# Atti

– 32 –

Universitat Politècnica de Catalunya
Barcelona - Spain

# AXMEDIS 2007

# Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution

Barcelona, Spain 28th – 30th November 2007

*Edited by*

Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini

*Sponsored by*

Universitat Politècnica de Catalunya
All AXMEDIS Project partners
IST FP6, European Commission

Firenze University Press
2007

AXMEDIS 2007 : proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution: Barcelona, Spain, 28-30 November 2006 / edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini. – Firenze : Firenze university press, 2007.
(Atti; 32)
http://digital.casalini.it/9788884536785

ISBN 13: 978-88-8453-678-5 (online)
ISBN 13: 978-88-8453-677-8 (print)

658.514 (ed. 20)
Produzione - Automazione



Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
http://epress.unifi.it/

*Printed in Italy*

*AXMEDIS 2007*
www.axmedis.org/axmedis2007

*Universitat Politècnica de Catalunya*
*Barcelona, Spain*
*28th – 30th November 2007*

*Sponsored by*

ACIT - Advance Concepts for Interactive Technology GmbH
(Germany)

Associazione dei Fonografici Italiani
(Italy)

Accademia Nazionale di Santa Cecilia, Fondazione
(Italy)

British Broadcasting Corporation
(United Kingdom)

Second Major Department: Dipartimento di Italianistica, Università degli Studi di Firenze
(Italy)

Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, DISIT Lab.
(Italy)

ELION Enterprises Ltd
(Estonia)

École Polytechnique Fédérale de Lausanne
(Switzerland)

Electronics and Telecommunication Research Institute (Republic of Korea)

EUTELSAT S.A.
(France)

EXITECH S.r.l.
(Italy)

Fraunhofer Gesellschaft zur Foerderung der Angewandten Forschung E.V.
(Germany)

Hewlett Packard Italiana S.r.l.
(Italy)

GIUNTI Interactive Labs S.r.l.
(Italy)

Kaunas University of Technology
(Lithuania)

MBIGROUP
(Italy)

Peking University
(China)

Sociedad Digital de Autores y Editores
(Spain)

Sejer Representing Bordas and Nathan
(France)

spin-off of Consorzio Pisa Ricerche, CPR
(Italy)

TEO LT, AB (formerly Lithuanian Telecom)
(Lithuania)

Telecom Italia
(Italy)

TISCALI Services
(Italy)

ICSRiM, University of Leeds
(United Kingdom)

Universitat Politècnica de Catalunya
(Spain)

University of Reading
(United Kingdom)

Video Reklamos Studija (Video Advertising Studio)
(Lithuania)

XIM Ltd
(United Kingdom)

# Table of Contents




*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5

# Message from the Chairs

In the digital age, progress and development in science and technology greatly impact our work and life at all levels. In response to the explosion of cross-media contents and distributions, the AXMEDIS FP6 IST project is co-supported by the European Commission to develop an inclusive framework to empower the growth and expansion of this domain both in terms of research as well as large scale industrial applications. The AXMEDIS 2007 International Conference seeks to promote discussion and interaction between researchers, practitioners, developers and users of tools, technology transfer experts, and project managers. In line with the AXMEDIS conference series, the AXMEDIS 2007 brings together a variety of participants from the academic, business and industrial worlds, to address different technical and commercial issues. Particular interests include the exchange of concepts, prototypes, research ideas, industrial experiences and other results. The conference focuses on the challenges in the cross-media domain, including production, protection, management, representation, formats, aggregation, workflow, distribution, business and transaction models. Additionally it explores the integration of new forms of content and content management systems and distribution chains, with particular emphasis on the reduction of costs and innovative solutions for complex cross-domain issues and multi-channel distribution.

The AXMEDIS conference has brought together the experiences and communities coming from the WEDELMUSIC conference series, the MUSICNETWORK and other co-located workshops. Together, the AXMEDIS conference has widened the scope and enlarges the communities to share and cross-fertilise new developments and latest innovations.

The first AXMEDIS International Conference was held in Florence, Italy, in 2005 with over 230 attendees from 22 countries, with 48% from research and academic sectors, 37 % from the industry, 7,4% from government, 4% from cultural institutions, etc. The event included 2 collocated Workshops, 2 panels, and 4 Tutorials. Last year, the Conference was held in Leeds, UK (13-15 Dec 2006 with a pre-conference tutorial day on the 12th Dec 2006), with similar amount of delegates from 25 countries, with 57% from research and academic, 34% from the industries, 5% from government and 4% from cultural institutions. AXMEDIS2006 also hosted 4 co-located workshops, 2 panel, 4 tutorials and 4 keynotes.

This year, the program committee has received a relevant number of submissions for research and applications, industrial panels and workshops. The selection has not been easy due to the amount of high quality submissions and the limited time slots of the conference. The technical programme produced is very dense with high quality presentations, including a large number of scientific and industrial presentations, industrial panels, workshops and tutorials. Same as previous years, the AXMEDIS2007 conference has produced two volumes of proceedings. This is the second volume of the proceedings and it contains selected submissions for workshops, industrial panels and additional papers.

We are very grateful to many people without whom this conference would not be possible. Thanks to old and new friends, collaborators, institutions, organisations, and the European Commission, who has supported AXMEDIS. A special thank to all the sponsors and supporters including the EC IST FP6. Thanks to members of the International Program Committee for their invaluable contributions and insightful work. Thanks to Florence University Press for the organisation of this proceedings. Last but not least, many thanks to the many people behind the scene and to all participants of AXMEDIS 2007. We look forward to welcoming you to Barcelona and wish you an exciting, enjoyable, excellent conference.

**General Chair:** Jaime Delgado, *Universitat Politècnica de Catalunya, Barcelona, Spain*
**Programme Co-Chairs:** Pierfrancesco Bellini, *DSI-DISIT, University of Florence, Italy*
Paolo Nesi, *DSI-DISIT, University of Florence, Italy*
Kia Ng, ICSRiM - University of Leeds, Leeds, UK

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

## Conference Organizing Committee

**General chair**

Jaime Delgado
*DMAG – Dept. Arquitectura de Computadors*
*Universitat Politècnica de Catalunya, Spain*
*jaime.delgado@ac.upc.edu, http://research.ac.upc.edu/dmag/*

**Research co-chairs**

Kia Ng
*ICSRiM - University of Leeds, UK*
*kia@computer.org, www.kcng.org*

Pierfrancesco Bellini
*DSI-DISIT, University of Florence, Italy*
*pbellini@dsi.unifi.it*

**Tutorial co-chairs**

Paolo Nesi
*DSI-DISIT, University of Florence, Italy*
*nesi@dsi.unifi.it, www.dsi.unifi.it/~nesi*

David Fuschi
*Giunti Interactive Labs, Italy*
*d.fuschi@giuntilabs.it*

**Industrial co-chairs**

Domenico Dato
*Tiscali, S.p.A, Italy*
*ddato@tiscali.com*

**Workshops co-chairs**

Nigel Earnshaw
*BBC, UK*
*nigel.earnshaw@rd.bbc.co.uk*

Silvia Llorente
*Universitat Politècnica de Catalunya, Spain*
*silviall@ac.upc.edu*

**Local Committee co-chairs:**

Víctor Torres
*Universitat Pompeu Fabra, Spain*
*victor.torres@upf.edu*

Anna Carreras, Lidia Castillejo, Xavier Maroñas
*Universitat Politècnica de Catalunya, Spain*
*annac@ac.upc.edu*
*lidac@ac.upc.edu*
*xmaronas@ac.upc.edu*

## International Programme Committee

P. Bellini – University of Florence (IT)
I. Burnett – University of Wollongong (AU)
M. Campanai – Exitech (IT)
D. Crombie - FNB (NL)
R. Dannenberg - Carnegie Mellon University (US)
D. Dato - Tiscali (IT)
J. Delgado - Universitat Politècnica de Catalunya (ES)
J. Domingo-Ferrer – Universitat Rovira i Virgili (ES)
C. Dorai - IBM T.J. Watson Research Center (US)
N. Earnshaw - BBC (UK)
D. Fuschi - Giunti Interactive Labs (IT)
M. Gauvin - SDAE (ES)
S. Gradmann – University of Hamburg (DE)
R. Grimm – University of Koblenz-Landau (DE)
K. Hirata - NTT (JP)
T. Huang - Peking University (CN)
P. Kafno – University of Nottingham & HDThames (UK)
T. Kalker - HP (US)
Z. Liu - AT&T Labs Research (US)
S. Llorente - Universitat Politècnica de Catalunya (ES)
R. Martí - Universitat Autònoma de Barcelona (ES)
A. Matellanes - Motorola Labs (UK)
H. Mayer - Joanneum Research (AT)
M. Merabti - Liverpool John Mores University (UK)
P. Nesi – University of Florence (IT)
K. Ng – University of Leeds (UK)
J. Nuetzel - 4FriendsOnly.com (DE)
L. Pearce - XIM (UK)
E. Peig - Universitat Pompeu Fabra (ES)
E. Rodríguez - Universitat Politècnica de Catalunya (ES)
C. Serrao - ISCTE/DCTI/Adetti (P)
Y. Song – University of Michigan at Dearborn (US)
C. Timmerer – University of Klagenfurt (AT)
A. Waller - Thales Research (UK)

# AXMEDIS 2007

# Industrial and Application Session

*Coordinated by*

Domenico Dato, TISCALI Services, Italy

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

# Implementation on a Real-time SVC Encoder for Mobile Broadcasting

Sangjin Hahm, ByungSun Kim, Changseob Park
*Technical Research Institute*
*Korean Broadcasting System*
*Seoul, Korea*
*cashy@kbs.co.kr*

Munchurl Kim
*Multimedia Computing,Communications and broadcasting Lab*
*Information and Communication University*
*Deajeon, Korea*

## Abstract

*Scalable Video Coding (SVC) can be applicable in mobile broadcasting environment due to the flexibility of spatial, temporal and quality scalability. Recently, SVC technology becomes mature rapidly but its reference SW encoder isn't optimized yet. Therefore, we have developed a real-time SW SVC encoder for broadcasting. In this paper, we show our SVC encoder that can provide two spatial layers: QVGA(320x240) and VGA(640x480). The base layer can be fully compatible with H.264/AVC. Our encoder is performing real-time operation on a normal PC by optimizing SVC algorithm*

***Keywords-component; SVC, H.264/AVC***

## 1. Introduction

Advanced video coding (AVC) is used in mobile broadcasting for high compression rate and good video quality. In Korea, Digital Multimedia Broadcasting (DMB) also adopted AVC. The video resolution for mobile broadcasting is usually low because the mobile terminal such as cellular phone has small display for mobility and low power consumption. Recently, demands for higher video resolution in mobile broadcasting environment are increasing. This requirement leads to the use of scalable video coding which can provide both low and high display resolution together.

Now, the standard reference software of SVC called Joint Scalable Video Model (JSVM)[1] isn't implemented efficiently just only to verify SVC tools in the perspective of standard conformance. It is far from real-time encoding. So, we designed a SVC encoder only with spatial scalability for the real-time application of mobile broadcasting. Our SVC encoder meets the requirements of real-time implementation and acceptable performance with tolerable PSNR value drop and bitstream increment by optimizing SVC algorithm.

## 2. SVC encoder

*A. SVC*

SVC is a scalable extension of H.264/AVC being developed by JVT(Joint Video Team) co-established by MPEG(Moving Picture Expert Group) under ISO/IEC and VCEG(Video Coding Expert Group) under ITU-T[2]. The SVC standard aims at providing the technologies for flexible representation of its compressed bitstream to make it possible to cope with various display sizes and wide range of network bandwidth etc. SVC provides three scalabilities: spatial, temporal and quality. SVC represents three scalabilities using layer structure. In each scalability layer, the first layer is called the base layer and all higher layers, called enhancement layers, are built on top of the base layer.

*B. Proposed real-time SVC encoder*

In this paper, we only considered the spatial scalability for mobile broadcasting application. For spatial scalability coding, SVC incorporates interlayer prediction and independent AVC between the base and enhancement layer. The base layer of SVC is compatible with AVC. For the enhancement layer, the inter-layer prediction coding includes inter-layer texture prediction, inter-layer motion prediction and inter-layer residual prediction between two layers. The encoding information such as texture, motion and residual data from base layer is also used in encoding enhancement layer.

For real-time encoding, we optimized H.264/AVC (the base layer) encoding algorithms using our developed fast intra-prediction, fast sub-pel motion prediction, fast zero motion block detection and fast mode decision between intra and inter mode.

Moreover, we utilize fast processing unit and multi-core or multi-threading architecture of CPU by using multi-thread programming and Single Instruction Multiple Data (SIMD) assembly techniques.

Figure 1 shows the architecture of developed SVC encoder for real-time application.

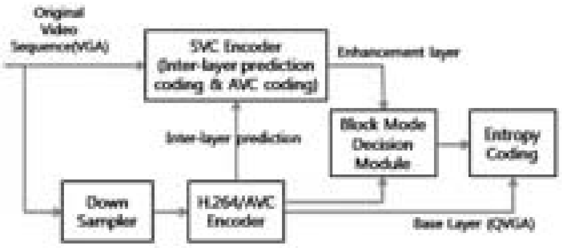



**Figura 1. SVC encoder Stucture**

Input video is down-sampled and the down-sampled video sequence is fed into the AVC encoder for the base layer.

The spatial enhancement layer is encoded using the inter-layer prediction coding and independent AVC coding. The AVC coding in enhancement layer is the same as that of the base layer. The inter-layer prediction is performed based on Macro Block (MB) mode selected in the base layer. Between modes in the independent AVC coding and inter-layer prediction coding, the mode with the least rate-distortion cost is chosen as MB mode in the enhancement layer.

In this SVC encoder structure, we process the base layer encoding, up-sampling and enhancement layer encoding in the same thread to reduce processing time. This method simplifies the data exchange and synchronization among threads.

*C. Implementation of the proposed real-time SVC encoding*

**Figure 2. User interface of SVC encoder**

Figure 2 shows user interface of developed SVC encoder. During encoding, the left two windows show input video for encoder and the right two windows show decoded video. Therefore you can compare the quality between original and SVC decoded video. You can also recognize real-time processing. Our encoder has several encoding options: quantization parameter, method of motion estimation, search range for motion estimation, number of frames for encoding and etc.

## 3. Experimental result

The experiment was performed with video sequences with various motions and textures. For the performance evaluation of our SVC encoder, the encoding time (FPS: Frame per Second) was measured for each test video in conjunction with PSNR drop and bitstream increment.

The Conditions for our experiments:

- CPU : Intel Core 2 Dui E6600, 2.40Ghz
- Memory: DDR2 800Mhz 4GB
- Video Resolution : Base layer-QVGA(320x240)
  Enhancement layer-VGA(640x480)
- Encoding threads : 4
- Frame number : 3000
- Motion estimation method : Diamond search

| QP | 20 | 24 | 26 | 28 | 30 | 40 |
|---|---|---|---|---|---|---|
| FPS | 30.21 | 32.61 | 35.82 | 36.42 | 39.37 | 39.54 |
| PSNR | 43.31 | 40.68 | 38.85 | 36.98 | 35.56 | 27.87 |

**Table 1. Test results of SVC encoder**

Table 1 shows that the real-time processing of SVC encoding is successfully achieved with good performance in terms of PSNR values and encoding speed.

## 4. Conclusion

In this paper, we show a real-time SVC encoder with two layers of spatial scalability. In order to achieve the real-time encoding capability, we optimized H.264/AVC encoding algorithm with our fast encoding methods and utilized multi-thread and SIMD techniques. The experimental results show the encoding speeds from 50 to 70 frames per second with acceptable PSNR quality.

## References

[1] ISO/IEC JTC 1/SC29/WG 11, Joint Scalable Video Model JSVM-10, N8963, San Jose, CA, USA, April 2007.

[2] ISO/IEC JTC 1/SC29/WG11:'Text of ISO/IEC 14496-10:2006/PDA M3 Scalable Video Coding', MPEG76/N8015, Montreux, April 2006.

# The PrestoSpace Metadata Access and Delivery Platform

L. Boch G. Dimino A. Messina
RAI - Centre for Research and Technological Innovation
Corso Giambone 68, I-10135 Turin, Italy
{a.messina, l.boch, g.dimino}@rai.it

W. Bailer
JOANNEUM Research
Steyrergasse 17, 8010 Graz, Austria
werner.bailer@joanneum.at

R. Basili
Università Roma Torvergata
Via del Politecnico 1, I-00133 Roma, Italy
basili@info.uniroma2.it

W. Allasia M. Vigilante
euriX s.r.l.
c.so Tortona 17, I-10153 Turin, Italy
{allasia,vigilante}@eurix.it

C. Bauer
ORF
Würzburggasse 30, A-1136 Vienna, Austria
christoph.bauer@orf.at

## 1. Introduction

The PrestoSpace Integrated Project was launched in 2004 under the Information Society Technologies priority of the Sixth Framework Programme of the European Community (IST FP6 507336). Several European broadcasters and audiovisual archive owners, universities and research centres and industry representatives are part of the PrestoSpace consortium. The project website is www.prestospace.org. The objective of the project is to provide technical devices and systems for digital preservation of all types of audio-visual collections, by building up preservation factories providing affordable digitisation, management and distribution services. This demonstration concerns the Metadata Access and Delivery subsystem of the PrestoSpace infrastructure, which aims at collecting descriptive metadata from the numerical analysis of audiovisual material, and providing a search and retrieve interface for archivists to this information [9]. The former task is performed by the Documentation Platform. The latter task is the function of the Publication Platform, a Web application collecting the extracted metadata in an organised and ergonomic way, for a fast and efficient browsing of the documented material. The Publication Platform is the object of the proposed demonstration.

## 2. Summary of the metadata extraction tools

Metadata extraction tools operate in the metadata collection phase. There are two families of tools: content analysis tools and semantic analysis tools. The PrestoSpace Documentation Platform includes the following audiovisual content analysis tools:

- Shot boundary detection. The shot boundary detection tool segments a video to its primary building blocks, i.e. its shots, and is capable of detecting both abrupt (cuts) and gradual transitions (such as dissolves, fades, wipes, etc.) [1].

- Key frame and stripe image extraction. The key frame detector extracts a number of key frames per shot, depending on the amount of visual change. Stripe images are spatiotemporal representations of the visual essence, created from the content of a fixed or moving column of the visual essence over time.

- Camera motion detection. The camera motion detector analytically describes four basic types of camera motion in the content (pan, tilt, zoom, roll), a rough quantisation of the amount of motion, and the length of the segments in which they appear [2].

- Speech to text transcription. An automatic speech-to-text engine is used, developed by ITC-IRST [5], capable of extracting text from English and Italian spoken content.

- Audio structuring and segmentation. This analysis consists in classifying segments of audio in four principal categories (silence, music, speech, noise).

- Editorial parts segmentation. Editorial parts are the constituent parts of the programme from the editorial point of view. The Documentation Platform uses an automatic editorial segmentation technique in the news domain, choosing a multi-layer approach that merges video and audio information.

The PrestoSpace Documentation Platform includes the following semantic analysis tools:

- Linguistic Processing. The linguistic processing is carried out by a Natural language parser called CHAOS [3][4], which includes several language processing components used to extract semantic entities from text.

- News Categorization. Semantic categories are automatically assigned by a text classifier based on a traditional supervised machine learning model. We used an extended version of the profile-based classifier, known as the Rocchio model.

- Web Alignment. A spidering process is employed to retrieve all the documents published in a target temporal window in the main news websites. This is centred on the broadcasting day by adopting a symmetric span.

- Ontological integration. In MAD, the KIM platform [8] is in charge of making available extensive ontological knowledge about the news domain, and supporting indexing and navigation functionalities. It provides novel Knowledge and Information Management infrastructure and services for automatic semantic annotation, indexing, and retrieval.

## 3. Architecture of the Publication Platform

The extracted metadata are represented by a single XML-based document format, taking the best from each of two metadata standards natively orientated to the description of audiovisual objects, MPEG-7 [7] and P_META [6]. P_META was adopted due to its complete set of information structures for the identification, classification and publication-related features of a programme, while MPEG-7 standard was adopted due to its powerful temporal segmentation tools and for its comprehensive set of standard audiovisual descriptors. The rich variety of information extracted by different analysis modules poses several requirements to the Information Retrieval functionalities in the publication phase. The user interface should model access methods according to different (and integrated) capabilities: a) full text search as usually applied by mostly popular search engines; b) Natural Languages Questions; c) Semantic browsing as navigation through concepts, relations and instances of the ontology. The Publication Platform architecture is based on a Web application as user interface, a DBMS storing the available information related to programmes, and the KIM indexing and search engine. The search interface supports the various retrieval approaches. The user can choose the target of his/her search (e.g. a programme or a news item), which can be filtered by title, broadcast date and service, contributions (e.g. authors, journalists, directors), classification (topics, categories), text of description. The browsing interface is made up of four frames: a video preview, the editorial parts tree, the key frames, and an extensible multi-tab frame, each of which is representing a specific elaboration result. The content of all the frames is synchronised during user interaction.

## 4. Demonstration contents

The Publication Platform has been extensively tested during the project lifetime, through the organisation of structured users workshops in which all tested tools got an overall positive result. The demonstration is based on a collection of several hours of audiovisual material, taken from RAI, BBC and ORF archives. Search and retrieve functionalities and browsing capabilities are illustrated in detail during the demonstration, with queries covering the full set of the available modalities.

## References


[1] W. Bailer, H. Mayer, H. Neuschmied, W. Haas, M. Lux, and W. Klieber. Content-based video retrieval and summarization using MPEG-7. In *In Proc. of Internet Imaging V*, 2004.

[2] W. Bailer, P. Schallauer, and G. Thallinger. Joanneum research at TRECVID 2005 - camera motion detection. In *In Proc. of TRECVID Workshop*, 2005.

[3] R. Basili, M. T. Pazienza, and F. M. Zanzotto. Efficient parsing for information extraction. In *In Proc. of European Conference on Artificial Intelligence (ECAI98)*, 1998.

[4] R. Basili and F. Zanzotto. Parsing engineering and empirical robustness. *Journal of Language Engineering*, 8(2/3):97–120, 2002.

[5] F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. A system for the segmentation and transcription of italian radio news. In *In Proc. of RIAO, Content-Based Multimedia Information Access,*, 2000.

[6] EBU. P_META 2.0 metadata library. Technical specification, European Broadcasting Union. Metadata Advisory Group, 2007.

[7] ISO/IEC 15938. Multimedia content description interface. Iso/iec international standard.

[8] A. Kiryakov, B. Popov, D. Ognyanoff, D. Manov, A. Kiriliov, and M. Goranov. Semantic annotation, indexing, and retrieval. *Elsevier's Journal of Web Semantics*, 2(1), 2005.

[9] A. Messina, L. Boch, G. Dimino, R. Basili, W. Bailer, P. Schallauer, W. Allasia, M. Vigilante, and M. Groppo. Creating rich metadata in the TV broadcast archives environment: the PrestoSpace project. In *Proc. of AXMEDIS06 Conference*, 2006.

