel holistic approach for the safeguarding and transmission of ICH was presented, relying on the use of multisensory technology for the generation of a completely novel cultural content. Since existing digital cultural heritage platforms/repositories (e.g., Europeana) currently contain very limited ICH content (audio/video) and metadata, the proposed system can possibly fill this gap, by capturing and analyzing ICH content in new ways and by supporting new applications for ICH research and education. Experiments with a number of users, both experts and learners, from various ICH domains showed the great potential of ICT technologies in capturing, analysing, and modeling ICH data. Additional pre and post-tests, designed by each teacher, could better determine whether the users actually learnt an ICH and improve their knowledge and skills. Future

Post-print: K. Dimitropoulos *et al.*, "A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures," in *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 3-16, Nov.-Dec. 2018, doi: 10.1109/MIS.2018.111144858. © IEEE
<https://ieeexplore.ieee.org/document/8255779/authors#authors>

advances on sensor technologies are expected to further boost the use of ICT in the preservation of a plethora of ICH domains.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's under Grant Agreement FP7-ICT-600676, i-Treasures.

REFERENCES

1. V. Hafstein, "Intangible heritage as list: From masterpieces to representation," *Intangible Heritage*. Abingdon, U.K.: Routledge, 2009, pp. 93–111.
2. M. Alivizatou-Barakou *et al.*, "Intangible cultural heritage and new technologies: Challenges and opportunities for cultural preservation and development," *Mixed Reality and Gamification for Cultural Heritage*. New York, NY, USA: Springer, 2017.
3. M. Mortara, C. E. Catalano, F. Bellotti, G. Fiucci, M. Houry-Panchetti, and P. Petridis, "Learning cultural heritage by serious games," *J. Cultural Heritage*, vol. 15, no. 3, pp. 318–325, 2014.
4. K. Dimitropoulos, P. Barmpoutis, A. Kitsikidis, and N. Grammalidis, "Classification of multidimensional time-evolving data using histograms of grassmannian points," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 4, pp. 892–905, Apr. 2018. doi: [10.1109/TCSVT.2016.2631719](https://doi.org/10.1109/TCSVT.2016.2631719).
5. K. Xu, Y. Yang, C. Leboullenger, P. Roussel, and B. Denby, "Contour-based 3D tongue motion visualization using the ultrasound image sequences," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 5380–5384.
6. B. Picart, S. Brognaux, and S. Dupont, "Analysis and automatic recognition of human beatbox sounds: A comparative study," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 4255–4259.
7. J. Tilmanne and N. d'Alessandro, "MotionMachine: A new framework for motion capture signal feature prototyping," in *Proc. Eur. Signal Process. Conf.*, Aug./Sep. 2015, pp. 2401–2405.
8. G. Chantas *et al.*, "Multi-entity Bayesian networks for knowledge-driven analysis of ICH content," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 355–369.
9. A. Kitsikidis *et al.*, "A game-like application for dance learning using a natural human computer interface," in *Proc. HCI Int.*, Aug. 2015, pp. 472–482.
10. M. Dagnino, L. J. Hadjileontiadis, M. Ott, and F. Pozzi, "An integrated platform supporting intangible cultural heritage learning and transmission: Definition of requirements and evaluation criteria," *J. Comput. Inf. Technol.*, vol. 22, no. 4, pp. 277–292, 2014.