# Ontology-Driven Digital Preservation of Interactive Multimedia Performances

Alexander Mikroyannidis,[1] Bee Ong,[1] Kia Ng,[1] and David Giaretta[2]
[1]*ICSRiM - University of Leeds,*
*School of Computing & School of Music, Leeds LS2 9JT, UK*
[2]*STFC, Rutherford Appleton Laboratory, Oxfordshire OX11 0QX, UK*
*caspar@icsrim.org.uk   www.icsrim.org.uk/caspar*

## Abstract

*Interactive multimedia performances are rapidly gaining ground within performing arts communities nowadays, mainly due to breakthroughs in human-computer interaction technologies, such as human motion capture and analysis. This has brought forward the issue of digital preservation of these performances, so that they can be reconstructed in the future. This paper presents an ontology-driven approach to the digital preservation of interactive multimedia performances. An ontology model is proposed for describing the complex relationships amongst different components of a performance, as well as their temporal aspects and evolution over time.*

## 1. Introduction

An Interactive Multimedia Performance (IMP) involves one or more performers who interact with a computer based multimedia system making use of multimedia contents that may be prepared as well as generated in real-time including music, manipulated sound, animation, video, graphics, etc.

An example of an IMP process is the one adopted in the MvM (Music via Motion) interactive performance system, which produces music by capturing user motions [1]. The system captures user motions using motion capture devices and stores them in a 3D format. These motions are then mapped into music by using a mapping strategy, with parameters provided through a GUI. The motion-music map is forwarded to the generation component which produces multimedia content.

IMP preservation is a challenging issue. In addition to the output multimedia contents, related digital contents such as mapping strategies, processing software and intermediate data created during the production process (e.g. data translated from "signals" captured) have to be preserved, together with all the configuration, setting of the software, changes (and time), etc. The most challenging problem is to preserve the knowledge about the logical and temporal relationships amongst individual components so that they can be properly assembled into a performance during the reconstruction process of the original IMP.

The preservation of IMPs produced by the MvM system comprises part of the Contemporary Arts testbed dealing with preservation of artistic contents, which is one of the three testbeds of the EU project CASPAR (www.casparpreserves.eu). The other two are Scientific and Cultural testbeds which are focused on very high volume and complex scientific data objects and virtual cultural digital objects respectively.

This paper introduces an ontology approach to describing IMPs for their preservation. A set of extensions for the CIDOC-CRM standard [2-4] are proposed, together with an ontology model for describing temporal facts. The remainder of this paper is organized as follows: Section 2 presents some metadata approaches to digital preservation. The applicability of CIDOC-CRM in digital preservation is discussed in section 3. Section 4 introduces the proposed extensions to CIDOC-CRM and section 5 presents the ontology model for the temporal enrichment of metadata. Finally, the paper is concluded and some plans for future work are provided.

## 2. Related Work

Metadata and ontologies have been proven an important factor in digital preservation. Metadata element sets designed specifically for preservation purposes include those developed by RLG Working Group on Preservation Issues of Metadata (RLG) [5], CURL Exemplars in Digital Archives (CEDARS) project (www.leeds.ac.uk/cedars) [6], the metadata of the National Library of Australia (NLA) [7] and the Networked European Deposit Library (NEDLIB) [8]. A consensus effort was carried out by the OCLC/RLG Working Group on Preservation Metadata to develop a common metadata framework to support the preservation of digital objects, which was based largely on CEDARS, NEDLIB and NLA element sets [9]. The Preservation Metadata Implementation Strategies (PREMIS) Working Group later built on this framework a PREMIS data model and a data dictionary for preservation metadata [10].

The CIDOC Conceptual Reference Model (CRM) has been proposed as a standard ontology for enabling interoperability amongst digital libraries [2-4]. CIDOC-CRM defines a core set concepts for physical as well as temporal entities. This is very important for describing temporal dependencies amongst different objects in a preservation archive. A combination of core concepts defined in CIDOC-CRM and multimedia content specific concepts of MPEG-7 for describing multimedia objects in museums has also been introduced. A harmonisation effort has also been carried out to align the Functional Requirements for Bibliographic Records (FRBR) [11] to CIDOC-CRM for describing artistic contents. The result is an object oriented version of FRBR, named FRBRoo [12].

## 3. CIDOC-CRM for Digital Preservation

CIDOC-CRM was originally designed to describe cultural heritage collections in museum archives. The meta-schema of CIDOC-CRM is illustrated in Fig. 1. CIDOC-CRM's conceptualisation of the past is centred on Temporal Entities (e.g. events). People (Actors) and objects (Conceptual Objects and Physical Objects) involved, time (Time-Spans) and Places are documented via their relationships with the Temporal Entities.

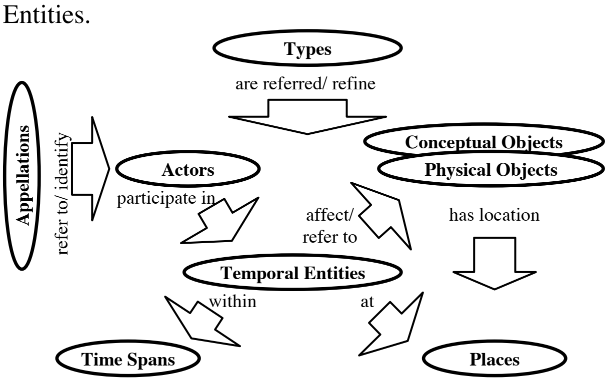



**Figure 1. The meta-schema of CIDOC-CRM [4]**

The CIDOC-CRM vocabulary can be used to describe a performance at a high level. However, more specialised vocabularies are necessary for the interactive performing art domain to precisely describe the relationship amongst the elements of a performance. For example, it is necessary to model how equipments are connected together in the performance. Some concepts representing digital objects have been very recently introduced in CIDOC-CRM for digital preservation purposes. Nevertheless, there is a need for documenting the relationships amongst software applications, data, and operating systems, as well as the operations performed on them. In addition, CIDOC-CRM is designed primarily for the documentation of what has happened, whereas in digital preservation, it is also required to document the reconstruction of a past event from preserved components.

## 4. CIDOC-CRM Extensions for Preservation of IMPs

In order to address the preservation of IMPs, we propose a set of extensions to CIDOC-CRM. These extensions have the following objectives:

- To provide a domain specific vocabulary for describing objects related with IMPs.
- To provide a vocabulary for describing the interrelationships between digital objects and the operations performed on them in the digital preservation context.

Fig. 2 shows the set of concepts that we have introduced in CIDOC-CRM to describe IMP objects. The extended concepts are prefixed by IMP and an identification number. The original CIDOC-CRM entity and property names are prefixed by E and P respectively.

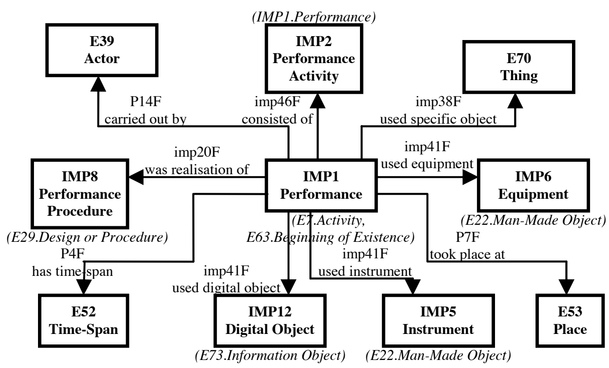



**Figure 2. CIDOC-CRM extensions for describing IMP objects**

More specifically, the following concepts have been introduced:

- *"IMP2.Performance Activity"*: for describing activities of a performance.
- *"IMP5.Instrument"*: a specialisation of CIDOC-CRM "E22.Man-Made Object" for modelling musical instruments (e.g. cellos, violins, drums, etc.) used in a performance.
- *"IMP6.Equipment"*: a specialisation of CIDOC-CRM "E22.Man-Made Object" for modelling equipment (e.g. a microphone, a sound mixer or a computer, etc.) used in a performance.
- *"IMP8.Performance Procedure"*: a specialisation of CIDOC-CRM "E29.Design or Procedure" for describing the procedure in which a performance should be carried out.
- *"IMP12.Digital Object"*: a specialisation of "E73.Information Object" for describing digital objects.

As shown in Fig. 3, "IMP12.Digital Object" has two subclasses: "IMP17.Digital Data Container" and "IMP18.Digital Data Object". A digital data container ("IMP17.Digital Data Container") is a container of one or more digital data objects ("IMP18.Digital Data Object"). An example of digital data container is a file. The bit stream contained within the file is considered as a digital data object. This separation is necessary to model a bit stream in memory or in cases where multiple bit streams carrying different information carried by a single digital data container. A special type of digital data object is a computer program (IMP13.Computer Program). In this case, the bit stream is a set of instructions to be executed by a computer. There are two specialisations of computer programs: "IMP14.Operating System" and "IMP15.Software Application".

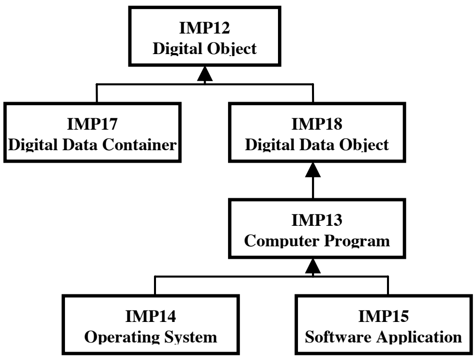



**Figure 3. Classification of digital objects**

Operations on digital objects can be described using "IMP26.Digital Object Operation", which is a specialisation of CIDOC-CRM "E5.Event". A number of subclasses of "IMP26.Digital Object Operation" have also been defined to deal with common operations such as creation, duplication, transformation, modification, access and deletion. This is necessary in the preservation context, where the history of a digital object needs to be documented.

## 5. Bitemporal Ontology Modelling

In order to capture temporal facts in our preservation ontologies, we propose the use of the bitemporal ontology model of Heraclitus II [13]. The Heraclitus II framework considers ontologies as a semantically rich knowledge base for information management and proposes ways for the management and evolution of this knowledge base. Heraclitus II uses an ontology model that is based on the object model defined by the Object Data Management Group (ODMG) [14] and more specifically on TAU [15]. The TAU model is an extended version of ODMG that supports modelling and reasoning about time and evolution.

Ontology modelling in Heraclitus II is bitemporal, allowing for ontology representation over two dimensions of time: *valid* and *transaction* time. The valid time of a fact is defined as the time when that fact is true in the modelled reality. The transaction time of a fact is defined as the time when that fact is current in the knowledge base and may be retrieved. Valid times can belong in the past, present or future and are usually supplied by the ontology author. Transaction times are provided by the ontology management system, cannot change and are bounded between the knowledge base creation time and the current transaction time.

Ontology objects, namely concepts, relations and instances, can be associated with transaction time, valid time, both (bitemporal), or none (static). This modelling allows for *retro-active* as well as *pro-active* changes to be captured and represented on the knowledge base. A retro-active change occurs when a fact that is entered at a certain transaction time in the knowledge base, has been valid in the real world before this transaction time. On the other hand, when the valid time of a fact is greater than its transaction time, then a pro-active change is captured in the knowledge base.

In digital preservation, certain changes have to be monitored in order to keep the archived IMP up-to-date and be able to reconstruct it at any time. These changes mainly regard the hardware and software the IMP was produced with, as well as changes in the environment of the IMP, such as changes in the performers' behaviour or in the environment setting of the performance. Modelling these aspects in a bitemporal ontological form can help capture and document any occurring changes more systematically and efficiently.

For example, the behaviour of the performers is an essential factor for maintaining the authenticity of an IMP. It is very likely though, that changes will occur over time in the fashion a performer plays an instrument, thus compromising the authenticity of the reconstructed performance. One way to address this issue can be modelling the original behaviour of the performer within an ontology, with the use of bitemporal concepts and relations between them. We can then capture any changes in the performer's behaviour and represent them on the corresponding bitemporal ontology objects, thus keeping the evolution history of the performer's behaviour in our knowledge base. In this way, we should be able to re-interpret the performer's behaviour in order to adapt the rest of the IMP accordingly.

## 6. Conclusions and Future Work

The present paper explored the area of digital preservation of IMPs. An ontology-driven approach has been proposed, by extending the current concepts defined in the CIDOC-CRM standard, for preservation of IMPs. A number of concepts describing the performing art domain, as well as digital objects have been proposed. In addition, a bitemporal ontology model has been presented, addressing the temporal aspect of performances and their evolution over time.

As future work, the authors are planning to evaluate the proposed CIDOC-CRM extensions using MvM performance data. The proposed ontology will also be integrated with the architecture of CASPAR project for use by its software components.

## 7. Acknowledgements

Work partially supported by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project CASPAR. The authors are solely responsible for the content of this paper. It does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

## 8. References

[1] K. C. Ng, "Music via Motion: Transdomain Mapping of Motion and Sound for Interactive Performances," *Proceedings of the IEEE*, vol. 92, 2004.

[2] M. Doerr, "Increasing the Power of Semantic Interoperability for the European Library," in *ERCIM News*, vol. 66, 2006.

[3] T. Gill, "Building semantic bridges between museums, libraries and archives: The CIDOC Conceptual Reference Model," *First Monday*, vol. 9, 2004.

[4] M. Doerr, "The CIDOC CRM - an Ontological Approach to Semantic Interoperability of Metadata," *AI Magazine*, vol. 24, 2003.

[5] B. Berger, J. Coleman, W. Cromwell-Kessler, R. Dale, B. DeCandido, C. Montori, and S. Ross, "RLG Working Group on Preservation Issues of Metadata - Final Report," RLG and Preservation 1998.

[6] M. Day, "Metadata for Preservation - CEDARS Project Document AIW01," UKOLN 1998.

[7] National Library of Australia, "Preservation Metadata for Digital Collections," 1999.

[8] C. Lupovici and J. Masanes, "Metadata for long term-preservation," NEDLIB 2000.

[9] OCLC/RLG Working Group on Preservation Metadata, "Preservation Metadata and the OAIS Information Model - A metadata framework to support the Preservation of digital objects," 2002.

[10] PREMIS, "Data Dictionary for Preservation Metadata - Final Report of the PREMIS Working Group," OCLC and RLG 2005.

[11] FRBR, "Functional Requirements for Bibliographic Records - Final Report," International Federation of Library Associations and Institutions (IFLA), Frankfurt am Main, Germany 1997.

[12] M. Doerr and P. LeBoeuf, "FRBRoo Introduction," 2006.

[13] A. Mikroyannidis and B. Theodoulidis, "Heraclitus II: A Framework for Ontology Management and Evolution," in *Proc. 2006 IEEE /WIC/ACM International Conference on Web Intelligence (WI 2006)*, Hong Kong, China, 2006, pp. 514-521.

[14] R. G. G. Cattell and T. Atwood, "The Object Database Standard: Odmg-93: Release 1.2." San Francisco: Morgan Kaufmann Pub, 1996.

[15] I. Kakoudakis and B. Theodoulidis, "TAU: Towards a Temporal Object-Oriented DBMS," in *Proc. 9th International Database Conference (IDC)*, Hong Kong, 1999.

# Panel on Enabling DRM Interoperability with Interoperable Rights Expression Languages

*Organized by*

MPEG-21 REL Experts, ISO/IEC JTC1/SC29/WG11

*Coordinated by*

Jaime Delgado, UPC (Spain)

Xin Wang, ContenGuard, Inc. (USA)

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

# Enabling DRM Interoperability with Interoperable Rights Expression Languages

Jaime Delgado, Universitat Politècnica de Catalunya, Spain
Xin Wang, ContentGuard, Inc., USA
Taehyun Kim, DRM inside, South Korea
Florian Schreiner, TU Munich, Germany
Satoshi Ito, Toshiba, Japan
Nigel Earnshaw, BBC, UK

## Panel objective

DRM (Digital Rights Management) technologies have been around for about 10 years - a rather long way. There are some successfully deployed systems in practice that start to affect people's daily content consumption experiences and advance digital content marketplace. However, the interoperability issues among the DRM systems give consumers poor user experience and hinder even wider adoption of DRM in the marketplace. The panel is to discuss how to easy those DRM interoperability issues to facilitate the acceleration of adoption of DRM technologies in the marketplace.

Specific issues are:

- System issues around DRM and REL (Rights Expression Languages) interoperability
- Future of RELs
- What the industry needs to do with interoperability solutions

Among many DRM interoperability issues, the one with rights expression languages has been brought up lately. The issue is, as content moves from one DRM system into another DRM system, how to make rights expressed in one REL for the first DRM system understandable in the second DRM system that uses another REL to express its rights. One possible solution to this issue is to have a reference REL, to which others should relate in order to obtain interoperability between systems using different RELs. However, other solutions are possible. To facilitate the interoperability among RELs, MPEG-21 REL, for instance, is defining profiles (extensions and subsets of the original ISO/IEC IS 21000-5) that try to match specific scenarios that might have alternative RELs. Examples include the MAM (Mobile And optical Media) profile, that tries to facilitate interoperability with, for example, OMA (Open Mobile Alliance) DRM, and the DAC (Dissemination And Capture) profile, one of its goals being to facilitate interoperability with several other systems in the broadcast area, such as TV-Anytime.

Several experts from the MPEG Committee (ISO/IEC JTC1/SC29/WG11) have organized an open Industrial Panel to discuss how to enable DRM interoperability through resolving interoperability issues around RELs, not only from the MPEG point of view but also from others, such as ODRL, OMA, TV-Anytime, Coral, DMP, etc. For this reason, experts from some of these groups have collaborated in the organization of the Panel.

Apart from the MPEG-21 and ODRL Rights Expression Languages, the panel will review two publicly available 'rights expression' signalling conventions targeted at the broadcast industry. The first, TV-Anytime's Rights Management and Protection Information (RMPI), was the result of work by participants of the TV-Anytime forum and published in 2005. The second more recent publication is from the Digital Video Broadcasting Project (DVB) Consortium, known as DVB Usage State Information (USI) and has been released in the DVB 'Blue Book' as part of the DVB specification for the DVB Copy Protection and Copy Management (CPCM) system. These two specifications share some similarities in approach, scope and design and will be reviewed in the context of the business models they potentially support.

The target audience is those involved in the analysis, use and development of multimedia content creation, management and distribution systems using different DRM systems and solutions.

## Panel questions

Some of the questions that will be discussed in the panel are:

- Is MPEG-21 REL profiling a way to achieve REL interoperability?
- Is it reasonable to use a limited REL?
- Is it feasible a business model based on the interoperation of different RELs?
- Do we need a dictionary or ontology?
- How is the market evolving and what is requiring on this?
- What the industry needs to do with the interoperability solutions?
- Is it necessary to standardize protocols to communicate between servers (such as a license server) and their clients?
- Is Trusted Computing a possibility to achieve Interoperability? Can it be a base for a secure conversion of licenses in decentralized systems?

## Presentations

- MPEG-21 REL and its profiles.
- ODRL version 2 and OMA REL.
- TV-Anytime RMPI and DVB USI, RELs for broadcast related applications.
- Implementing REL.

# Panel on Issues in Security for Digital Rights Management

*Organized by*

WP3.2 partners from VISNET-II European NoE

*Coordinated by*

Eva Rodriguez and Jaime Delgado, UPC (Spain)

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

# Issues in Security for Digital Rights Management

*Organised by the WP3.2 partners from VISNET-II European NoE*

Eva Rodríguez, Jaime Delgado, *Universitat Politècnica de Catalunya (Spain)*
Adrian Waller, *Thales Research & Technology (UK)*
Carlos Serrao, *ISCTE/DCTI/ADETTI (Portugal), Universitat Pompeu Fabra (Spain)*

## Panel objective

Security and Digital Rights Management (DRM) are key topics in many contexts. However, when we put both together and we try to specify an architecture for them, we confront new research issues that are based on current ones. This is the focus of this Panel.

In particular, the topics to be discussed include:

- Architectures to manage digital rights for audiovisual content considering its whole life cycle, from its creation to its consumption by end users.
- Definition of mechanisms for the secure management of licenses.
- Definition of key storage and distribution architectures considering web services, secure storage for the secure management of keys, authentication, certificates, etc.
- Event Reporting security.
- Security mechanisms compatible with content adaptation.
- Access control frameworks and their integration in broader architectures and applications.
- Architectures for Access Control and DRM interoperability.
- Interoperability of the representation information for DRM (licenses) and SAML (security) and other possible standards.
- Use of Semantic Web tools to formalise semantics for security services.
- Use of Ontologies for policy management.
- DRM and Security solutions: Benefits to publishers.

The VISNET-II (Networked Audiovisual Media Technologies) Network of Excellence co-funded by the European Commission is working, in one of its workpackages, on the topics mentioned above. Results will be presented in order to start discussions with the aim of identifying research issues and services and products to develop.

Furthermore, other projects, as OpenSDRM, from ADETTI, or CENIT Segur@, co-funded by the Spanish Ministry of Industry, will also present their approach to the issues.

The OpenSDRM project has its roots on former MOSES EC RTD project, where a open DRM platform had to be build to support the emerging MPEG-4 IPMP eXtensions framework. OpenSDRM is an open-source, secure and distributed DRM architecture to support the complete digital content value chain, from content provider to the end-user. OpenSDRM source-code is publicly available and can be used to test and establish DRM governed digital content scenarios.

The Segur@ project has started working in architectures for the development of semantic trust and for the development of trust based on data. Then, in this project, ontologies for basic security services will be defined and how ontologies can help in policy management will be analysed.

## Panel questions

Some of the questions that will be discussed in the panel are:

- Are there pending security issues in existing DRM systems?
- How security can help to the management and distribution of licenses and keys?
- Can security (or access control) and DRM architectures interoperate?
- How they interoperate in specific environments, such as Virtual Collaboration?
- Can intellectual property rights information be adequately represented and processed using SAML, XACML or other standards?
- How semantic web tools can be used to formalise basic security services?
- How ontologies can help to the specification and management of policies?
- How DRM supports publishing business models?
- Are publishers taking benefit from existing DRM and Security solutions?
- Is open-source DRM a solution for DRM?
- How the DRM market will shape for the future?
- DRM - rights management vs. copy-protection?

## Presentations

- Architecture, Access Control and Security in DRM: The VISNET-II Project approach.
- Open-source DRM: The OpenSDRM Project approach.
- Use of Web semantic tools for trust and security services: The SEGUR@ Project approach.

# Panel on Digital context, collective licensing and cultural diversity: the way forward

*Organized by*

AFI, Italian Association of Phonographic Producers

*Coordinated by*

Massimo Baldinato, AFI (Brussels)

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

# **“Digital context, collective licensing and cultural diversity: the way forward”**

Massimo Baldinato
*Italian Association of Phonographic Producers*
*CONFINDUSTRIA*
*Brussels Office*

Is there a form of cultural diversity specific to today’s digital world?

How collective licensing is being adapted to the digital framework?

What are the challenges to be met by participating in and appropriating digital diversity in a context marked by the economisation of cultural resources?

In a study presented to the EU’s Ministers of Culture on 13 November 2006, the European Commission put in evidence the importance of the culture sector for the EU economy underlining its potential for creating more and better jobs in the future. With its 5.8 million employees, the Culture sector employs more persons than the total employed in Greece and Ireland put together. Further, the culture sector accounted for 2.6% of EU GDP in 2003, and experiences higher growth rates than the average of other sectors of the economy.

A further Europe-wide survey has revealed that two-thirds of Europeans feel that they share elements of a collective culture. Nearly nine out of ten Europeans say that culture, cultural exchanges and intercultural dialogue should have an important place in the EU.
These findings emerged from the recent Eurobarometer survey of people's views on culture, which was carried out during the spring 2007. The survey covered 26,000 persons from all over Europe and from all walks of life.

On the other hand information society is booming.

In a recent speech Mrs Viviane Reding, Commissioner for Information Society and Media, said: *“Convergence of audiovisual media, broadband networks and electronic devices is generating new opportunities in the ICT and content sectors. It is both creating new delivery channels for traditional formats and opening the path to the development of interactive content and services.*
*Furthermore, we have a fantastic opportunity to make our cultural heritage accessible online.”*

The issue is how finding solutions agreed among all the stakeholders, respecting the rights of the content owners while boosting the digital content’s market.

In this framework regulatory aspects at European level are of great importance, also under a competition point of view. This aspect will be deepened during this panel thanks to the participation of Mr Carlo Toffolon, of the Unit Media of the DG COMP of the European Commission.

Collective licensing and the role of collecting societies is another interesting issue and Mrs Silvina Munich from CISAC, the umbrella organization of the collecting societies worldwide, will contribute in clarifying this aspect.

Finally the approach of the electronic industry will be presented as well by a representative of this sector.

# I-MAESTRO 3rd Workshop on Technology Enhanced Music Education

*Organised by*

i-Maestro EC IST Project
www.i-maestro.org

*Coordinated by*

Kia Ng, ICSRiM, University of Leeds, UK

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 © Firenze university press, 2007

# The EASAIER Project: Creating a 'Virtual Oral/Aural Learning Environment' for Scottish and Irish Traditional Music

Joseph Harrop & Celia Duffy
EASAIER project, Royal Scottish Academy of Music and Drama
j.harrop@rsamd.ac.uk, c.duffy@rsamd.ac.uk

## Abstract

*EASAIER (Enabling Access to Sound Archives through Interaction, Enrichment and Retrieval) is an EU funded project that brings together seven partners, from academic and commercial sectors, to develop licence-free innovative access and interactive software for sound archives, according to specified user needs. EASAIER is a 36 month-long project, due to finish development in November 2008.*

*Some of the EASAIER software tools digitally replicate techniques and methods of oral/aural learning observed in Scottish and Irish traditional music. This short paper discusses these techniques alongside the software tools that allow for them to be recreated in a virtual learning environment.*

*The EASAIER project is developing a software package that will enable the important cultural content stored in sound archives to be accessed as an invaluable, interactive digital educational resource.*

## 1. Introduction

> 'It is like learning from a real person except you're in remote control of the scenario, but the other thing is that tradition bearers only last their lifetime and I imagine that this would be incredibly useful for learning from players who aren't here anymore.'
>
> Lori Watson, Scottish traditional fiddler [1]

EASAIER stands for Enabling Access to Sound Archives through Interaction, Enrichment and Retrieval. Funded by the EU FP6 Information Society Technologies strand, EASAIER brings together a total of seven partners from the audio and electronic engineering sectors, both academic and commercial, to develop licence free software for sound archives. EASAIER is a 36 month-long project, due to finish development in November 2008 [2]. The authors are part of the EASAIER project team, and their specific responsibilities concern the project's user needs specifications and evaluation of the final software package.

EASAIER's enhanced access tools will improve access to, and enable interaction with, sound archive content. The example discussed in this paper is a demonstration of the benefits of practice-based online interaction for educational purposes. When viewed alongside the powerful search and retrieval functions being developed for the same software package, the potential for sound archive content as a large-scale educational resource becomes clear.

Some of the EASAIER software tools can be employed to digitally recreate the oral/aural learning techniques (i.e. learning 'by ear') observed in Scottish and Irish traditional music. In this paper, basic techniques and methods of oral knowledge transfer in Scottish and Irish traditional music will be discussed, followed by a description of how EASAIER's enhanced access tools can be seen to empower these same methods and techniques in a digital environment.

## 2. Oral/aural learning in Scottish and Irish traditional music

The Royal Scottish Academy of Music and Drama (RSAMD) offers the only honours degree in Scottish traditional music in the world and is the centre for the recently launched Scottish Traditional Music Graded Exams. The RSAMD also houses the HOTBED learning resource database (HOTBED stands for Handing On Tradition By Electronic Dissemination) [3]. This database of audio and video content was developed with the empowering of methods of oral/aural learning in mind. The EASAIER user tools build on the outcomes of HOTBED. RSAMD Scottish traditional music staff and students regard the manner in which a particular tune or song is learned as of paramount importance, compared to their classical music colleagues. Dr Frances Morton, research officer for the ESRC-funded project, Investigating Musical Performance, elaborates:

> 'Scottish traditional music students in our IMP case studies and questionnaires consistently rated skills such as reading music as unimportant, whereas classical musicians rated these as of the highest importance to their genre. Also, memorisation and improvisation were rated highly by Scottish traditional musicians, but unimportant by their classical music counterparts' [4]

Learning from written notation is an accepted approach on the Scottish traditional music course, however, oral/aural learning is 'the norm' [5].

So what exactly is meant by 'learning by ear'? There is, no doubt, a variety of techniques. However, in Scottish and Irish traditional music, specific methods of instruction and demonstration are coupled with a keen ear and ability to memorise on behalf of the learner. In addition to this, it is accepted that the pupil will often assimilate some features of the teacher's playing. For example, the influential Irish fiddler, Junior Crehan, described his teacher John 'Scully' Casey as 'a great fiddle player and the best practitioner of the ornamental style that I ever heard ... he'd play the tune and I'd play with him and I'd take his style' [6]. Michael Downes learnt fiddle from Crehan, and describes his approach to teaching a tune 'by ear':

> 'At first, Junior would finger them out to you – by time, he'd just play them out and you might play them away with him, depending how quick you was to pick up, depending how quick you'd be' [7].

Dr Joshua Dickson, highland piper and writer on the sociology of Scottish traditional music, provides some clarification on oral/aural learning in the Scottish piping tradition:

> 'A young piper will ask his or her tutor to go over a piece, again and again, to slow a piece down, to concentrate on a particular phrase in order to eek out that which makes that particular phrase a whole piece of music in itself: a whole universe of rhythmic dynamism. That pupil will then be able to take this on, to mimic, and then ultimately put something of themselves into it' [8].

Alongside the ability to memorise and retain tunes, the learner will engage in one or more from the following list of methods are used in oral/aural learning in Scottish and Irish traditional music (see Table 1). This table features those methods digitally recreated by EASAIER project's software tools. As a list, it does not aspire to be a comprehensive inventory of all possible methods of oral/aural learning, but shows the general methods of interaction evident when a learner and a tradition bearer of Scottish and/or Irish music get together to play music, or sing songs. How these methods are recreated by the EASAIER project is discussed below.

**Table 1. Methods of oral/aural learning in Scottish and Irish traditional music**

- Playing/singing along
- Vocalisation or 'diddling' of tunes (in the case of instrumentalists)
- Repetition of tunes/songs
- Slow repetition of tunes/songs
- Repetition of small and large excerpts of tunes/songs
- Slow repetition of small and large excerpts of tunes/songs
- Slow repetition of tune/song excerpt to study smaller features of delivery (e.g. ornamentation, bowing or breathing)

### 3. Oral/aural learning in recent times

The informal, community-based nature of oral/aural learning, in the case of Scottish and Irish traditional music, formerly involved two or more people physically present in the same place at the same time. However, with the advent of recording, it became possible for singers and instrumentalists to interact with music reproduced by phonographs, record, cassette and CD players, and latterly MP3 players and music played via the internet. For Scottish and Irish traditional musicians, this has given rise to new methods of oral/aural learning, different in approach to those techniques prevalent in the 'face-to-face' interaction of learner and tradition bearer.

Firstly, the tradition bearer is no longer physically present, rather present in the recording being played.

This means that all decisions in the oral/aural learning process are made by the learner. Secondly – because of the nature of 'playing' a record, tape, CD or MP3 – musical interaction is reduced to playing or singing along with a tune or song in its entirety, or laborious repetition of musical excerpts by way of stopping playback and 'rewinding' to an appropriate place in the recording.

This is not to say that advances in technology have been detrimental to the transmission of Irish and Scottish traditional music. As Mary McCarthy observes in her book *Passing It On* (1999):

> 'Multiple forms of media were used in the transmission of music in Ireland, from a variety of literacy based media to mass media such as radio and television, to contemporary technologies that are transforming the landscape of music teaching and learning. The technology of music learning has changed radically over the last century, from a Curwen modulator of the 1890s that provided a visual to assist in learning scales and intervals, to interactive computer software of the 1990s that allows a student to hear and explore sounds in unprecedented ways' [9].

## 4. EASAIER's virtual 'oral/aural learning' environment

The EASAIER project is developing a software package with playback functionalities that enable those methods prevalent in a 'tradition bearer-to-learner' interaction to be visited on that of a 'recording-to-learner'. This becomes apparent when the attributes of a particular EASAIER functionality is discussed alongside the methods of oral/aural learning listed above.

EASAIER playback facilities employ, as a base, the Sonic Visualiser audio visualisation software developed by Chris Cannam at the Centre for Digital Music at Queen Mary, University of London. Installed on a computer or browser-based, playing and singing along with a recording would be much the same for a learner as when using, for example, an MP3 player (i.e. using 'play' and 'stop' buttons). Vocalisation, or 'diddling', of tunes is an exclusive feature of the face-to-face interaction.

It is the way the EASAIER software enables the different types of repetition prevalent in oral/aural learning that makes it of particular use to the learner. Additionally, it is this quick and easy access to repetitious methods of learning that sets an EASAIER interaction apart from the recording-to-learner previously discussed. The learner, rather than stopping the recording and rewinding/fast-forwarding to find a specific tune amongst a set of reels or jigs, instead uses the 'Looping and Marking' functionality. This feature gives the learner the ability to place markers at specific points of interest along a digital timeline of the recording. These points might be, for example, the start of a certain tune, or a particular ornamentation of a phrase. They can then choose to 'loop' the music between these marks, eliminating the need for stopping and starting again. This 'looping' feature can be applied to small or large sections of a recording.

The tradition-bearer-to-learner interaction will often feature the rendition of a tune at a slower tempo. Playing a recording at a slower tempo using a record or cassette will result in the overall pitch of the recording going down. CD and MP3 players do not have the facility for slowing down a recording. The EASAIER software's time and pitch scale modification functionality enables slowing down and speeding up without altering the overall pitch of the recording. Undesirable sounds generated by digital modification, or 'artefacts', are rendered near to nonexistent – something particularly difficult to achieve in the case percussive audio, such as drums or guitars. The ability to slow the tempo of a recording down without altering the pitch is useful for learners when studying the more detailed features, such as tune ornamentation, bowing or breathing.

This same feature, which can be applied to downloaded or internet 'streamed' audio content, alters the pitch of a recording without a change in tempo. While playing a tune slower to a learner is an obvious method of oral/aural learning, the pitch altering allows for interaction with recordings of tradition bearers whose instrument may not be set at a practical pitch. For example, using this feature, a fiddle learner can interact with a recording of a highland piper (pipes have notoriously variable levels of pitch) by shifting it to a key more suitable to his/her instrument. Pitch alteration, without changing the tempo, also has obvious applications for the oral/aural learning of traditional songs, where voice range and register can be different between the learner and the recording.

This is arguably a new manner of interaction made accessible to the EASAIER software user. That is to say, while this kind of software is available commercially, EASAIER will be free to the public. In combination with EASAIER searching and retrieval features (see below), this functionality will empower methods of oral/aural learning to a greater extent than when using a CD or record. A clear example of this is the ability to use the time and pitch scale modification functionality on video content. This allows the learner to interact with the visual prompts normally only prevalent in a face-to-face interaction. While videos of tradition bearers were available, learners could not slow down or alter the pitch of the recording to observe salient features. The visual aspect is particularly useful for musicians who do not read music notation – fiddlers and guitarists, for example, will often watch the fingers of the left hand as a guide.

Additionally, the EASAIER project is developing a number of enhanced searching and retrieval functions that supplement the tools to help oral/aural learning. These include a cross-media retrieval function, where different types of media files related to a search will be displayed. For example, were a learner to search for a fiddle tune played by Frankie Gavin, audio files would be displayed alongside other related media such as images of the musician and video of his playing. When searching for songs, the lyrics would be displayed as well as related audio recordings. This 'cross media' retrieval is made possible through the developments in the semantic web. Another useful retrieval function in the EASAIER package is the 'similarity search', where a recording will be submitted to the search engine, and after extracting certain features from the audio file, such as beats per minute and instrumentation, recordings that display similar features are retrieved.

## 5. Conclusion

The package being developed by the EASAIER project aims to enable access to, and empower interaction with, content of important cultural value stored in sound archives. The discussion here has been an illustration of how that interaction and access can enhance oral/aural learning in Scottish and Irish traditional music. It is important to note, however, that EASAIER software has been developed in direct response to the user needs of sound archive managers and those end-users who wish to access sound archive content. Our work has involved discussion with several large sound archives from around the United Kingdom and Europe, including the British Library National Sound Archive, the Institut National de l'Audiovisuel in France, the National Archive of Recorded Sound and Moving Images in Sweden, and the Irish Traditional Music Archives. It was in testing the prototypes for the tools outlined above with traditional musicians – both student and professional – that the idea for this paper emerged.

It may not be a surprise that those wishing to engage with audio and audio-visual recordings would want to learn from it in a manner that reflect the 'non-lexical' nature of the content itself. Alongside the innovative ways of searching an archive, EASAIER software being developed empowers those types of learning and interaction not wholly dependent on literacy, invigorating techniques of oral/aural learning that were not previously available digitally. As any oral/aural learner will confirm, close interaction with the recording can communicate levels of subtlety that the written word and note cannot.

The quote from traditional Scottish fiddler Lori Watson at the beginning of this paper rightly points out that time, and feasibly distance also, place no restraints on the techniques and methods of oral/aural learning when using the EASAIER functionalities. This outcome is in keeping with the EASAIER project's aim of making the cultural content available in sound archives more accessible and usable. In the context discussed here, it allows that cultural content to be used as an educational source and resource for oral/aural learning, imparting to the learner invaluable and practical means of interaction with which to inform their practice.

---

### 6. References


[1] Transcribed from video interview, EASAIER prototype testing, RSAMD, Glasgow, 28 May 2007.

[2] www.easaier.org

[3] www.hotbed.ac.uk

[4] J. Harrop and F. Morton, email correspondence, August 2007

[5] S. Barrett, C. Duffy, and K. Marshalsay, "Final Report: HOTBED (Handing On Tradition By Electronic Dissemination)", RSAMD, Glasgow, 2004.

[6] Quoted in B. Taylor, "Junior Crehan of Ballymackea Beg", *Musical Traditions*, Spring 1992, Rochford, p. 29.

[7] B. Taylor, "A Contrast in Styles", *Dal gCais*, Vol. 4, 1978, pp. 98–99.

[8] J. Harrop and J. Dickson, email correspondence, September 2007.

[9] M. McCarthy, *Passing It On: The Transmission of Music in Irish Culture*, Cork University Press, Cork, 1999, p. 190.

# Using a Cognitive Model of Irish Traditional Fiddle Playing in the Development of a Teaching Aid for Improvisation


Flaithrí Neff
*Computer Science Dept.*
*University College Cork*
*Ireland*
*fn2@cs.ucc.ie*

Eoghan Neff
*School of Music*
*Cardiff University*
*Wales, UK*
*fiddling@eircom.net*

Ian Pitt
*Computer Science Dept.*
*University College Cork*
*Ireland*
*i.pitt@cs.ucc.ie*



## Abstract

*The art of improvisation is an essential part of many musical genres and can take musicians many years to master. Within Irish traditional music, improvisation is not specifically taught, even though it is fundamental to the spontaneous creativity exhibited by experienced traditional musicians. There may be many different reasons why the teaching of improvisation is not emphasized in Irish traditional music, one perhaps being a lack of understanding of the perceptual elements associated with improvisation.*

*We introduce a cognitive model that attempts to highlight the differences between experienced and inexperienced musicians during moments of improvisation in Irish traditional fiddle playing. We propose that our model gives a clearer understanding of the perceptual processes involved and hope to further develop and implement the model as an improvisation teaching aid.*


## 1. Introduction

Little has been written on improvisation in comparison to the endless volumes on its apparent counterpart - composition. Perhaps the most useful definition of improvisation involves the concept of invention and spontaneity. Invention assumes a conscious manipulation of things or events while spontaneity implies an unconscious impulse. Where the musician in a literate tradition may rely on the composer's will to present such ideas to him/her, the oral musician is held solely responsible for his/her creativity.

We acknowledge that the cognitive processes involved during an instance of improvisation are very complex and not yet fully understood. However, we believe that a useful model can be based on contemporary perceptual theory [1][2][3][4][5][6] and the constraints of the Irish fiddle tradition [7] in an attempt to design an Irish fiddle improvisation teaching aid.

The intention is that our Irish Fiddle Improvisation Model (IFIM) will sufficiently describe the general musical behaviours of an inexperienced musician during attempts at improvisation, both during solo performance and group interaction. We believe that taking the initial approach of devising a cognitive model will help us understand the human mechanisms involved before embarking on psychophysical testing and eventual software development.

## 2. Improvisation within the Irish Fiddle Tradition

We acknowledge that some of the very innovative spontaneous inventions of experienced musicians may not be accurately modeled. However, we believe that some of the more basic concepts of improvisation within the fiddle tradition can be implemented using various constraints.

Irish traditional music incorporates strict musical and aesthetic structures. Some of the constraints include:

- Maintaining strict timing
- Structure of phrasing
- Structure of the tune (i.e. AABB etc.)

Because of the physical construction of the fiddle, and the tradition of playing within the first position, a variety of constraints particular to the instrument are also present. These include:

- Triplet ornamentation using the bow
- Using the fourth finger only on the E string or during rolls
- Playing an octave lower where the lowest note of the tune is G

## 3. The Irish Fiddle Improvisation Model

The IFIM is closely based on the Auditory User Interface Model developed by Neff and Pitt [8][9] and inherits its three-block approach, with the addition of a fourth block. However, a major difference with regard to the three inherited blocks are the Schema rules which are heavily determined by the structure and traditions of Irish music, and more specifically Irish traditional fiddle playing. All other areas where model rules apply have also been oriented toward the specific task of Irish fiddle playing.

The three blocks of the IFIM include:

- The Sensory Filter
- The Subtask Attention and Inhibition Manager
- The Higher Processing Mechanism

The fourth, additional block, simply referred to as 'Musician Reaction' relates to the musician's output, as this subsequently has an affect on the Auditory Scene where the cycle begins again (figure 1).

An auditory scene consisting of many streams is filtered at various stages by the human perceptual system. This ensures that the most relevant streams gain access to the limited cognitive resources. Perceptual interference and interaction may occur at any of the stages. According to our model, the process is cyclical with the Higher Processing mechanisms also influencing both the Sensory Filter stage and the Subtask Attention and Inhibition Manager. The player's musical reaction to what he/she hears is added to the auditory scene, and therefore becomes part of the cycle.

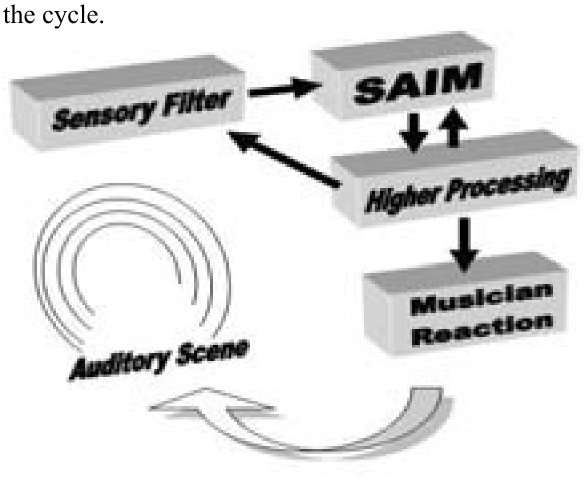



**Figure 1. An auditory scene consisting of many streams is filtered at various stages. The Higher Processing mechanism influences the Sensory Filter, the Subtask Attention and Inhibition Manager and Musician's Reaction.**

### 3.1. The Auditory Scene

An auditory scene may consist of an individual musician or an ensemble of musicians. This complex scene is primitively organized by the most peripheral of the perceptual system's processes – Auditory Scene Analysis [2]. Prior to the Sensory Filter the auditory scene is automatically segregated into auditory streams based on the various acoustic attributes of the sounds involved [2].

In relation to non-musical auditory scenes, such as those depicted in [8] and [9], the organization of streams determined by ASA rules remains robust while travelling through the rest of the model. In a musical scene however, the situation is more complex as musical organization is strongly imposed by the relevant Schema. This musical 'macro' organisation of streams over primitive organisation of streams is determined by musical experience, tradition and training. Still, it is important to point out that some of the primitive ASA rules strongly maintain their stream organisation such as those imposed by spatialization.

Just as there is competition between ASA rules when sounds are associating with two or more streams, it is also possible that ASA rules and the musical 'macro' rules of the Schema may compete. Such competition for stream segregation/association has been explored by musicians and composers of many different genres, (Bach's *Prelude for Partita Number 3* for solo violin, Javanese Gamelan).

### 3.2. The Sensory Filter

Individual streams formulated during ASA are filtered by the Sensory Filter block in the IFIM. The result is that some streams will pass through to the next stage while others will be blocked from continuing along the auditory pathway. This feature is necessary for simplifying the complex scene that is presented and is an important asset in reducing cognitive overload.

The Sensory Filter block (figure 2) is structured in accordance with the musician's Schema [5] representing a particular situation. Schema theory explains the formation of a perceptual template through experience, and so specific rules based on the musician's level of experience, training, musical knowledge and expectations of Irish fiddle playing are applied to the incoming streams previously organised by ASA. Schemas originate from the Higher Processing Mechanism and so insufficient musical experience will result in inaccurate management of incoming auditory streams at this stage. The inexperienced musician will apply an inappropriate schema resulting in fragmentary rule-sets.

The schema rules will allow content compatible with its rules to pass and may even add or subtract content from the stream based on subliminal expectations. In our model, anomalies are also allowed to pass as their segregation from the mainstream content is so distinct that they merit further cognitive investigation. New variables to the musician's schema (including anomalies) are necessary for new schemas to be constructed, or for rule-sets to be augmented through experience.

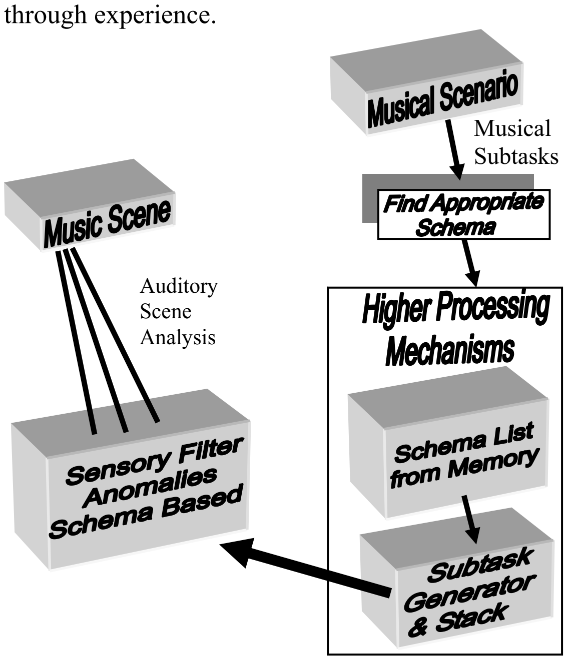



**Figure 2. The musical scene is segregated into multiple streams before reaching the Sensory Filter. The Sensory Filter is constructed under the influence of Schemas in the Higher Processing Mechanism.**

### 3.3. Subtask Attention & Inhibition

The Subtask Attention and Inhibition Manager (SAIM) further constrains the auditory streams of each musical subtask. Its primary purpose is to facilitate exclusive access to human memory for the most crucial auditory stream. Only one stream at a time has the privilege to be in Focused Attention [6], while all other streams that have managed to pass the Sensory Filter are in Peripheral Attention (figure 3). Our interpretation of this part of the human perceptual system is that Focused Attention is the exclusive gateway to human memory via the Focal Buffer (figure 3) where auditory/musical information has access to important high-level processes such as rehearsal.

Data in the Peripheral Attention component are temporarily stored in the highly volatile Peripheral Loop. This loop does not have access to human memory and therefore the information stored in the loop will not be processed accurately or retained for later use.

Both particular acoustic traits (such as sudden loudness/timbre/pitch/location change) and musical variations (expert or inexpert) draw focused attention. Streams that do not pull on focused attention are inhibited and remain in peripheral attention. The SAIM is itself a top-down process and it administers these processes.

When focused attention is gating from one stream to another, it needs time to readjust to the new stream. Therefore, a non-critical stream that pulls focused attention away from a critical stream for just a short time-period, may have a dramatic impact as the SAIM requires at least 600ms [10] to accurately swap between streams. For inexperienced musicians this effect may be emphasized.

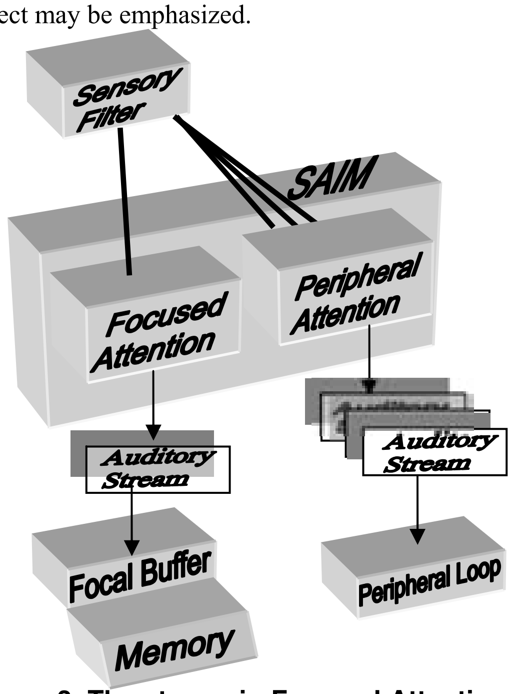



**Figure 3. The stream in Focused Attention has access to Memory via the Focal Buffer (FB). Streams in Peripheral Attention are stored in the volatile Peripheral Loop (PL).**

### 3.4 Higher Processing Mechanism

In our model, we identify the Higher Processing Mechanism as predominantly including memory-based activities such as storage, rehearsal and Schema formation. We acknowledge that many other nonmemory-based activities are also present, and that these dictate top-down influence such as SAIM.

Our fundamental view of memory is based on the Changing-State Hypothesis [3][4]. This hypothesis is constructed around the notion of competing relevant and irrelevant streams for the high-level process associated with seriation. The focus of this hypothesis is on the Irrelevant Sound Effect [3][4] but we extend this theory to include all memory-related activity.

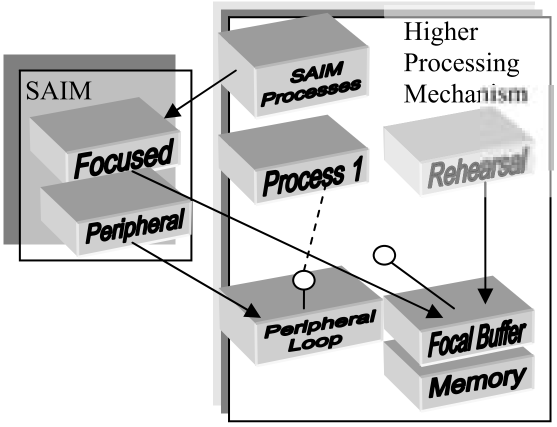



**Figure 4. The SAIM is a top-down process governing attention mechanisms. Content in both the PL and the FB may compete for the same process causing interference.**

## 4. Future Work

Working within the confines of the Irish Fiddle Tradition helps to constrain the implementation of an improvisation teaching tool. A generic improvisation teaching tool is not attainable at this stage but we hope that this is a step toward achieving that goal.

Any implementation will need to incorporate rules at every stage of the IFIM as well as rules governed by the music and instrument itself. However, it also needs to be flexible so that a user can be innovative and add new methods of improvisation to the rule sets.

We also need to consider various techniques involving realtime analysis of the student's playing. For this we will investigate the potential of Automatic Music Audio Summary Generation tools [11].

## 5. Summary

We have presented a cognitive model (IFIM) attempting to describe improvisation in Irish Traditional Fiddle playing. Our model describes the perceptual pathway from the peripheral grouping of sounds through to higher level processes involving Schema-based filtering, Attention Management and Memory activity. Along with using the constraints inherent within this particular genre of music and this particular instrument, we hope to use a combination of perceptual-based rules defined at each stage of our model and musical-based rules to devise an improvisation teaching tool for the Irish Fiddle.

## 6. References

[1] R. Vilimek, and T. Hempel, "Effects of Speech and Non-Speech Sounds on Short-Term Memory and Possible Implications for In-Vehicle Use", *International Conference on Auditory Display*, Limerick, 2005.

[2] A.S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 2002.

[3] D.M. Jones, and W. J. Macken, "Irrelevant Tones Produce an Irrelevant Speech Effect: Implications for Phonological Coding in Working Memory", *Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol. 19, No.2,* 1993, pp. 369-381

[4] D.M. Jones, C. Madden, and C. Miles, "Privileged access by irrelevant speech to short-term memory: The role of changing state", *Quarterly Journal of Experimental Psychology, 44A,* 1992, pp. 645-669.

[5] A.D. Baddeley, *Your Memory – A User's Guide*. Prion, London, 1996.

[6] S.N. Wrigley, and G.J. Brown, *A model of auditory attention*. Technical Report CS-00-07, Speech and Hearing Research Group, University of Sheffield, 2000.

[7] E. Neff, and J. O'Connell, "A History of the Irish Solo Violin Tradition", Ongoing PhD Thesis, Cardiff University.

[8] F. Neff, A. Kehoe, and I. Pitt, "User Modeling to Support the Development of an Auditory Help System", *Text, Speech and Dialogue, 10th International Conference, TSD 2007*, Springer, Germany, 2007, pp. 390-397.

[9] F. Neff, I. Pitt, and A. Kehoe, "A Consideration of Perceptual Interaction in an Auditory Prolific Mobile Device", *Workshop on Spatial Audio for Mobile Devices, Mobile HCI 07, 9th International Conference on Human Interaction with Mobile Devices and Services*, Singapore, 9th September 2007.

[10] T. A. Mondor, and R. J. Zatorre, "Shifting and Focusing Auditory Spatial Attention", *Journal of Experimental Psychology: Human Perception and Performance*, Vol .21, No. 2, 1995, pp. 387-409.

[11] G. Peeters, A. La Burthe, and X. Rodet, "Toward Automatic Music Audio Summary Generation from Signal Analysis", *Centre Pompidou, IRCAM,* Paris, France, 2002.

# Computer Analysis of the Indirect Piano Touch: Analysis Methods and Results


Aristotelis Hadjakos
*TU Darmstadt*
*telis@tk.informatik.tu-darmstadt.de*

Erwin Aitenbichler
*TU Darmstadt*
*erwin@tk.informatik.tu-darmstadt.de*

Bernhard Wetz
*HfMDK Frankfurt*
*b.wetz@inspiration-b.de*

Max Mühlhäuser
*TU Darmstadt*
*max@tk.informatik.tu-darmstadt.de*



## Abstract

*Indirect touches - touches that originate from above the key - play an important role in piano technique. Analysis methods are presented and applied in a study with piano students performing different touches in slow motion. Colored markers that were attached to the players' fingers were tracked and the angles in the joints were determined. Methods to judge the regularity of the measured movements are introduced and applied to the obtained dataset. Further, phenomena that we found in the motion graphs are discussed.*


## 1. Introduction

Analyzing the technique of a piano player, two types of touch are distinguished [4]:

- Direct touch and
- Indirect touch.

A direct touch begins with the finger in contact with the key. At the starting point of a direct touch, the finger is at rest. The finger then continuously accelerates the key. On the contrary, the indirect touch begins with the finger above the key. When the finger hits the key, it has already attained a considerable speed. This is a key difference to direct touch with implications on the sound because of noise being generated when the finger hits the key. In this paper we examine the indirect touch.

For normal touch, the finger is flexed in the knuckle (1st joint). The 2nd and 3rd joints contribute to the finger's motion by flexion or extension [4]. Henceforth, we will call a touch with flexion of the 2nd and 3rd joint a flexion-touch and a touch with extension of the 2nd and 3rd joint an extension-touch.

The remaining paper is organized as follows. In section 2 we discuss related work. Next, we describe the design goals and the approach of our touch analysis software. In section 4, we describe the study, which was made with piano students. Typical results of the analysis are shown in section 5. We present formal methods for the analysis (section 6) and apply them on the user study (section 7). Conclusion and future work sections follow.

## 2. Related work

Music via Motion (MvM) [9] is a framework that allows mappings from physical movement to multimedia events. MvM uses video tracking and other sensor technology for the acquisition of movement data.

The Conductor's Jacket [8] is used to gather data from conductors. It consists of various sensors integrated into clothing. The Conductor's Jacket measures physical motion and physiological activity. It has EMG-sensors, sensors for breathing, body temperature, galvanic skin reaction, and heart rate. Additionally, there are position and orientation sensors attached at various points.

Schoonderwaldt et al. [10] developed a bow tracking system that consists of a combination of optical tracking and acceleration sensors. The system uses EyesWeb [2] as a framework for tracking colored markers on the bow.

The Hyperbow [11] is a commercial carbon fiber bow with attached sensors. The tracking of the bow position is done by oscillators attached to the ends of the bow and an antenna at the frog of the instrument. Flexion sensors measure the force that is applied to the bow when pressing it against a string.

A commercial visual tracking system (Selspot) was used by Dahl [3] to track movement trajectories of drummers performing a rhythmical pattern. Individual movement habits were found, which the players kept consistently in all playing conditions.

The Piano Pedagogy Research Lab of the University of Ottawa uses sensor technology to analyze piano playing with the computer [1]. The researchers create tools for teaching, research, and for prevention of piano playing related health problems [7].

To measure the behavior of grand piano actions, Goebl et al. [5] attached acceleration sensors to selected keys and corresponding hammers of the examined pianos. Direct and indirect touches were executed on the prepared instruments and resulted in different behaviors of the piano action. In direct touches the motion of the hammer starts immediately with the motion of the key. In indirect touches, the hammer motion starts several milliseconds after key motion. In contrast to Goebl's work we directly examine the finger motions of the players performing indirect touches.

## 3. Analysis software

### 3.1. Design goals

The design goals for the analysis software were:

**Touch Analysis:** The software should capture the finger motion and create a graphical representation.

**Automation:** The process should be automatic and require as little human intervention as possible.

**Low Cost:** The acquisition hardware should not be expensive. A system based on off-the-shelf components would also be affordable for hobbyists.

**Extensibility:** The software should support extending the analysis. Two areas are relevant: (1) the analysis of motions originating from the wrist or elbow and (2) the refinement of the analysis by using better and more expensive acquisition hardware.

### 3.2. Approach

The finger motion of the piano player is recorded as a video. To enable tracking, colored markers are attached to the player's finger. The video is analyzed in three steps. First, the positions of the markers are tracked. This is done by *MotionAnalyzer*, which was developed for this purpose. The second step is to compute the angles between phalanges of the finger with *AngleExtractor*, a command line program. Third, a graphical representation is generated. This is done with *ToDat* that generates a file that is used as input for Gnuplot.

### 3.3. MotionAnalyzer

*MotionAnalyzer* (see figure 1) is used for tracking colored markers attached to the player's finger. *MotionAnalyzer* displays a still image from the video. The user first defines the reference colors by clicking


**Figure 1. MotionAnalyzer**

on the markers to be tracked in the image. The program will then search for these markers and extract their coordinates. The user can also adjust the sensibility of the recognition for each color individually.

In each still image, which is extracted from the video, *MotionAnalyzer* searches for the marker colors. It computes the Euclidian distance of each pixel to the defined reference colors taking into account the RGB components of the color. If the Euclidian distance falls below a threshold, which the user defines for each color individually, the pixel is grouped to the reference color. The median of the x- and y-coordinates of the so collected pixel is than the recognized position. *MotionAnalyzer* was implemented using C++ and DirectShow on a Windows XP platform.

### 3.4. AngleExtractor

*AngleExtractor* computes the angles of the finger joints from the data provided by *MotionAnalyzer*. *AngleExtractor* computes the angle between two line segments that are formed by the positions of three successive markings. The order of the markings is given by the order in which they were defined in *MotionAnalyzer*. An additional angle is computed at the position of the first marking: It is the angle between the line segment of the first and second marking and an imaginary horizontal line intersecting the first marking. See figure 2 for an example of the computed angles.

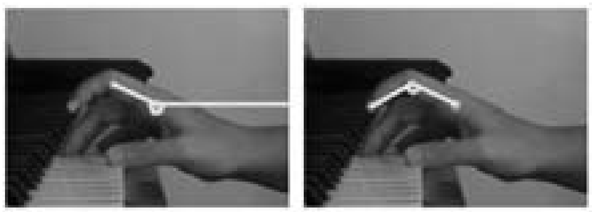


**Figure 2. Measured angles**

## 4. Experiment

A user study with five piano students was conducted at the HfMDK Frankfurt. The students were recorded playing indirect flexion- and extension-touches. They played the touches with the index and little fingers of both hands resulting in 40 samples of data. A Canon Ixus 30 digital camera was used, which provided a video in MJPEG format with a resolution of 320x240 pixels at a temporal resolution of 60 fps. The students were asked to slow down the motion artificially because the camera does not have enough temporal resolution to capture the motion at original speed. A metronome was used to make sure that the students executed the motions with a defined rhythm and to hint on the velocity of the motion. Markers were attached to the knuckle (1st joint), the 2nd joint, and the fingertip. The two resulting angles were measured (see figure 2). Because hand and arm are at rest, there is no need to add markers behind the 1st joint.

## 5. Observation

The figures generated from the angle measurements visualize two angles as a function of time. The upper graph represents the angle in the 1st joint; the lower graph represents the angle in the 2nd joint. Figure 3 shows a series of flexion-touches. It begins with the finger resting on the key. After a short time, the player lifts the finger and the angles in the 1st and 2nd joints increase. After reaching the maximum height, the angles in the joints decrease again until the finger hits the key.

Figure 4 shows a series of extension-touches. It starts with the finger resting on the key. After a short time, the player lifts the finger. While the angle in the 1st joint increases, the angle in the 2nd joint decreases. After the finger reaches the maximum height, the angles in the 1st joint decrease while the angles in the 2nd joint increase.

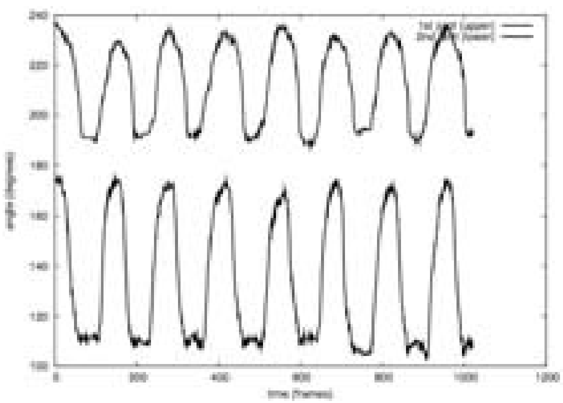


**Figure 3. Flexion-touch**

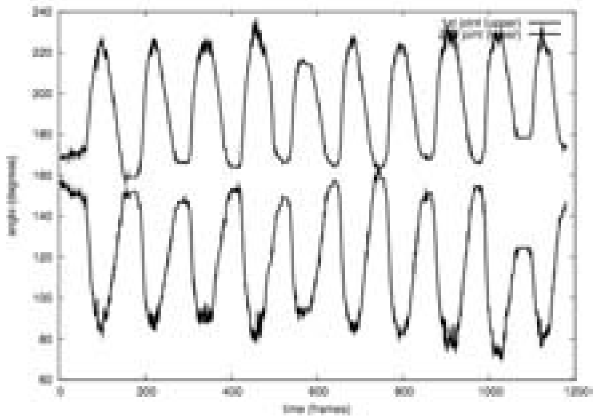


**Figure 4. Extension-touch**

## 6. Analysis Methods

### 6.1. Levels and ways

The gathered data was analyzed with formal methods. Before these methods can be applied, the motion curves have to be segmented. The following segments are distinguished (see figure 5):

- Preparation way,
- Preparation level,
- Hit way, and
- Hit level.

Preparation level and hit level are phases of relative stability. There is only little motion and the present motions extinguish each other. The hit level occurs when the finger hits the key. The preparation level occurs when the finger reaches its end position above the key. The preparation way is the transition from hit level to preparation level. The hit way is the transition from preparation level to hit level.

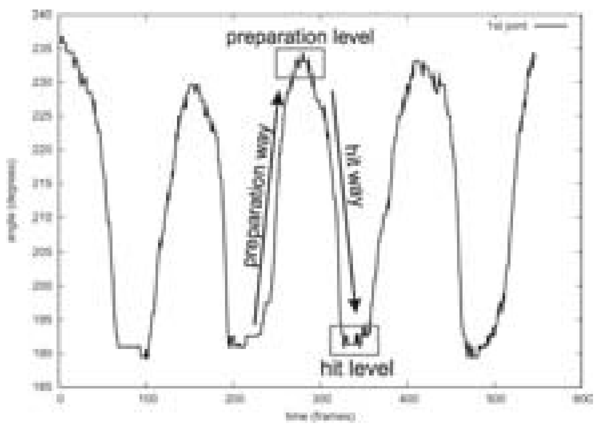



**Figure 5. Segments**

### 6.2. Properties of levels and ways

The height of a preparation level is the maximum angle achieved in the preparation level. The height of a hit level is the minimum angle achieved in the hit level. The above definitions of height apply for the 1st joint executing either flexion-touch or extension-touch and for the 2nd joint executing flexion-touch.

For the 2nd joint executing an extension-touch the preparation and the hit levels are vertically flipped because the hit level occurs when the 2nd joint is fully extended. For this case, the height of a preparation level is the minimum angle achieved in the preparation level and the height of the hit level is the maximum angle achieved in the hit level.

The length of the hit way is the difference between the heights of the connected levels.

### 6.3. Equality and translation measures

Equality and translation measure, which are defined in this section, are aids for judging the regularity of a series of touches.

The equality measure (E) is defined as the fraction of the shortest hit way to the longest hit way of a series of touches (see figure 6). For the translation measure (T) the heights of four levels of a series of touches have to be considered:

- Lowest hit level,
- Highest hit level,
- Lowest preparation level, and
- Highest preparation level.

The translation measure is the fraction between the minimum distance between a preparation and hit level and the maximum distance between a preparation and hit level (see figure 7).

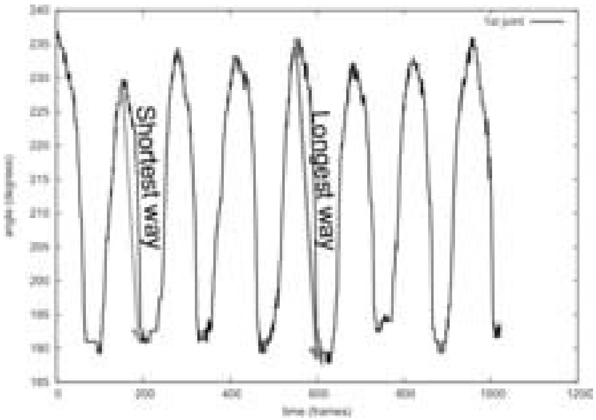



**Figure 6. Equality measure**

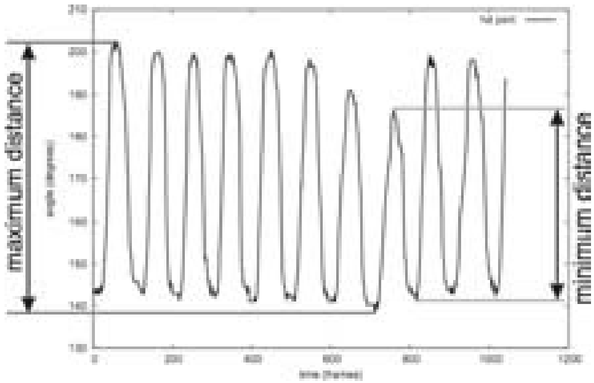



**Figure 7. Translation measure**

If the motion in a joint reaches similar preparation level and hit level heights in a series of touches, the equality measure of that joint's movement will have a value close to 1. However, an offset the preparation and hit level height of the same amount is not detected by the equality measure. It is, however, detected by the translation measure. The combination of equality (E values) and translation (T values) measure has implications for the analyzed motion as can be seen in table 1.

**Table 1. Implications of E and T values on the analyzed motion**

| | E big | E small |
|---|---|---|
| T big | Regular motion | Impossible, $T \leq E$ |
| T small | Translation | Irregular, possibly also translated |

To support the analysis process, the *ET* program was developed. It calculates the E and T values given the preparation and hit levels. The user provides this information by marking the preparation and hit levels with bounding boxes in the GUI of the *ET* program.

## 7. Analysis

### 7.1. Quantitative analysis

In our dataset of 40 samples, the motion of the 1st joint tended to be more regular than the motion of the 2nd joint in both, flexion-touches and extension-touches. In 80% of the cases the E value of the 1st joint was higher than the E value of the 2nd joint. In 87.5 % of the cases, the T value of the 1st joint was higher than the T value of the 2nd joint.

Extension-touches and flexion-touches of each finger were compared, giving 20 pairs to be considered. In 80 % of the pairs the T value of the 2nd joint was higher when executing flexion-touches. However, only in 60 % of the pairs the E value of the 2nd joint was higher when executing flexion-touches. Overall, in the combination of E and T values, it seems that the flexion-touch motion could be more regular.

A list of the measured E and T values and more details about the analysis process can be found in [6].

### 7.2. Empirical evaluation

In some graphs of the 2nd joint executing a flexion-touch phenomena could be seen in our dataset: enter-drop, leave-drop, early movement and complete irregularity.

An enter-drop occurs if the angles drop below the hit level before returning to the hit level again (see figure 8). This phenomenon is called enter-drop because it occurs when the hit level is entered.

A leave-drop occurs if the angles drop below the hit level when the preparation should begin (see figure 8). This phenomenon is called leave-drop because it occurs when the hit level is left.

A graph with distinct enter-drops and leave-drops can be seen in figure 9.

Early movement of the 2nd joint occurs if the beginning of the movement of the 2nd joint precedes the movement of the 1st by a substantial amount (see figure 9).

Some graphs of flexion-touches of our dataset show strong enter-drops and early movements of the 2nd joint, e.g. the graph in figure 9. The movement can be described as follows:

1. The finger is in preparation level. The 2nd joint is fully stretched.
2. The finger is flexed in the 2nd joint (early movement) while the first joint stays in rest.
3. The flexing of the finger in the 1st joint begins.
4. While the finger is still considerably far from the key, the 2nd joint has already reached the minimum angle.
5. The finger is stretched in the 2nd joint and flexed in the first joint until the finger reaches the key.

The described motion is not a correct execution of a flexing-touch. A beginning flexion-touch is aborted in favor of an extension-touch.

Some graphs of the 2nd joint were so irregular that the preparation and hit levels could not be identified (see figure 10).


**Figure 8. Enter-drop and leave drop**

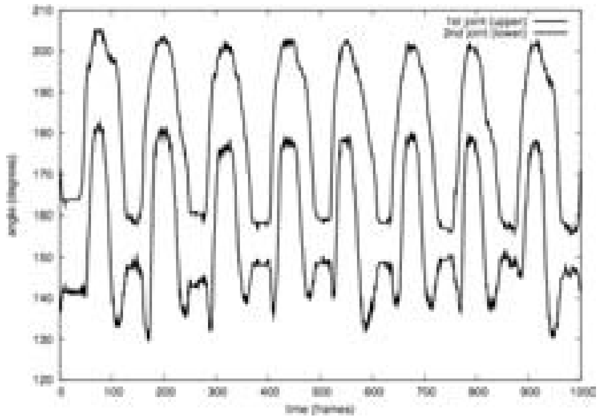


**Figure 9. Enter-drops, leave-drops, and early movement**

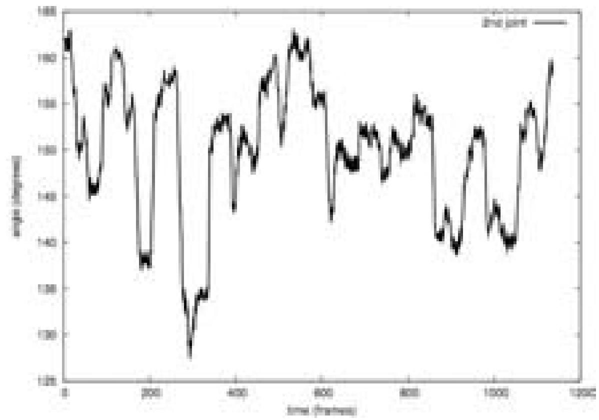


**Figure 10. Complete irregularity**

## 8. Conclusion

By tracking visual markers attached to players' fingers we calculated the angles in the joints and visualized them. For analyzing the graphs we introduced the E and T values that can be used for estimating the regularity of the movements. For the calculation of the E and T values, the graphs were segmented to preparation level, hit level, preparation way and hit way. Properties of these segments were defined.

The methods introduced in this paper help to interpret motion graphs and give a judgment about the regularity of the motion. Although our study can be expanded, for example by using a high frame rate camera, important tools for the analysis of the indirect piano touch were introduced and can serve as a basis for further research.

## 9. Future work

Our approach can be extended towards online generation of the graphical representations. If the graphs were generated in real-time they could serve as direct visual feedback about the regularity of the touches and could be used in a pedagogical setting.

If we could distinguish flexion-touches and extension-touches automatically and in real-time, this could be used to implement a special electronic piano. The flexion- and extension-touches would have different timbres. This piano could be useful for learning and teaching the different touches and as an instrument with an additional expressive parameter.

## 10. References

[1] Birkett, S., G. Comeau, M. Frize, C. Herry, R. Know, and P. Payeur, Challenging 300 years of piano teaching practices with 21st century technology: New teaching and research tools in piano pedagogy, Round table conducted at the Canadian University Music Society (CUMS) Conference, London, Ontario, 2005

[2] Camurri, A., P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Richetti, A. Ricci, and G. Volpe, Toward real-time multimodal processing: EyesWeb 4.0, AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, 2004

[3] Dahl, S., Playing the accent – comparing striking velocity and timing in an ostinato rhythm performed by four drummers, Acta Acustica United with Acustica vol. 90, 2004

[4] Gat, J., The technique of piano playing, 2nd edition, Collet's Holding, London, 1965

[5] Goebel, W., R. Bresin, A. Galembo, Touch and temporal begavior of grand piano actions, Technical Report, Austrian Institute for Artificial Intelligence, Vienna, 2005

[6] Hadjakos, A., Computeranalyse des Leggieroanschlags, Diploma Thesis, HfMDK Frankfurt, 2007

[7] Herry, C., L., M. Frize, R. A. Goubran, and G. Comeau, Evolution of the surface temperature of pianists' arm muscles using infrared thermography. Paper presented at the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Shanghai, 2005

[8] Marrin, T. and R. Picard, The Conductor’s Jacket: a device for recording expressive musical gestures, Proceedings of the International Computer Music Conference (ICMC 1998), Ann Arbor, USA, 1998

[9] Ng, K. C., Music via motion: transdomain mapping of motion and sound for interactive performances, Proceedings of the IEEE vol. 92(4), 2004

[10] Schoonderwaldt, E., N. Rasamimanana, and F. Bevilacqua, Combining accelerometer and video camera: Reconstruction of bow velocity profiles, Proceedings of the International Conference on New Interfaces for Musical Expression (NIME06), Paris, 2006

[11] Young, D., The Hyperbow controller: real-time dynamics measurement of violin performance, Proceedings of the International Conference on New Interfaces for Musical Expression (NIME02), Dublin, 2002

# Requirements and Application Scenarios in the Context of Network Based Music Collaboration

Chrisoula Alexandraki
*Department of Music Technology & Acoustics, Technological Educational Institute of Crete*
*chrisoula@stef.teicrete.gr*

Ioannis Kalantzis
*AKMI Vocational Training School*
*Athens, Greece*
*ikalantzis@hotmail.com*

## Abstract

*Recent advances in computer network technology have greatly enhanced the feasibility of networks that allow remote collaboration in performing music. This paper presents a research study on the user and technical requirements for systems in this context. User requirements have been gathered through a questionnaire-based survey, whereas the reported technical ones are the result of a qualitative study on the relevant research projects and the existing technological tools in the area of live streaming of multimedia content. Furthermore, the paper attempts one step further, by classifying the effective application scenarios that can emerge for remote music collaboration, when the reported requirements have been met.*

## 1. Introduction

The growing need for innovative network-collaboration environments for live music performance has been a challenging field for a number of academic and research institutions all over the world [1, 2, and 3]. An overview of the music and sound art projects involving the use of network infrastructures can be found at [4]. According to this article, the advent of computer network music dates back to the 1970s, when the commercialization of personal computers in the United States began.

Currently, the latest advancements in the field of broadband networking and of computer technology in general, have allowed for a variety of music collaboration scenarios to be considered feasible not only in research, but also in a commercial context. It is worth noticing for example, that live streaming of multimedia content is becoming so apparent that scenarios of network music collaboration are used by network providers to advertise the quality of the services they provide. Such scenarios, usually involving a popular Greek performer, have been used in a number of TV commercials in Greece.

In practice however, using computer networks for music collaboration is not trivial. The effectiveness of such attempts depends on various factors that range from the quality of service (QoS) provided by the underlying network, to a number of psychophysical, perceptual and artistic aspects [2]. Furthermore, the success of such experiments is strongly dependent on the means provided to the user in order to interface with the environment and communicate with other performers.

In this paper we attempt to enumerate the requirements of network based music collaboration environments and classify the application scenarios that emerge in this context.

## 2. Research context

The study reported in this paper has been carried out, in part requirement of a Greek national research project, which is currently in progress. The title of this project is “DIAMOUSES – distributed interactive communication environment for live music performance”.

The main objective of the DIAMOUSES project is the development of an integrated platform, which will allow for remote collaboration throughout a distributed live music performance environment. Musicians-members of an orchestra, whilst geographically spread, will be able to simultaneously perform the same piece of music. At the same time, this ‘network-performance’ will be witnessed by an audience located elsewhere, breaking the barriers set by geographical distance, thus resulting in a new network collaborative community.

The system under development will support signal transition in heterogeneous computer networks, including IP networks as well as a pilot DVB-T network platform which operates in the island of Crete. The combination of these two types of networking allows for simultaneous support of various routing schemes such as broadcasting, multicasting and unicasting. Moreover, it enables application scenarios which involve a broad range of target users with diverse skills and preferences, such as digital TV subscribers for interactive and non-interactive television services.

## 3. Research methodology

In this section of the paper we present the methodology which was adopted for performing the study whose results are reported in the sections that follow. The objective of the study was to define a set of requirements that must be met in the context of network based music collaboration. These requirements concern the ones set forth by users and also the technical ones for performing music through networks effectively.

In respect with user requirements, we followed a quantitative approach, by performing a questionnaire-based survey. This survey was targeted towards two groups of potential users of our system. The first group was concerned with users that have a high level of involvement in music. The users of the first group were performers, composers, conductors, instructors, as well as recording engineers and professionals from the area of music technology. The second group of users took into account the general public, which can act as an audience of a distributed music performance, having a general interest in music.

Audience involvement in distributed music performance has been taken into account since the early experiments of network performance. However, to the authors' awareness, these experiments silently assumed that all members of the audience were to be situated at the same location and therefore occupy a single node in the network, where high quality video projections and an appropriate sound reproduction system were provided [2]. In our analysis, we additionally consider the situation in which not only the various musicians, but also the members of the audience can be distributed in different locations (e.g. in the area of coverage of a broadcasting network infrastructure, such as a digital TV network).

The technical requirements were approached through a qualitative study which involved literature review, study of the relevant standardized technologies (e.g. RTP/RTSP protocols), and hands-on evaluation of the existing software tools that have been implemented in the area of live streaming of multimedia content.

## 4. User Requirements

This section presents the user requirements collected by sending questionnaires to potential users of the system under development. Two types of questionnaires were distributed: one for users actively involved in music and one for the general public which can be thought of as the audience of the distributed music performance.

Each question has a number of alternative responses. Users were asked to give a preference value to each response. So if for example a question had 3 alternative responses, then users would give a preference value 3 to the alternative response of their top preference, a preference value 2 to their second preference, and so on. According to the preference values, a normalised average was calculated for each alternative response-j, as follows:

$$W_j = \frac{P_j}{\sum_j P_j}$$

, where $P_j$ is the sum of the products of the preference values given, multiplied by the number of users that have assigned the particular value, for all preference values of alternative response-j. The analysis of the user requirements is based on the normalised average values $W_j$, which were calculated for each response. These values are given as a percentage in the diagrams that follow.

The analysis of the results takes into account aspects which are vertically related to the requirement in question. For example, questions regarding preferences in performing music are arranged according to music genre.

Each of the questionnaires was accompanied by a cover letter which was featuring the context of the research study and introducing the users to the concept of remote music collaboration.

### 4.1. Users actively involved in music

In this group of users a total of 58 replies was received. Requirements were classified according to the users' type of involvement in music and according to music genre. The form of the questionnaire was such that a user could have more than one type of involvement in music. However, if somebody was involved in more than one music genres, a separate questionnaire for each genre had to be completed. The following table shows the distribution of users among different music genres.

**Table 2: Distribution of the music genres**

| Music Genres | No. | Perc. |
|---|---|---|
| Classical/Contemporary | 22 | 38% |
| Jazz/Blues | 8 | 14% |
| Pop/Rock | 11 | 19% |
| Electronic/Electroacoustic | 10 | 17% |
| Ethnic/Folk | 7 | 12% |

The rest of this section is structured as follows. Firstly, we provide an English translation of the questions in the questionnaire, as these were originally formulated in Greek. Following, is the diagram which depicts the average values ($W_j$) of the declared preferences for each alternative answer and for each music genre. Finally, some observations on the resulting diagram are provided.

**Question1:** Give your preference in musical instruments and musical interfaces when performing music.

a) Acoustic instruments

b) Electric instruments

c) Electronic instruments

d) Computer (interaction solely through mouse or mouse pad and keyboard)

e) MIDI controllers (keyboards, sliders, knobs, etc.)

f) Experimental sensors for gesture recognition

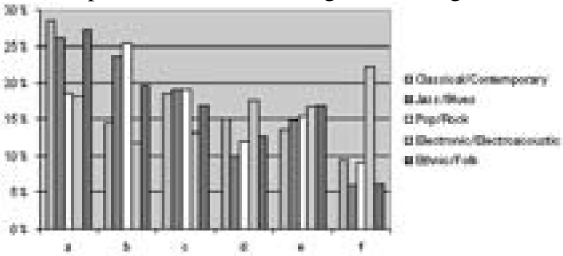


**Figure 1: Musicians' preference in musical instruments and musical interfaces**

As expected, the top preference is in acoustic instruments for all music genres, apart from musicians of pop/rock who prefer electric instruments, and musicians of electronic/electroacoustic music who prefer experimental sensors. It is interesting to notice that a) experimental sensors are top priority for musicians of electronic/electroacoustic music, and b) the use of MIDI controllers is almost equally preferred by all music genres.

**Question 2:** Rate your preference in deciphering the flow of a musical piece while performing with others.

a) Through a musical score

b) Performing from memory

c) Prima vista or performing according to a score that is dynamically generated

d) Performing musical patterns based on your choice or on indications by others

e) Performing a score comprised of predefined graphical symbols

f) Improvising on a given musical theme

g) Free improvisation based on movement or eye contact

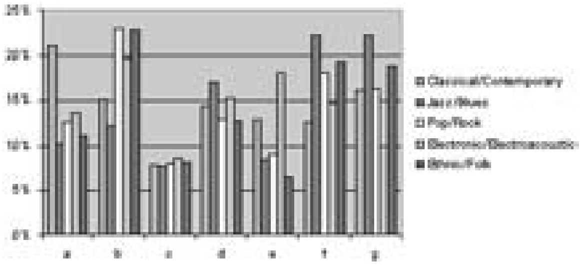


**Figure 2: Preference in deciphering the flow of a musical piece**

This question was included in the questionnaire in order to indicate requirements on the graphical user interface provided in circumstances of distributed music performance. It can be inferred from the diagram that musicians of classical and contemporary music have a strong preference in the presence of a score whereas jazz and folk musicians show a preference in improvisational music. It is interesting to notice that musicians of electronic/electroacoustic music would prefer to memorise the piece, rather than have to use any means for supporting them in following the flow of the music.

**Question 3:** Rate your preference in trying to synchronize with the other performers.

a) Conductor

b) Metronome

c) Visual metronome (usually a light, which flashes according to tempo and rhythm)

d) Score scrolling

e) Arithmetic visualization of tempo and rhythm (e.g. tempo: 120, bar: 27, rhythm: ¾, second quarter, would result in something like '120 27 ¾ 2')

f) No means of synchronization other than auditory and visual contact

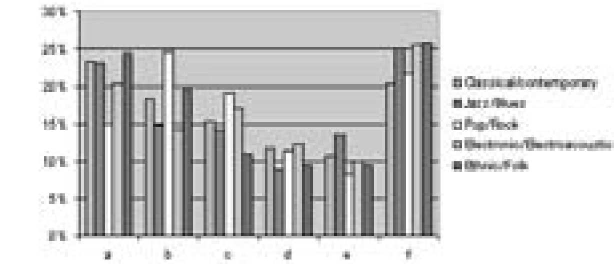


**Figure 3: Preferences in synchronising with the other performers**

All music genres show a very strong preference in visual contact with the other performers, which – in the perspective of a distributed performance – implies that video communication should be provided among the musicians. Another interesting conclusion is that musicians of Pop/Rock prefer the metronome more than any other means of synchronization. This should be provided as a utility of the client software when performing pop music in a distributed environment.

**Question 4:** Rate your preference in the sound reproduction system for listening the other performers in the absence of visual contact.

a) Headphones
b) Loudspeakers
c) Multichannel audio

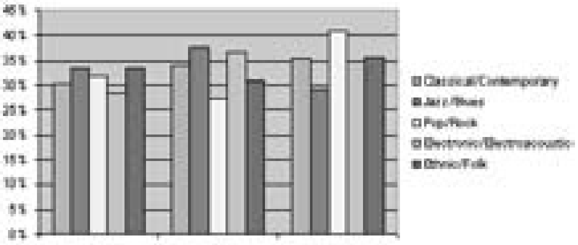



**Figure 4: Preference in the sound reproduction system**

It appears that there is a slight preference for multichannel audio. Although the majority of users questioned did not have an experience in distributed performance, it seems that musicians want to hear music reflected from the surrounding area, as it would do in a concert hall. There is strong evidence in prior experiments that sound reflections are desirable in this context [2].

**Question 5:** Rate your preference for special monitoring facilities in the absence of visual contact with the other performers

a) Monitor the dry mixed signal from participants
b) Monitor the mixed signal from participants after audio effects processing (e.g. reverberation)
c) Listen to one performer at a time with the possibility to choose another performer whenever needed (dry signal)
d) Listen to one performer at a time with the possibility to choose another performer whenever needed, after audio effects processing.

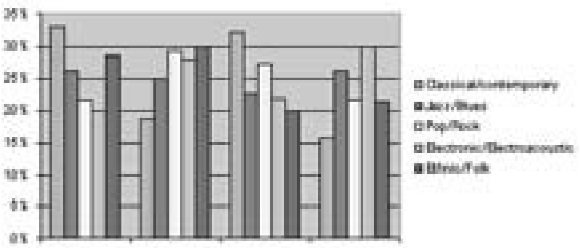



**Figure 5: Preference in sound monitoring**

In this diagram a preference for listening to all participants at the same time (mixed signal) is apparent for all music genres apart from the electronic and electroacoustic music. Furthermore, it appears that musicians of this genre find the presence of audio effects necessary, in contrast to musicians of classical/contemporary music who prefer to hear the dry signal.

**Question 6:** Suppose that you are remotely located from the other performers and that you are able to have visual contact with them through digital video. Rate your preference in the video communication provided.

a) One-way visual communication with the conductor
b) Bilateral visual communication with the conductor
c) One-way visual communication with one of the other performers, with the possibility to view another performer whenever needed
d) Bilateral visual communication with one of the other performers, with the possibility to view another performer whenever needed

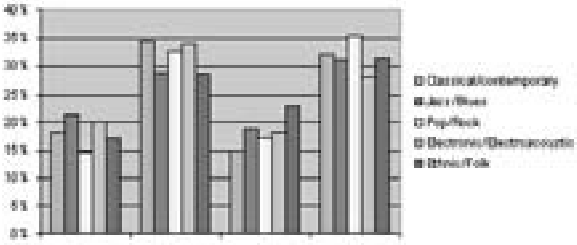



**Figure 6: Preferences in visual communication**

There is an obvious preference for bilateral visual communication for all music genres. The musicians of classical/contemporary and electronic/electroacoustic music prefer to have visual communication with the conductor than with the other musicians, which is not the case for the other music genres.

### 4.2. Members of the audience

Although users of this group were asked to rate their preference in different music genres, the analysis of their requirements is not arranged according to genres. The reason for this is that the audience have a more passive role than musicians who affect the outcome of a distributed performance scenario. This section will concentrate on the results of the survey, without getting in detail in formulation of questions or statistical data.

A total of 35 completed questionnaires was received, which were arranged according to users' education level and the kind of music of their top preference. Users were more or less evenly distributed among the different music genres. The provided questionnaire form allowed them to declare their favorite music genre if this was not included in the list provided. The answers in this field were the genres of Heavy Metal, Soul, Disco and Byzantine-Hymnology. The educational level of the users ranged from school graduates to PhD holders, with the majority of users holding a university degree.

Users were introduced to the concept of remote distributed music performance and they were asked about their preference in the following aspects: facilities for watching a performance, sound reproduction system, video information content, metadata provided, provision of video on demand services and provision of event rating services. Finally users were prompted to comment on the concept of distributed music performance and give their own suggestions.

Regarding facilities for watching a distributed performance, users exhibited equal preference in the alternatives provided, which were a computer terminal, a home television or a centralized screen projection. The preferred sound reproduction system appeared to be the multi-channel system instead of conventional stereo sound reproduction systems, with a higher preference in surround speaker systems (of type 5.1 or 7.1), although polyphony (e.g. 8-speaker system) was provided as a separate option. In respect with the content of the video information, users seemed to be interested in having the possibility to choose when to view each of the distributed performers alone and when to view all of them on separate frame portions of the same display.

The interest in metadata information about the performance and the music performed was rated as shown in figure 7. It can be seen that users are more interested in having information about the music performed, rather than having information about the performance itself or the performers.

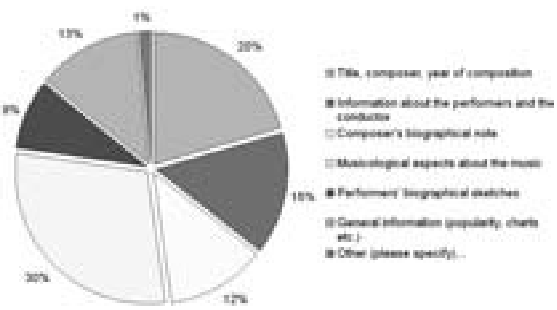



**Figure 7: Audience preferences in the information content of the provided metadata**

Users were also asked about their interest in a video on demand service which was related to the performance and could be offered to them, and a majority of 51% were highly interested. Finally, there was a higher interest in having the possibility to rate the performance in relation to its artistic aspects, rather than in relation to its technical coverage and the underlying technology.

## 5. Technical requirements

According to a number of scientific articles ([3] & [5]), real-time audio streaming is one of the most intensive applications in networking. The technological innovation of applications for network based music performance has been somewhat discredited due to the broad proliferation of teleconferencing technologies. However in music, accuracy in time and quality of the information delivered is a lot more crucial than in teleconferencing applications.

In respect with the network infrastructure, in order to accomplish network-based music collaboration a high level of QoS must be ensured, which requires cooperation at all network layers so as to minimize delay and quality variation of the information delivered. There have been a few scientific publications which enumerate the technical requirements in network based music collaboration. In this paper we will concentrate on latency sensitivity, bandwidth demand, synchronisation and error susceptibility.

### 5.1. Latency sensitivity

There are a number of factors causing latency in delivering live data streams in distributed music collaboration scenarios. These are due to the hardware equipment, the software applications involved, the operating system and the network infrastructure. If we concentrate on transmitting raw PCM audio streams and simplify the process of signal transmission between two participants, then we can identify causes of latency in the entire lifecycle of a data packet. Specifically, in a one-to-one transmission the lifecycle of this packet will involve the following steps: data capturing (analogue-to-digital conversion included), data packetisation, network transmission, data depacketisation and finally data playback (including digital-to-analogue conversion). What is more, an additional delay is caused by the process of loading the data buffer, which should be of adequate size in order to follow the above procedure and get reproduced at the receiver's playback equipment without producing additional distortion.

It appears to be a good analogy and has been suggested in a number of publications in this area, that the target of maximum tolerable round-trip delay ought to be comparable with the amount of the acoustic latency produced due to physical separation. Estimating latency according to the speed of sound in dry air (344m/sec) and assigning the spatial separation of musicians a value of the order of 10m result in a tolerable delay of approximately 30 milliseconds. According to prior evaluations and psychoacoustic experiments, this value is highly dependent on the music performed and the performing schema. A 20 to 30 millisecond delay is tolerable for traditional ensemble performance although this value will vary depending on the tempo of the music performed ([2] & [6]), as well as the acoustic properties and in particular the timbre of the musical instruments involved [2].

## 5.2. Bandwidth demand

Bandwidth demand is directly related to the information content of the transmitted data. Network music collaboration, may require apart from audio, also video transmission and possibly other types of information content (MIDI data, or gesture data, etc.)

In the case of audio information, transmission of CD-quality audio requires a data rate of 1.4Mbps. When employing multi-channel or better quality of audio (e.g. sampled at 48, 96 or 192 kHz, or providing 24-bit resolution), bandwidth demand is further increased. Therefore, it seems reasonable to find ways to minimise data overload for live audio streaming. In this direction, two main approaches are being discussed: audio compression and alternative encodings for representing sound and music.

It has to be taken into account, that lowering the bandwidth of sound information has major drawbacks, either in the quality of the reproduced sound or in the overall latency. For instance, sound compression algorithms that achieve sufficient compression ratios with decent audio quality result in a significant delay overhead, especially during the encoding process [7]. At the other end of the spectrum, a number of possibilities appear for low-bitrate representation of sound information, such as the conventional MIDI streams or the more recent OpenSound Control protocol, the standard for MPEG-4 Structured Audio and the IEEE standard for Symbolic Music Representation in MPEG [8]. The disadvantage in these approaches is that they cannot reproduce expressiveness in performing music, and that they are not appropriate for all types of music. Vocal music can be considered as an example.

In addition to sound information and according to the user requirements presented in this paper, it appears that video information is also necessary for remote music collaboration. Video information has two major advantages in this context. The first is related to the fact that video information can be recognisable, even when it has very low quality. For example, the Simple Profile of MPEG-4 Video supports bitrates, which are as low as 64kbps. The second advantage in employing video data is concerned with the directness of visual information in communication. The need for visual communication is evident in the user requirements section of this paper. Furthermore, the example of large orchestras, where performers synchronise by watching the conductor should be considered as a proof of the directness of visual communication. In this case, the delay of the visual information from the conductor to each of the performers is practically zero. In the context of the DIAMOUSES project, we are adopting an approach, in which musicians will receive low-fidelity video, for communicating with each other and the audience will receive high quality video. Sending high quality video to the audience is made feasible due to the fact that communication with the audience does not have to be synchronous.

## 5.3. Synchronisation

In respect with network based music collaboration, synchronisation refers to the time adjustments which need to be made when multiplexing multiple streams of audio or video data. There are two preconditions for achieving this type of synchronization. The first is that the clocks of the participants must agree with great accuracy and the second is that timing information must be sent along with the network stream.

The suggested approaches for synchronizing the clocks of multiple participants in a network music performance are to synchronise either by using the Network Time Protocol (NTP) [3], or via GPS signals [2]. The first solution offers an accuracy of 200μsec under optimal conditions in a LAN and a few milliseconds in WANs. The GPS solution offers an accuracy of approximately 10μsec or better. However, even if synchronizing the connected participants through the network, one must take into account clock inaccuracies caused by the operating system itself. This is in fact the main reason why some operating systems are considered inappropriate for network music performance.

Timing information sent along with the data packet can be ensured by the network protocols that operate at the application layer of the computer network. For example, protocols that are normally used in multimedia streaming (e.g. RTP/RTSP) ensure the delivery of NTP timestamps included in the header of the network packet, as a built-in functionality.

When the above conditions are met, synchronising multiple streams is only a matter of calculation.

### 5.4. Error Susceptibility

Sound information is particularly sensitive to errors. The major cause of transmission errors is packet loss over the network. Errors due to lost packets are inevitable while at the same the strict requirements in minimizing all sorts of latencies renders the task of data correction even more complicated.

Most applications that involve network based music collaboration facilitate the UDP protocol. Although a fast protocol, UDP offers no guarantee for the reliability of the data delivered, as packets may arrive out of order, appear duplicated, or go missing without notice. However the RTP protocol, which operates at a network layer above UDP, offers mechanisms for detecting packet loss. Such a mechanism is the provision of the 'RTP sequence number' (i.e. the index of the packet), which is included in the network packet and is increased by one for every new RTP packet.

In cases of excessive packet loss, there has to be a mechanism, which will compensate for this loss. As presented in article [9], data correction algorithms can be classified in two main categories: Automatic Repeat Request (ARQ), which requires retransmission of the lost packet, and Forward Error Correction (FEC), which is based on transmitting redundant information along with the original information. Obviously, ARQ mechanisms are not acceptable for live audio applications over the network, as they dramatically increase the end to end latency. However, FEC data correction algorithms have been used in network music performance before, as they offer data reliability, without causing significant overhead on the overall latency and the required bandwidth ([2] & [3]).

## 6. Application Scenarios

Different application scenarios, or different variants of application scenarios put forward different requirements, both from the perspective of the user and the one of the technological infrastructure needed to support the specific scenario. For instance, a piano master class distributed within a Campus Area Network (CAN), will have different requirements from a piano master class distributed among different continents (WAN), both from the perspective of instructor-to-student communication and the one of the underlying network infrastructure.

In this context, an application scenario may be formed by assigning attribute values to a number of parameters. These parameters will be referred to as 'interaction parameters' hereafter, due to the fact that they affect the type of interaction in an application scenario for remote music collaboration. This section follows by attempting to provide an overview of all the interaction parameters that comprise an application scenario for network-based music collaboration and which can have a direct impact on the requirements which need to be satisfied.

**Figure 8: The interaction parameters that comprise an application scenario for network-based music collaboration**

Obviously, one of the most determinant parameters is the operational intent of the scenario, namely the purpose of the event. Different requirements are raised in the context of a live concert, than in the context of a master class. As for recording in a remote studio for example, strict requirements are posed in terms of bandwidth and tolerable data loss. Although user roles are related to the operational intent, they are included in the above figure as a separate node because different user roles raise different requirements in the interaction environment. It was apparent from the user requirement analysis preceded, that user roles, similarly to music genres, significantly affect the requirements of the application scenario.

As mentioned before, different types of information content results in different requirements on the available network bandwidth. In the above figure, the term 'control data' is used to refer to the various alternative representations for sound and music that were mentioned at the section related to bandwidth demand. The interaction parameter 'networking' is included as a separate interaction parameter, because it is directly related to the type of services that can be supported in a certain scenario. Additionally, networking affects the scalability of application scenarios, not only in terms of their geographical spread (e.g. LAN or WAN) but also in terms of the number of participants that may be supported by the infrastructure without causing network congestion (e.g. DVB vs. WiFi).

## 7. Conclusions and future work

In this article, we presented an overview of the requirements for environments that enable network-based music collaboration. Although requirements in this context have been previously reported for specific research efforts, we targeted towards a more generalised approach that takes into account the majority of the variations that exist in distributed music performance scenarios.

The requirements study, as well as the unraveling of the possible variations of an application scenario for remotely performing music, is a part of a larger research project. In this project, DIAMOUSES, three of the possible scenarios have been selected for evaluating the system under development. We expect that user and expert evaluation of the selected scenarios will enlighten valuable findings in the area of network-based music collaboration.

## 8. Acknowledgements

The DIAMOUSES project is being implemented in the context of the Regional Operational Programme of Crete 2000 – 2006 and it is co-funded by the European Regional Development Fund (ERDF) and the Crete Region, coordinated by the General Secretariat for Research and Technology, of the Ministry of Development of Greece. The partners of the DIAMOUSES consortium are: Department of Music Technology and Acoustics, Technological Educational Institute (TEI) of Crete – Project Coordinator; Department of Applied Informatics and Multimedia, TEI of Crete; Department of Electronics, TEI of Crete; Department of Computer Engineers and Informatics, University of Patras; FORTHnet S.A.; AKMI, School of Vocational Training.

## 9. References

[1] J. Lazzaro and J. Wawrzynek, “A case for Network Musical Performance”, *in Proceedings of ACM NOSSDAV '01*, Port Jefferson, NY, June 2001, pp. 157–66.

[2] A.A Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos and C. Kyriakakis, “From Remote Media Immersion to Distributed Immersive Performance”, *in Proceedings of the ACM SIGMM 2003 Workshop on Experiential Telepresence,* November 7, 2003, Berkeley, California, USA

[3] X. Gu, M. Dick, Z. Kurtisi, U. Noyer, and L. Wolf, "Network-centric Music Performance: Practice and Experiments", *IEEE Communications Magazine*, June 2005, pp.86-93.

[4] A. Barbosa, “Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation”, *Leonardo Music Journal* - Volume 13, MIT Press, 2003, pp. 53-59.

[5] W. T. C. Kramer, “SCinet: Testbed for High-Performance Networked Applications,” *IEEE Comp. Mag.*, vol. 35, no. 6, June 2002, pp. 47–55.

[6] R. Rowe, “Real Time and Unreal Time: Expression in Distributed Performance”, *Journal of New Music Research*, vol. 34, Routledge, 2005, pp. 87-95.

[7] K. Sooyeon, L. JeongKeun, W. Y. Tae, K. Kyoungae, and C. Yanghee, “Hat: A High-quality Audio Conferencing Tool using mp3 Codec”, *INET 2002*, Washington, DC, USA, June 2002

[8] P. Bellini, P. Nesi, G. Zoia, "Symbolic Music Representation in MPEG", *IEEE MultiMedia*, Oct-Dec, 2005, pp. 42-49.

[9] J.-C. Bolot, H. Crepin and A. Vega Garcia: “Analysis of audio packet loss in the Internet”, *Proceedings of the NOSSDAV'95*, 1995, pp. 154 – 166.

# Programming with Pitch: MidiBraille, a Composing Tool for Blind Musicians

Lydia Machell
*Interdisciplinary Centre for Scientific Research in Music, University of Leeds – External Member*
*lydia.machell@primavistamusic.com*

## Abstract

*The more sophisticated music technology becomes, the more difficult it can be to provide accessibility to course content. While sighted students benefit from technology-enhanced approaches to music pedagogy, blind students can be further marginalised by them.*

*MidiBraille is the composing tool component of the Prima Vista Braille Music System. All elements of a Braille score – text and symbols as well as notes – can be created from a Midi keyboard by a blind user, producing Braille and print scores simultaneously. MidiBraille uses pitch not only as aural feedback but as a programming language in a unique combination of step-time and 6-key Braille input. While embracing the culture of Braille music literacy, MidiBraille's output in both Braille and standard notation promotes blind students' inclusion in a range of activities open to their sighted peers, including cooperative projects, digital score distribution and e-learning.*

## 1. Introduction

Advances in multi-media approaches to music pedagogy provide valuable new tools to sighted students, but blind students are further marginalised by "visual-centric" methods. Attempts to make standard teaching tools accessible to the blind, the so-called "enabling technologies", are too often retrofitted add-ons requiring the blind user to mimic the actions of the sighted user. At the same time, some multi-media approaches might be described as "disabling technologies", widening the accessibility gap more than ever before.

This article focuses on one aspect of the Prima Vista Braille Music System, MidiBraille. It traces the development of an approach that uses pitch as a programming language and also introduces the concept of Adoptive Design.

It is not within the scope of this article to make a comparative study of tools for accessible music education; this subject is covered by the i-Maestro publication, "Accessibility Aspects in Music Tuition" [1].

## 2. Adoptive vs. Adaptive Design

When sighted musicians had only their ears and the printed score to work with, Braille-literate blind musicians were less disadvantaged than they are now, despite the rigours of memorisation and the scarcity of Braille scores available. Braille music notation is a comprehensive system of music representation which has been sidelined by the revolution in score-writing software such as Sibelius. Attempts to bridge the accessibility gap have concentrated on verbal description. Sibelius Speaking [2], for instance, uses JAWS scripts to navigate Sibelius software so that a blind user can create a print score. These and other systems take stave notation as their starting point. Enabling a blind person to create a print score has its uses, just as providing a sighted person with the means to produce a Braille score might, but it is not intrinsically meaningful to the user. Such approaches are referred to as adaptive but in fact it is not so much the software that has been adapted to the user, as the user who must adapt to the software environment.

An adoptive approach is one which adopts the user's abilities and needs as the basis for the underlying design concept. In the context of music education, this would mean taking a blind user's Braille and aural skills as a starting point and capitalising on them.

## 3. The Prima Vista Braille Music System

The Prima Vista Braille Music System takes an adoptive approach by using Braille input and non-verbal aural feedback wherever possible. The system consists of software and hardware designs which address the three main aspects of access to music for the blind: Braille copies of existing digital scores are created by the Print-to-Braille Transcriber; the Braille Music Interface generates SimBraille as a teaching tool for sighted users; and the MidiBraille Interpreter enables blind users to simultaneously create Braille and print scores, with the benefit of aural feedback. The system is described in detail in "An Introduction to the Prima Vista Braille Music System" [3].

## 4. Braille music conventions

Like stave notation, Braille music consists of notes, symbols and text. Unlike stave notation, it is depicted linearly rather than as a time-against-pitch graph. Pitch is shown as a combination of note name and occasional octave indications. The sign for each note incorporates note name and duration, while symbols and text all have fixed positions relative to the note to which they apply.

Braille itself is made up of 6-dot cells and can be produced on a 6-key Braille typewriter. These can be mechanical devices producing embossed hard copy such as the Perkins Brailler or electronic devices such as portable Braille note-takers which can store Braille files. Each key relates to a numbered dot of a Braille cell. Used in combinations, they produce all possible variations of Braille cell. Apart from singers, who have both hands free, blind musicians must memorise scores for performance.

## 5. The origins of MidiBraille

The concept of MidiBraille was inspired by terminology. Since multiple keys are depressed simultaneously, 6-key input is not referred to as "typing" but as "chording". To the author, this prompted an analogy between pitch and Braille dots.

Most score-creation applications offer a number of note input methods including mouse (clicking on the stave to place a note), computer keyboard (typing note names), real-time input (inputting notes from a Midi keyboard, playing strictly to a metronome click) and step-time input. Step-time combines the input of pitch from the Midi keyboard with an indication of duration from the computer keyboard. This method is a popular compromise as it has the advantages of speedy pitch input, particularly for chords, without the strict time-keeping demands of real-time input.

Each of these methods, however, covers only note and duration input. Text and symbols are added to the score from on-screen palettes, tool-bars or menus, or from the computer keyboard, either in full or as keyboard short-cuts. For the blind user, even with the assistance of navigation software, the options are either impossible (such as mouse input) or extremely difficult, and can require frequent changes between Midi and computer keyboards.

## 6. The development of MidiBraille

Braille typewriters are configured as two sets of three keys in a single horizontal row. Keys 1, 2 and 3 correspond to the index finger, middle finger and ring finger of the left hand, while keys 4, 5 and 6 mirror this in the right hand. The key numbers in turn relate to the dot numbers of a Braille cell. These are arranged as two parallel vertical rows of dots, with dots 1 to 3 in the left-hand column and dots 4 to 6 in the right-hand column. Braillists can speedily create literary, mathematical or music code Braille using the six keys of a Braille input device.

The MidiBraille design began with the black keys of the Midi keyboard. Looking at the group of three keys representing F#, G# and A#, it was decided to use this set in two octaves to correspond to the six keys of a Braille typewriter, three in the left hand and three in the right.

### 6.1 The Data Score

Having theoretically assigned Braille dots, as well as other functions such as Return and Space Bar, to Midi pitches, how could this information be compiled and output as Braille? In particular, adhering to adoptive rather than adaptive design principles, how could this be done with minimal recourse to the computer keyboard?

It was decided to keep the MidiBraille process entirely contained in a score-writing environment as the use of standard studio software and hardware would avoid the compatability issues which often arise when multiple applications are involved. This approach also recognises the need for educational institutions to stay within budget while ensuring that their course material is accessible.

Information input at the Midi keyboard would be held in a "Data Score". Although represented as notes on a stave, the Data Score has no musical meaning. However, when the MidiBraille software is run on this score, the correlation between Midi pitch and Braille dot is analysed with reference to a look-up table and a Braille text file is created.

## 6.2 Words and music

Using the 6-key method, digital Braille files, whether literary, mathematical or musical, can be created from a Midi keyboard. This may have its uses but creating music code Braille with 6-key input is similar to using note-name input from a computer keyboard. Further development was needed to devise a Braille equivalent of step-time.

This was achieved by extending the concept of Midi pitch as programming language and assigning different functions to different octaves of the Midi keyboard. Midi pitches were assigned meanings for note duration, octave shifts and rests, while other symbols and text could still be added to the score with the black note 6-key input method (see figure 1). MidiBraille became not just the Braille equivalent of step-time input but an improvement on it as text, symbols and note duration can all be input from the Midi keyboard, with pitches providing aural feedback at each step.

## 7. Bidirectionality

Accessibility is too often treated as a one-way street. In the case of Braille music provision for the blind, the difficulties of providing Braille transcriptions of print scores can obscure the other half of the problem: disseminating the works of blind musicians and giving them the means to participate fully in the activities of colleagues and fellow students.

When a score is created using MidiBraille, it is automatically output as both a Braille file and a Sibelius file (figure 2). It can be aurally proofed in Sibelius as well as, of course, distributed to sighted musicians, while the Braille file can be embossed, read on a Braille display or treated as any other digital file, for instance saved to disk or emailed. In an educational context, there is potential for an unprecedented degree of inclusion for blind students, not by asking them to sit in front of computer monitors and to adapt to an essentially graphic system, but by adopting a system based on Braille and aural skills.

With sighted musicians gaining access to Braille scores through MidiBraille and blind musicians accessing digital print scores through Prima Vista's Print-to-Braille Transcriber, bidirectionality is achieved within one system.

This two-way street extends into every area where digital music scores are used, beyond the classroom and studio and into sheet music sales sites, platforms such as Yamaha's Digital Music Notebook [4] and e-learning programs.

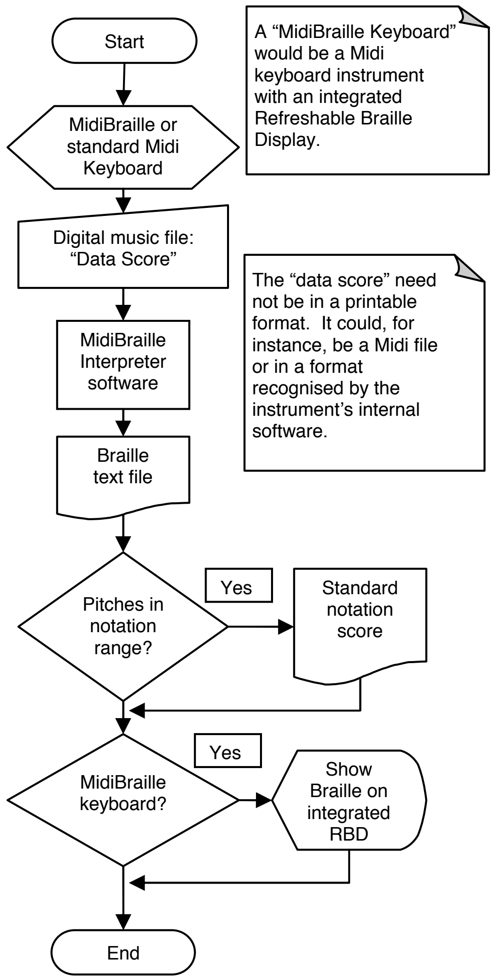



**Figure 1. MidiBraille Workflow**


**Figure 2. Data Score and Print and Braille output**

## 8. Future developments

Future developments would include the production of a MidiBraille keyboard prototype, and the development of a range of Prima Vista outreach programs, including the transcription of educational and test material and the design of e-learning courseware.

### 8.1 The MidiBraille keyboard

The project has benefitted from the loan of a PSR 1500 keyboard from the Yamaha Corporation. Areas for future research include the use of a foot pedal in MidiBraille input and the development of an interface to allow the creation of user-defined sets of sound events for different input functions. Although this keyboard, and many like it, has a menu-driven graphic display that would make navigation by a blind user difficult, it would be worth exploring its many functions to see what aspects of MidiBraille could be extended or pre-programmed.

At its current stage of development, the Prima Vista system uses standard software and hardware found in most music teaching environments. While this benefits institutions aiming to provide accessibility on a budget, there is a need for the additional development of a dedicated MidiBraille keyboard for professional users or institutions targeting visually impaired students. This would have an integrated Refreshable Braille Display (RBD) and would use the keyboard's own internal software to compile the MidiBraille data, dispensing with the need for an overt Data Score step. The RBD would provide convenient access to compositions recorded on the keyboard as well as to commercially-available music data such as the files available for download from the Digital Music Notebook website. Combining Braille input and display with the existing functions of high-end keyboards, the MidiBraille keyboard would essentially become a stand-alone Braille Music Workstation. This would entail further liaison with a number of keyboard manufacturers in order to find an appropriate partner for prototype design.

### 8.2 Prima Vista outreach

Outreach projects using the Prima Vista system could include the development of Braille music e-learning courseware aimed at both blind and sighted users; the creation of a Braille transcription service for music publishers and examination boards; and a continuous assessment of multi-media music teaching tools as they emerge.

## 9. Prima Vista in context

Although the project is the sole work of the author, development has not taken place in a vacuum. The system was first introduced in Zurich in 2004, when the Braille Music Subcommittee of the World Blind Union met to discuss developments in Braille music and further standardisation of the Braille music code. Valuable feedback from other Braille music experts was also gained at the International Symposium on Braille Music at the German Library for the Blind in Leipzig in 2005. Trinity College of Music, London, has used the Print-to-Braille transcriber component to provide scores for visually impaired students since the beginning of 2006 and the Homai National School for the Blind and Vision Impaired in New Zealand has been a test site since 2005.

## 10. The Gutenberg Fallacy

Braille music literacy is the subject of some debate. Its decline in recent years has been linked to the closure of specialist schools for the blind, and it has been argued that it is difficult to learn.

The difficulty isn't inherent in the system but in the fact that it is hard to motivate students if they are not rewarded by easy access to the music they want; by having the option to browse for music; and by having the means to compose and distribute their own music.

These obstacles are now at last surmountable. Asking whether the project is worth pursuing because the current market is so small is like questioning the potential of Gutenberg's moveable type. Fortunately, the future for the mass production of books wasn't determined by the literacy rates of the time. Supply can create demand.

## 11. Conclusion

Advances in music technology can either distance the blind music student or be used to breach the accessibility gap. The development of the MidiBraille composing tool was based on adoptive rather than adaptive design principles. Central to this approach was the use of pitch as a programming language, enabling the blind user to create scores, including text and symbols, entirely from a Midi keyboard with the benefit of aural feedback.

Potential areas of future development were discussed. These include comparing and maximising the functionality of different Midi keyboards, and the development of a MidiBraille keyboard prototype.

Outreach programs based on the Prima Vista system were suggested, involving multi-media designers, music educators, examination boards and music publishers, as well as the production of e-learning courseware.

It was argued that the current level of Braille music literacy is not a reliable indicator of the need for Braille music provision, but that – as is the case for literacy in general – supply will promote demand.

## 12. Acknowledgments

The Prima Vista system is beta-tested by James Hitchins, Disability Projects Officer, Trinity College of Music, London; Jane Ware, compiler and editor of *The Braille Music Training Manual* and *The Braille Music Layout Manual* [5]; and John Eshuis, Homai National School for the Blind and Vision Impaired, Auckland, New Zealand. Thanks to them all for their continuing feedback and support.

Thanks also to Thuy Mallalieu, Royal National Institute of the Blind, for his input at the early stages of the project, and to Toshihisa Sadakata and Gary Gregson, Yamaha Corporation R&D, for the loan of the PSR1500 keyboard.

## 13. References

[1] *DE4.5.1– Accessibility Aspects in Music Tuition* Available at the i-Maestro project website: http://www.i-maestro.org

[2] Sibelius Speaking http://www.technovision.co.uk/Sibelius.htm

[3] Machell, L., *An Introduction to the Prima Vista Braille Music System*. Available from info@primavistamusic.com

[4] Digital Music Notebook from Yamaha http://www.digitalmusicnotebook.com/home/

[5] Ware, J., *Braille Music Training Manual,* Royal National Institute of the Blind, Peterborough, UK, 1997, revised 2001.
Ware, J., *Braille Music Layout Manual,* Royal National Institute of the Blind, Peterborough, UK, 1998, revised 2002.

# Improving Access to Music Education for the Blind and Visually Impaired

David Crombie, Benjie Marwick Johnstone, Neil McKenzie
*Dedicon*
*dcrombie@dedicon.nl, bjohnstone@dedicon.nl, nmckenzie@dedicon.nl*

## Abstract

*The i-Maestro project aims to develop new technologies to enhance the quality and accessibility of music education available across Europe. Dedicon are developing accessibility features for i-Maestro to allow students to interact with lessons and music notation using Braille and spoken music formats.*

*This paper will explore the problems faced by blind students when learning music, the motivations and aims of the i-Maestro project, and the development of accessible music notation editors and other accessibility features within i-Maestro.*

## 1. Introduction

Learning music has always provided many challenges for blind and visually impaired students over and above those met by their sighted counterparts. One of the biggest is simply access to music notation, which is traditionally recorded in visual form using the well-known five-line staff system. While various accessible music notation formats exist, transcription into these formats is time-consuming and expensive. Libraries of accessible music scores are limited in volume, and the availability of accessible music teaching materials is even more so. Transcription on request can take several months, and few teachers know what material they will need several months in advance. Thus it has always been very difficult for blind and visually impaired students to follow the same curriculum as sighted pupils.

## 2. About i-Maestro

I-Maestro is a European research project into the use of technology to enhance music education. [1] It aims to both provide technological support for traditional methods of teaching and to develop and enable new music teaching methods such as co-operative / competitive work and distance learning.

I-Maestro aims to provide exciting new tools while automating some of the more boring aspects of traditional music education. Posture and gesture analysis allow a pupil to watch a 3D representation of their bowing technique while playing a violin or to perfect their conducting skills by conducting a “virtual orchestra” that follows their instructions, mistakes and all. Meanwhile, if a pupil is having trouble with a specific passage of music, the computer can help by automatically creating various exercises that focus on different aspects of that passage, for example by focusing on the notes and the rhythm separately.

Accessibility is a central aim of the i-Maestro project. Traditionally it has been very difficult for sighted and visually impaired students to follow the same curriculum due to the difficulty of converting suitable materials into an accessible format. However, where lessons are provided digitally, accessible lesson material can be created on-demand and customised very closely to the user's personal requirements. [2]

## 3. Accessibility in i-Maestro

Work has been carried out on a state of the art analysis of the user requirements, file formats, standards and technologies for accessible music tuition such that the field of accessible music learning can be established to an extent that analysis of the accessibility issues and their impact on music education for regular learners and those with special needs can be made for the i-Maestro project. The assistive technologies relating to accessible music and accessibility in general (Sonification, screen readers, gesture and posture analysis, alternative representation and devices, zooming, spoken music, etc.) have been overviewed in an extensive document such that the technology is available for incorporation into technologies developed in the other work packages of i-Maestro. This report is available from the i-Maestro web site. [3]

Two major obstacles prevent easy use of the i-Maestro software by visually impaired users. The first is the decision to write all software in Max/MSP, a graphical programming language from Cycling 74. [4] Max/MSP is excellent for audio and video processing, but was not designed for creating accessible user interfaces and does not work well with JAWS or other screen readers. To work around this problem, scripts will be written for the i-Maestro software to co-ordinate control, navigation and output with JAWS.

The other main accessibility challenge is that many of the pieces of software produced for i-Maestro use music notation to some extent. A graphical music viewer / editor has been developed for this purpose by the University of Florence. It is Dedicon's role to augment this editor to allow the use of alternative, accessible music notation formats.

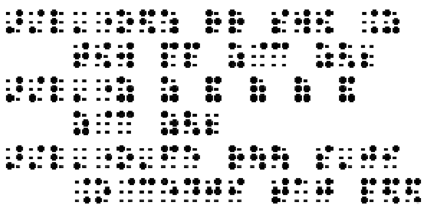
## 4. Accessible Music Notation Formats

Three alternative music notation formats are available for visually impaired users who are unable to use conventional western music notation.

### 4.1. Large Print Music

By far the most common format of music notation for visually impaired people is Large Print music. Here, conventional music notation is magnified, usually by around 200 – 400%. It is common to scale small symbols such as dots by a greater amount proportional to larger symbols such as note heads and clef signs. It is also common to move symbols relatively closer together to save space on screen or paper. [5]

### 4.2. Braille Music



Louis Braille himself was a musician, and invented a method of writing music notation using the 64 standard Braille characters. He assigned new meanings to various combinations of these characters, and the system is still in use today.

Over the years various countries have developed their own additions and conventions for Braille Music. In the early 1990s, attempts to agree an international Braille Music standard resulted in a definitive manual. [6] However, this has still not been fully adopted, with many libraries containing older music and many users preferring their national conventions.

### 4.3. Talking Music

```
Key Signature: One sharp.
Time Signature: 4 fourths time.
  Section 1.  Bar 1.
    third octave d whole.
  in agreement with
    fourth octave d half dotted.
    e.
    d.
...
```

Talking Music is a format developed by Dedicon, where a user hears a short excerpt of music played, followed by every detail of that section from the printed page being read out. [7] It is commonly used with the Daisy Talking Book [8] format, which allows easy navigation around the score.

The musical examples are produced automatically by software and as far as possible with a non-interpreted version of the music. All notes are played at the same volume, at the same velocity, and all notes of a certain duration have exactly the same duration. The musical example is there to provide the user with a context for the spoken material, making it easier to understand and play. The musical example is non-interpreted in order to afford the Talking Music user the same level of subjectivity to the musical content as the sighted user.

## 5. Methodology

### 5.1. The Music Notation Editor

The i-Maestro project has produced several pieces of software, many of which require the facility to edit musical notation. This involves varying degrees of complexity, from a single student completing a simple theory exercise to a group editing multiple parts on a score collaboratively. A music editor module has been developed for this purpose by the University of Florence's Dipartimento di Sistemi e Informatica (DSI.) [9] This editor is based on the new MPEG-SMR (Symbolic Music Representation) format. [10]

Dedicon are responsible for augmenting this notation editor to be accessible to users of various alternative musical notation formats, such as Braille Music and Talking Music. Large print output has already been implemented by DSI, with user-selectable magnification ratios.

### 5.2. The AccessMusic Finale Plug-in

The accessible music editors are based on code produced by Dedicon as part of the AccessMusic project. [11, 12]

The AccessMusic project provided a set of tools for creating accessible music which are freely available to download. These tools allow you to convert music scores from traditional western music notation to formats for the Blind and visually impaired. Currently the project provides software plug-ins which convert from Finale Music to Braille Music and Talking Music formats.

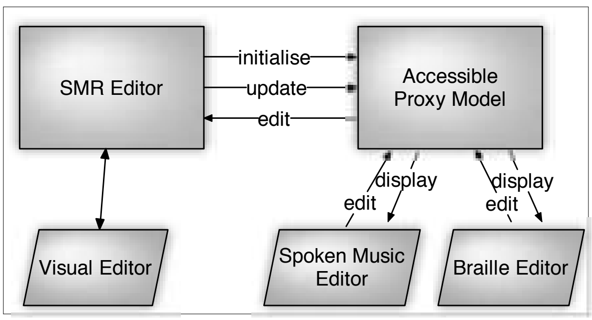
### 5.3. Linking the two together

The SMR editor and the AccessMusic code are both written in C++. The topmost layer of the SMR editor code is a simple wrapper to turn this code into a Max/MSP external known as the "IMED" (interactive music editor) module. We propose to modify this wrapper layer, interfacing with the SMR editor at the most abstract level possible within C++.

The accessible editors are turned on and off via Max messages. These messages can then be sent by the software in which the IMED is used based on the user's profile. As future work, output preferences could be initialised by this method as well, and saved back to the user's profile at the end of a session. [13]

## 6. Work

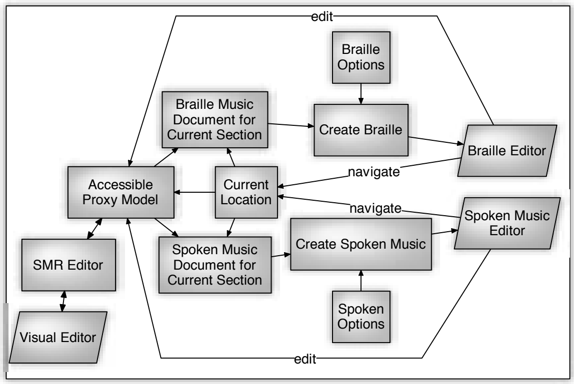
### 6.1. Architecture




The above diagram displays the SMR editor in the top left. The SMR editor is a music editor which was produced as MPEG reference software as part of the standardisation procedure within MPEG. For the purposes of i-Maestro it has been provided with a Max/MSP wrapper called the "IMED plug-in" which allow various modules running under the Max/MSP runtime environment to access functionality and information within the SMR editor.

In order to provide accessible music format extensions to this work, a similar architecture to that employed in the AccessMusic project has been used. [14] A much simpler model of the music is created, called the "accessible music proxy model." From this "proxy model," accessible representations such as Braille Music and Talking Music can be created. One advantage of this architecture is that with an appropriate wrapper to load the "proxy model" from another source, it becomes easy to reuse the accessible editing interfaces with other music editors.

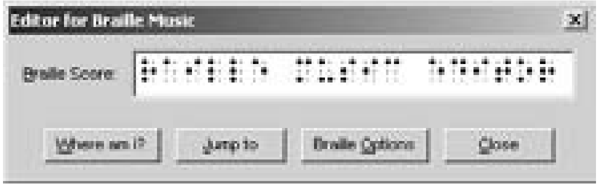
The first aim of the Accessible Music extensions is to display i-Maestro lessons in the form of Braille Music and Talking Music. Once this is successful, it is hoped that two way feedback between the various modules can take place, allowing Braille and Talking Music representations to be edited and the changes fed back into i-Maestro music lessons.




As in the AccessMusic code, the generic proxy model is parsed to produce a specific proxy model for the format required, i.e. Braille or Talking Music. However, the accessible editors keep track of the section currently being displayed, and only this section is copied into the specific proxy model, in order to reduce processing overheads. This is then converted into ASCII text for output, as in the AccessMusic code, and according to the user's processing options. Upon navigation within the accessible editing interface, this process is repeated with the section around the new location.

Changes made to the Braille representation are used to directly update the main proxy model. A pointer into the proxy model is maintained at all times, showing the location corresponding to the current location in the Braille view. When characters are deleted, the corresponding entity in the proxy model is removed, and when characters are inserted, the appropriate entity is inserted as soon as a valid sequence is recognised. The process above is then repeated to update the display.

### 6.2. Editor for Braille Music




The user interface for the editor for Braille Music is essentially a dialog box with a single text box displaying ASCII characters. These are displayed as Braille by the user's Braille Display Bar and can be edited in the same way as a conventional text box. They can also be displayed in Braille on-screen for use by sighted users if a Braille font is installed on the User's system.

Braille characters are entered by the user using six keys on a conventional keyboard. Sequences of characters are identified and the correct musical entity inserted into the proxy model. This approach was used successfully in the Braille Music Editor by Dodiesis. [15]

The rules for mapping character sequences to musical entities are stored in external XML / XSLT files, to allow for future modification. The standard international Braille Music conventions have been used during development.

Navigation around the score is performed by using the arrow keys. A navigation dialog offers larger jumps, such as previous/next bar, section, part etc.

"Display" options such as the fragment length can be configured via a further dialog.

### 6.3. Viewer for Talking Music


The user interface for the Talking Music viewer is essentially a dialog box with a single text box displaying ASCII text. This is read aloud by the user's Screen Reader. This can also be viewed directly by sighted users.

Navigation around the score is performed by using the arrow keys. A navigation dialog is provided for larger jumps, such as previous/next bar, section, part etc.

"Display" options such as language and default note length can be configured via a further dialog.

## 7. Progress

The first prototype system is scheduled for completion by late October 2007, when we will be looking for some initial feedback from potential end users and anyone else interested in the project.

The software will be developed to a beta stage and released open-source on SourceForge by 14th December 2007. This will allow anyone interested to make use of the software or develop it further.

As the use of accessible music notation formats within computer-based music education is a niche market at present, we will be looking to conduct user testing and also gain more informal feedback from a wide variety of people, many of whom may not at present be involved in one or more of these areas.

## 8. Further Work

Whilst the work described in this paper will allow those familiar with Braille Music to learn more about music theory, the biggest challenge is perhaps to help the user to learn Braille Music first. This should be relatively simple to implement in i-Maestro "lessons" by showing Braille Music alongside a more "intuitive" format like Talking Music.

The ability to edit a score "displayed" as Talking Music would be beneficial to those who do not read Braille Music. The exact details of how insertion, deletion and editing would be performed present an interesting research challenge.

It has been mentioned that while the international standard for Braille Music has been used throughout this project, national variations are still widely used and highly popular. Due to the modular architecture of the accessible editors, it would be possible to cater for these national Braille Music conventions simply by altering the XML / XSLT files from which the Braille Music rules are loaded. [16]

## 9. Conclusions

The i-Maestro project aims to develop new technologies to enhance the quality and accessibility of music education available across Europe. This project will increase the opportunities for blind and visually impaired students to follow the same curriculum as their sighted counterparts.

Dedicon are developing accessibility features for i-Maestro to allow students to interact with lessons and music notation using Braille and Talking Music formats. The accessible music editor software will be released open-source.

## 10. Acknowledgements

The i-Maestro project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST 026883, www.i-Maestro.net, www.i-Maestro.org). Thanks to all the i-Maestro project partners and participants for their interest, contributions and collaboration.

## 11. References

[1] i-Maestro Project Website
http://www.i-maestro.org

[2] Crombie, D., Lenoir, R., and McKenzie, N., (2007) Designing accessible software for music applications, in Hersh, M., & Johnson, M. (eds) Assistive Technology for vision impaired and blind people, Ch. 16, Springer-Verlag, Berlin Heidelberg New York

[3] i-Maestro Deliverable 4.5.1 "Accessibility aspects in Music Tuition" available from:
http://www.i-maestro.org/documenti/view_documenti.php?doc_id=629

[4] Cycling 74 Website
http://www.cycling74.com

[5] Music in Large Print, RNIB
http://www.rnib.org.uk/xpedio/groups/public/documents/visugate/public_muslprnt.hcsp

[6] Krolick, B., (1996) "New International Manual of Braille Music Notation", World Blind Union, Dedicon, Amsterdam, Netherlands

[7] Talking Music Website
http://projects.dedicon.nl/Talking%20Music/default.htm

[8] DAISY Consortium Website
http://www.daisy.org/

[9] DSI Website
http://www.dsi.unifi.it/

[10] P. Bellini, F. Frosini, G. Liguori, N. Mitolo, P. Nesi, "MPEG -Symbolic Music Representation Editor and Viewer for Max/MSP" Presented at the i-Maestro Second Workshop, Axmedis 2006, Leeds

[11] AccessMusic project website
http://accessmusic.sourceforge.net

[12] Crombie, D., Lenoir R., McKenzie N., (2005) Accessible Music Technology for People with a Print Impairment, Journal of New Music Research, Vol.34, No.2, Routledge

[13] Stephanidis, C., 2001. "User Interfaces for All: New perspectives into Human-Computer Interaction" in: Stephanidis, C. (Ed), User Interfaces for All – Concepts, Methods, and Tools (pp 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

[14] Crombie, D., Marwick Johnstone, B., McKenzie, N., "Incorporating accessibility within pedagogical environments"(2007) 7th CVHI Conference, Granada, Spain

[15] Dodiesis website
http://www.dodiesis.com

[16] Dardailler, D., & Palmer, S. B., & McCathieNevile, C. (Eds.) (2001). XML Accessibility Guidelines (XMLGL). W3C Working Draft, 03 October 2002 [On-line]. Available: http://www.w3.org/TR/2002/WD-xag-20021003.

# Automatic Generation of Test Objects for Music Self-Study

Kerstin Neubarth, Tillman Weyde[1]
Pierfrancesco Bellini, Francesco Frosini, Nicola Mitolo[2]
[1]*City University London, School of Informatics, Department of Computing, London, UK*
[2]*DISIT-DSI – University of Florence, Florence, Italy*
*{Kerstin.Neubarth.1, t.e.weyde}@city.ac.uk, {pbellini, ffrosini, mitolo}@dsi.unifi.it*

## Abstract

*Recent research and development of adaptive assessment addresses the automatic construction of test items. Test items are created on test delivery, by randomly setting variables in test item code or by selecting from a superset of test elements. The code or superset must be provided by the test author. In this paper we present an alternative approach, which requires the author only to select a test type and set some parameters for test generation. This reduces the author's workload, and enhances adaptation and personalisation, because variables in the generation are not assigned random values but pedagogically motivated values.*

## 1. Introduction

The i-Maestro project [9], which is partially funded by the European Commission (FP6 IST), is developing an interactive multimedia environment for technology-enhanced music education, covering music theory and practice learning with a focus on string instruments. The i-Maestro environment includes a production module with authoring and generation tools to create music exercises, lessons and courses. In this paper we focus on the automatic generation of exercises that can be automatically assessed. This group of exercises we call tests.

Tests form but a part of music exercises. While many music-learning activities involve skills rather than knowledge and are explorative and creative in nature, thus making assessment difficult or even detrimental to creativity [16], there are some areas of music education which are well defined in terms of expected student performance, like in basic music-theory training. For these areas, educational technology can support students' self-learning at home or in a music lab by offering exercises with automatic assessment. Providing students with knowledge of results (KR) and immediate informational feedback has been shown to increase their learning motivation and help improving learning achievements (e.g. [10], [12], [13], [15]).

Technology supported creation and evaluation of tests has been addressed by authoring tools for adaptive assessment systems (e.g. [7], [8]). To adapt to a student such systems select the next test question according to the previous student response. Underlying this selection is an update of the student's estimated knowledge level.

Guzmán and Conejo 2004 [7] propose a library of templates to create tests automatically. Templates cover true/false, multiple choice, multiple response, self corrected, fill-in-the blank, ordered response, inset, matching, word search and puzzle tasks. Two approaches to automatic generation are used: generative test items contain code, like numerical expressions or JSP, with variables that are set randomly when the question is created. The student only sees the output of the code. Alternatively, the author defines more elements for the test than will be shown to the student; automatic generation of the test item then consists of selecting a subset of the provided elements.

An ontology-based approach to the automatic creation of multiple choice questions is taken by Fischer 2000 [5] and Fischer and Steinmetz 2000 [6]. In the subject area of multimedia systems questions of two kinds are used: part-of (e.g. "what are the parts of an adaptive hypermedia system?") and application-of questions (e.g. "what are the application areas for Intelligent Tutoring Systems?"). The system selects one true and several false text options from a terminological ontology of multimedia systems.

The described generation techniques focus on knowledge tests and require additional authoring workload in terms of coding or overloading test items [7] or the initial definition of a domain ontology ([5], [6]). The ontology design may be partly automated as in semi-automatic authoring of hypermedia learning systems (e.g. [4]). The adaptation in adaptive assessment systems lies in the sequencing of test items rather than their generation. The system by Fischer 2000 [5] and Fischer and Steinmetz 2000 [6] does not yet realize adaptation.

The i-Maestro music exercise generation tool does not only support knowledge-oriented, but also activity-oriented exercises, although the tests we describe in this paper are mainly directed at music-theoretical knowledge and basic perceptual and music editing skills. The author's task is reduced to set options for the generation, by which he can control the generation process and outcome. Generation algorithms, as compared to the coding in generative test items described above, are entirely provided by the tool; and while the variables in generative test items are given random values, the algorithm parameters, based on the user-set options, are given pedagogically motivated values. The parameterisation of algorithms allows adaptation and personalisation in the generation of test questions, not only in their sequencing. Sequencing is not further addressed in this paper.

In the terminology of Guzmán and Conejo 2004 [7], "test" refers to a sequence of test items or questions. In contrast, in the remainder of this paper we use "test" for individual questions, to avoid cumbersome expressions (like "test item types") and because here we deal only with the generation and evaluation of single questions.

## 2. Test Types

For tests with automatic assessment, we currently consider seven core test types in i-Maestro. True/false tests require students to decide whether a presented statement is correct or incorrect. The statement can be given as text, music score, graphics, audio or video, or a combination of media types. For example, in a music score a chord is labelled by a chord symbol which describes the chord either correctly or incorrectly.

In multiple choice tests students are shown a target and several options; from these options they have to select the one that matches the target. In multiple response tasks more than one of the options match the target and should be selected. Target and options can be of several types, like text, music score, MIDI, audio or video. For example, the student listens to a pitch interval played by audio (target) and determines the interval class by selecting from intervals given in musical notation (options).

In fill-in-the-blank exercises the student completes missing parts of a presented text, music score, diagram, MIDI playback or audio recording. In music training such exercises can be designed as reconstruction or composition tasks. For reconstruction the completion has to match the original parts; reconstruction tasks can be automatically assessed. For composition the student creates new musical material (score, MIDI or audio) according to the presented context; by default, we assume that composition exercises will be assessed by a teacher.

In ordered response tests the student arranges presented elements, which can be text, music score, graphical, MIDI or audio items. Ordered response tests can be of two forms: With-target tests provide the student with the solution in a media type different from the type of the elements. For example, the student listens to an audio recording of a piece of music and realigns blocks of music score to match the perceived music. Without-target tests present only the elements themselves. For example, the student orders four chords to produce a cadence.

Matching tests ask the student to link items of two sets so that each element of the first set is associated with one element of the second set. Again, items can be of several media types, where all elements within a set have the same type. For example, the first set consists of chords in musical notation, the second set lists chord symbols. Or: the first set gives the chords in musical notation, the second set contains the corresponding MIDI or audio renderings.

Common tasks in music ear training are dictation and imitation. In both tasks the students listen to a music excerpt. In dictation they write the excerpt down in music notation. In imitation they play it on an instrument, sing it or, in rhythm tests, respond by tapping or clapping. As dictation and imitation differ only in the media type of the student response, but not in their general structure, we count them as one test type.

The described test types can be reused for several subjects in music learning, e.g. notes names, pitch intervals, harmony, rhythm, form or counterpoint (see e.g. [1], [11]).

## 3. Generation Algorithms

The automatic generation of tests in i-Maestro currently focuses on the creation of score, MIDI and audio items as elements in true/false, multiple choice, multiple response, ordered response and matching tasks. Reconstruction tests are at the moment limited to music scores to facilitate the evaluation of the student's response. We plan to extend automatic generation and evaluation of reconstruction exercises to audio, using audio processing and score following technology [14]. Similarly, we now include dictation (score response), but intend to add imitation tests (audio response).

Our generation of score, MIDI and audio items is based on MPEG Symbolic Music Representation (MPEG-SMR, [3]). MPEG-SMR data can be rendered as music score, MIDI or audio, providing presentations in different media types from one single representation format. In addition MPEG-SMR allows for score annotations of any type, including e.g. graphics or video. The SMR test elements are embedded in a test object. The test object determines the structure and interactivity of a test and is represented in Training Specification Language (TSL, [2]).

Figure 1 shows algorithms used in test generation. TSL algorithms process the test object as a whole. SMR synthesis algorithms create music data, while SMR variation algorithms modify music data. SMR analysis algorithms support the application of SMR variation according to the given musical context.

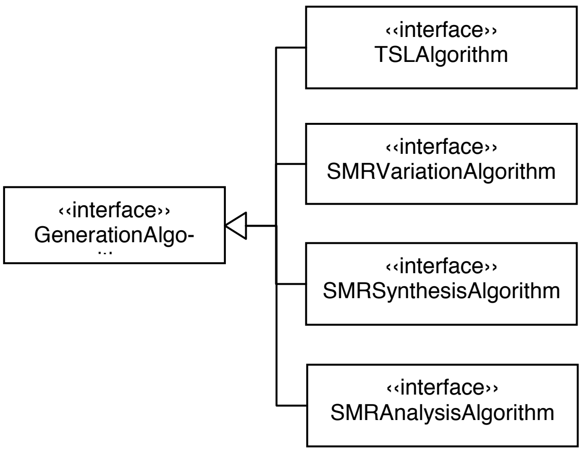



Figure 1. Generation algorithms

Figure 2 summarises the TSL algorithms used in generating objects of the test types described above. Apart from the InsertBlank and ShuffleSegments algorithms, the algorithms create one (CreateStatement) or more items of the specified resource types. Currently we create music objects which are presented as score, MIDI or audio. The difference parameter controls in which way a false statement is not correct (true/false tests), or how the false options differ from the target (multiple choice and multiple response tests). For matching tests the difference refers to the relation between elements within the two sets of items. This allows a flexible and pedagogically controlled generation of tests. When the difference is derived from information on individual learners, personalised tests can be created. Similarly, the structure parameter in the InsertBlank algorithm provides semantic flexibility because musical structures of different kinds (e.g. motives, accompanying voices or time signatures) can be cut out of the original piece of music, depending on the test objective. In the ShuffleSegments algorithm for creating ordered response tests, the musical structures are limited to horizontal structures (e.g. notes, measures, sections). The segments which are reordered consist of as many elements of the specified structure type as indicated by the length parameter. The segments are arranged in random order. Advanced versions of the ShuffleSegments algorithm could aim at an intelligent reordering of segments, taking into account the transitions between segments; but this is currently outside our design scope.

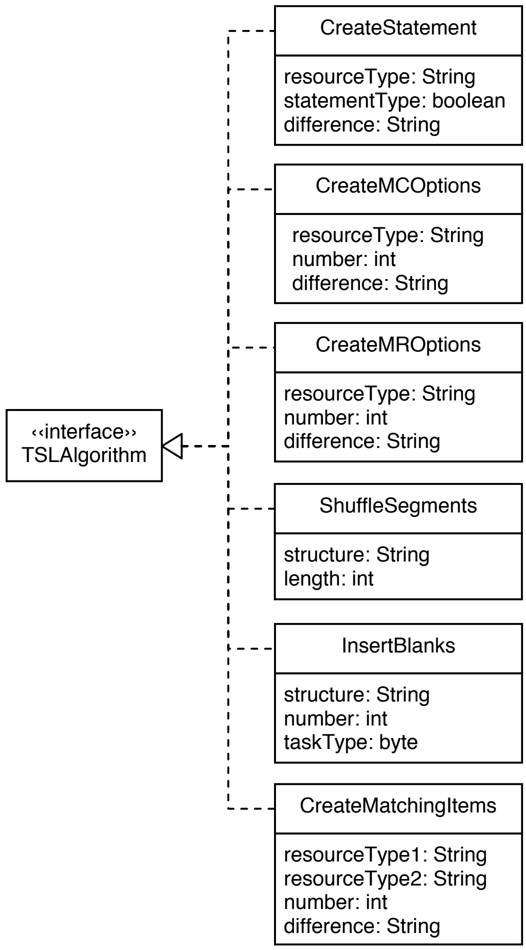



Figure 2. Algorithms for test object creation

Dictation tests do not require TSL processing in the same way as the other test types do. Generation of dictation exercises mainly consists of creating the music material which is presented to the student as MIDI or audio and which he has to write down as a score. This music material can be either taken from an existing SMR file or created from scratch by music synthesis algorithms.

Internally the algorithms in Figure 2 use SMR algorithms when creating items of the resource type MPEG-SMR, which are subsequently rendered as score, MIDI or audio. In particular, they apply music variation algorithms according to the difference parameters. The ShuffleSegments and InsertBlanks algorithms determine the structural elements to be reordered or cut out by music analysis.

To illustrate the use of music variation in creating test objects, we here give an example. A multiple choice test for ear training is generated, using the CreateMCOptions algorithm. Students will listen to one chord (audio target) and select from a list of chords given in musical notation (options as music scores). The same MPEG-SMR elements can be used to present students with a score target and audio options, where they listen to several options and have to select the option that matches the written chord.

For these multiple choice tests, the options could differ in the chord's mode (major, minor, diminished or augmented for triads) or inversion (root position, first inversion and second inversion for triads). In principle, chords can also differ in their fundamental (e.g. C major triad vs. D major triad), but this is rarely tested in ear training (unless a reference pitch is given). Thus, the difference parameter in the CreateMCOptions algorithm will be "mode" or "inversion", set by the user of the generation tool or by the tool processor based on e.g. a student's learning history. For a difference in mode, the options are derived from the target by applying the music variation algorithm TriadModeChange, which takes the desired mode as a parameter; for the difference in position the options are created by the TriadInversion algorithm, which takes the desired position as a parameter. The music variation algorithm is applied as many times as required by the number parameter in the CreateMCOptions algorithm. Figure 3 shows the target chord (top), user settings for the CreateMCOptions algorithm (centre) and resulting additional options (bottom).

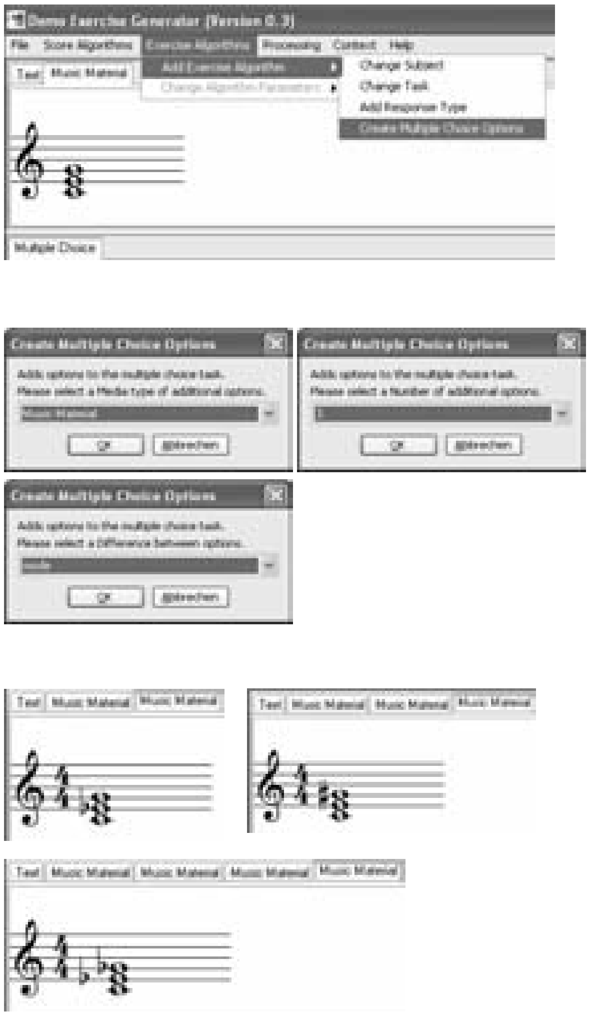



Figure 3: Creation of multiple choice options (screenshots)

Figure 4 illustrates the generation process for this example. The user sets the parameters of the CreateMCOptions algorithm. The algorithm then delegates to a TriadModeChange algorithm, provides it with the target triad, and lets it modify the major target triad to get the minor, augmented and diminished versions of the triad.

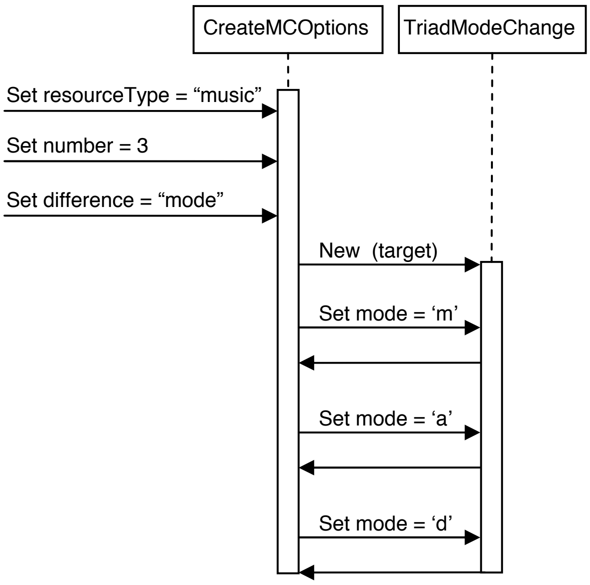



Figure 4: Creation of multiple choice options (processing)

The same music variation algorithms could be used to create a false statement in a true/false test (chord differing from the text annotation in its mode or position), items in a matching test (chords differing in their mode or position, shown as score in one set and rendered as audio in the other set) or incorrect options in a multiple response test. The additional correct options in a multiple response test are derived by a music variation algorithm that does not correspond to the difference parameter (e.g. transposition to change the chord's fundamental while maintaining its mode or position).

## 4. Evaluation Algorithms

The evaluation of students' performance on tests depends on the test type as well as on the media types used within the test. In addition, evaluation algorithms can provide different kinds of feedback to students: Controlling feedback for the considered test types can consist of a correct/incorrect statement (all test types) or an error count (ordered response, matching, reconstruction and dictation tests). In multiple response tests the error count includes false hits (not selected correct options) and false fails (selected incorrect options). However, different error counts will result, depending on whether errors are identified independently or dependently of each other. For example, one incorrect match in a matching test will automatically lead to a second incorrect match. In reconstruction and dictation tasks, elements following an incorrect element might be incorrect if compared to the target, but correct in relation to the preceding elements. This issue applies for automatic assessment as well as for human assessment. Pedagogically, instead of the number of errors, students could be told the percentage of correct performance. Informational feedback can easily be given by showing the correct answer on request. More advanced informational feedback would provide additional feedback on the kind of errors made.

Algorithms for evaluating i-Maestro tests operate on two levels. When the correct answer is stored as an object in the TSL representation of the test, the student's response should be identical to this object. In multiple choice and multiple response tests, the correct option object(s) can be marked in TSL. The student's response is correct if he has selected the option object(s) marked as correct in TSL. In matching tests, the student has to select the correct pairs of option objects. For ordered response tests the student response forms a list, whose element objects should be in the same order as in the list stored as solution in TSL. Figure 5 summarises these methods for evaluating the student response against data stored in TSL.

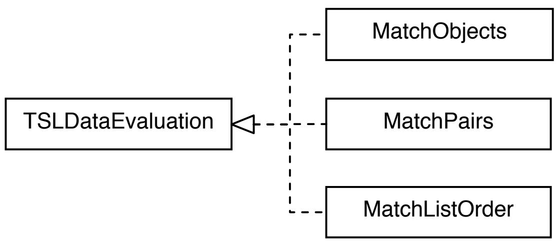



Figure 5: TSL-based evaluation algorithms

For MPEG-SMR targets and responses the evaluation is performed by SMR matching algorithms. While this evaluation approach could also be used for true/false, multiple choice, multiple response, ordered and matching tests based on SMR statements, targets and options, it is mainly applied for reconstruction and dictation tests. In these two test types the student does not select an object or object constellation already contained in the TSL representation, but creates a new object. The evaluation algorithm compares this new object with the target object. Depending on the specific task, only selected aspects of the SMR data are required to match.

The SMR library being developed in i-Maestro allows reading and writing music scores in terms of music symbols. This means that by using functions from the library, anyone can realize algorithms for the evaluation and comparison of music symbols. Some evaluation and comparison algorithms have already been included in the SMR library, allowing the following types of control:

- Interval detection: calculate the difference between two notes in terms of number of semitones.
- Duration detection: compare the ratio between two notes in terms of fraction.
- Consistency checking: check if the symbols contained in a measure are consistent with the measure time signature or vice versa.
- Retrograde comparison: compare two voices to check if the second is a retrograde version of the first.
- Inversion comparison: compare two voices to check if the second is an inversion of the first.
- Diminution comparison: compare two voices to check if the second is a diminution of the first.
- Augmentation comparison: compare two voices to check if the second is an augmentation of the first.
- Text annotation matching: compare two texts (for example the numbers of a figured bass) to check if they are compliant with each other.
- Harmonic relationships: check if a score observes the harmonic rules. In any case these rules will apply to one or more of the algorithms listed before.
- Counterpoint relationships: check if a score observes the counterpoint rules. In any case these rules will apply to one or more of the algorithms listed before.

## 5. Conclusion

In this paper we introduce an approach to automatic generation and evaluation of tests objects for music education, which we are developing in the European i-Maestro project. Such tests can support students in acquiring basic music theory knowledge, listening and music editing skills and in monitoring their learning progress. An automatic evaluation of the student's response provides immediate feedback, which makes these tests particularly suitable for music self-study. We are currently considering seven test types, which can be reused for different subject areas of music training: true/false, multiple choice, multiple response, fill-in-the-blank, ordered response, matching, and dictation and imitation tasks. The generated test object will define the test form, interactivity and assessment model, represented in Training Specification Language (TSL), and will most often contain music content, represented in MPEG Symbolic Music Representation (MPEG-SMR).

This paper focuses on the algorithms for test generation and evaluation: TSL algorithms for the generation of the test structure and a not music-specific evaluation of the student's response against information stored in the TSL object; music synthesis algorithms for creating music content from scratch; music variation algorithms for modifying music material, created by music synthesis or taken from an existing music score; and music analysis algorithms for a context-sensitive application of music variation and for a musical evaluation of the student's response.

The use of exercise and in particular music algorithms and thus domain-specific content processing is considered to reduce the author's workload and to increase the flexibility in test generation as compared to existing approaches to automatic test creation. For i-Maestro test generation, the author has to merely set a few generation options (like test type and subject parameters), instead of programming elements in generative test items or creating additional alternative test elements as in previous approaches. By setting options according to the individual student's learning needs or preferences, personalised tests can be created, allowing a finer-grained adaptivity than does selecting from already existing educational resources. Automatic evaluation of the student's response using music analysis algorithms enables detailed informational feedback, when it points to the position and kind of errors made. This information will subsequently be used to create new exercises and tests focusing on the emerging learning needs and to support adaptive sequencing of test objects.

## 6. Acknowledgement

The i-Maestro project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i-maestro.org). We thank all i-Maestro project partners and participants for their interest, contributions and collaboration.

## 7. References

[1] Auralia. http://sibelius.com/products/auralia/index.html

[2] A. Badii, "Mediating Representations: Domain Knowledge to Pedagogical Content Knowledge (PCK): supporting the instantiation of dynamic pedagogic flow responsive to the evolving situated chemistry of Pedagogic Experiences", *Second I-MAESTRO Workshop on Technology Enhanced Music Education*, collocated with the AXMEDIS06 International Conference, Leeds, 13-15 December 2006.

[3] P. Bellini, P. Nesi, M. Campanai, and G. Zoia, ISO/IEC FCD 14496-23:200x, Symbolic Music Representation, MPEG2006/N8632.

[4] A. Cristea, “Is semi-automatic authoring of adaptive educational hypermedia possible?”, *Advanced Technology for Learning* 1(4), ACTA Press, Calgary, 2004, pp. 227-236.

[5] S. Fischer, “Course and exercise sequencing using metadata in adaptive hypermedia learning systems, *Journal of Educational Resources in Computing* 1(1), ACM, Electronic Publication, 2000.

[6] S. Fischer, and R. Steinmetz, “Automatic creation of exercises in adaptive hypermedia learning systems”, *Proceedings of the 11th ACM Conference on Hypertext and Hypermedia, San Antonio, TX, May 30 – June 4*, SCM Press, New York, 2000, pp. 49-55.

[7] E. Guzmán, and R. Conejo, “A library of templates for exercise construction in an adaptive assessment system, *Technology, Instruction, Cognition and Learning* 1, Old City Publishing, Philadelphia, 2004, pp. 303-321.

[8] E. Guzmán, R. Conejo, and E. García-Hervás, “An authoring environment for adaptive testing, *Journal of Educational Technology & Society* 8(3), International Forum of Educational Technology & Society, Online Version, 2005, pp. 66-76.

[9] i-Maestro project, http://www.i-maestro.org.

[10] R. MacDonald, C. Byrne, and L. Carlton, “Creativity and flow in musical composition: an empirical investigation”, *Psychology of Music* 34, SAGE, 2006, pp. 292-306.

[11] Music Ace. http://www.harmonicvision.com

[12] R. Pekrun, and U. Schiefele, “Emotions- und motivationspsychologische Bedingungen der Lernleistung”, in F.E. Weinert (ed.), Psychologie des Lernens und der Instruktion, Verlag für Psychologie, Hogrefe, 1995, pp. 153-179.

[13] T. Priest, “Self-evaluation, creativity, and musical achievement”, *Psychology of Music* 34, SAGE, 2006, pp. 47-61.

[14] D. Schwarz, A. Cont, and N. Orio, “Score Following at IRCAM”, *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR), 8-12 October 2006, Victoria, Canada*, 2006.

[15] M. Sheridan, and C. Byrne, “Ebb and flow of assessment in music”, *British Journal of Music Education* 19, Cambridge University Press, Cambridge, 2002, pp. 135-143.

[16] P.J. Silvia, and A.G. Phillips, “Self-awareness, self-evaluation, and creativity”, *Personality and Social Psychology Bulletin* 30, SAGE, pp. 1009-1017.

# Visualization of bowing gestures for feedback: The Hodgson plot

Erwin Schoonderwaldt[†,‡] & Marcelo M. Wanderley[‡]
[†]*Dept. of Speech, Music and Hearing, KTH, Stockholm, Sweden*
[‡]*Input Devices and Music Interaction Laboratory, McGill University, Montreal, Canada*
*E-mail: schoondw@kth.se; marcelo.wanderley@mcgill.ca*

## Abstract

*A set of displays is proposed for the visualization of bowing gestures measured using motion capture techniques. The main displays (Hodgson plots) show the spatial trajectory followed by the bow frog in time in two different projections. The bridge and the strings of the instrument are shown in the background, forming a functional context for the displayed bowing gestures. The main purpose of the visualizations is to provide informative feedback to players regarding their use of the bow, making them suitable for pedagogical use.*

## 1. Introduction

Motion capture (MoCap) techniques have been proven useful for the analysis of bowing gestures in bowed-string instrument playing. The obtained data is charactarized by a high temporal and spatial resolution allowing for detailed analysis of timing and coordination [1, 2], extraction of bowing parameters such as bow speed and bow-bridge distance [3, 4], and the study of the kinematics and kinetics of players in relation with the development of injuries [5, 6, 7].

As is well known the player of a bowed-string instrument exerts direct control over the produced sound with the bow, mainly by varying bow speed, bow-bridge distance and bow force (i.e., the normal force exerted by the bow on the string). The production of a good tone under voluntary control of the player requires a subtle coordination of these bowing parameters. The creation and maintenance of a regular string vibration (Helmholtz motion) imposes physical constraints on the possible combinations of bowing parameters [8, 9]. Furthermore, the serial execution of bow strokes in the context of a musical piece requires planning ahead to optimize bow distribution, not to mention the wide variety of different bowing techniques the player has to master. Gaining control over the bow is therefore one of the major goals in learning to play a bowed-string instrument.

In the light of the previous the possibilities offered by MoCap are potentially interesting for bowed-string instrument teaching to provide feedback to the player on his/her use of the bow. The most obvious way to achieve that is by means of visualization. The cyclographs made by Hodgson [10] represent – as far as known by the author – the first photographic images of bow motion during violin performance used for pedagogical purposes. More recently, visual displays of quantitatively measured bowing gestures have been developed by Ho [11], Rabbath & Sturm [12], and Ng et al. [13]. These contributions show clearly the potential of the use of technology in instrumental teaching. However, most of these approaches are based on a rather implicit notion of feedback, assuming that what is shown will somehow have the desired effect. What still fails is a vision of how the feedback can be understood and utilized by students, and related to their playing skills.

In this paper visualization methods of bowing gestures recorded using MoCap and/or other quantitative sensing techniques are proposed. During the design special effort was made to make the displays as accessible and informative as possible to enhance the communication of feedback to the student.

## 2. Visualization of bowing gestures

### 2.1. Measurement of bowing gestures

The proposed visualization methods require an accurate measurement of position and orientation (6 degrees-of-freedom) of both the bow and the instrument. In addition, the positions of important landmarks on the bow and the violin (bridge, strings, hair ribbon) must be known, either via direct measurement or reconstruction. The visualization methods can in principle be applied on data obtained via different measurement techniques (e.g., [3, 4, 13,

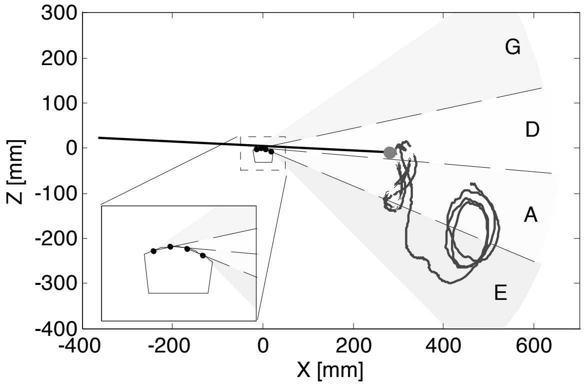



**Figure 1. Hodgson plot in orthographic back projection (xz-plane). The red dot indicates the position of the frog at the "present" moment (i.e., at the end of the selected time interval). The solid black line corresponds approximately to the bow-hair ribbon from the frog to the tip, ignoring the bending of the hair at the bow-string contact point. The trajectory history of the bow frog is indicated by a blue line, shown as solid and fat when the bow was in contact with the string, thin and dotted otherwise. In the background, the bridge, string positions and string crossing angles are shown (see close-up for more detail), forming the functional context of the displayed bowing gestures. The string crossing angles (dashed lines) subdivide the space into 4 angular zones associated with the bowing of the different strings. The zones are indicated with different pastel colors: blue (E string), green (A string), yellow (D string) and red (G string).**

14]). Technical details about the measurement methods are therefore out of the scope of this paper.

For the measurements shown in this paper a six-camera Vicon system was used for motion capture. Bow force was measured using a custom-made sensor, developed by Matthias Demoucron (IRCAM). For more details of the methods used the reader is referred to Schoonderwaldt et al. [3].

## 2.2. Design criteria

The major goal of the visualizations is to provide informative feedback on the use of the bow, as this forms an important element in the practicing process. According to Ericsson et al. [15] three requirements need to be fulfilled for deliberate practice: (1) a well-defined task, (2) informative feedback, and (3) opportunities for repetition and correction of errors. Feedback can hereby be understood as a "process by which an environment returns to individuals a portion of the information in their response output necessary to compare their present strategy with a representation of an ideal strategy" [16]. It has been shown that technology can successfully enhance teaching of complex musical tasks when implemented according to these criteria [17].

For a clear presentation of the feedback the following criteria were taken into account for the design of the visual display. Firstly, the display should be easy to understand for musicians without a scientific background. The visualizations should mainly be self-explanatory and the information should be presented in such a way that the player can easily relate it to his/her actual playing. Secondly, the display should contain relevant information giving the player an idea of how to improve his/her performance. A third criterion was that the display should not be normative in itself. The representation of an ideal strategy (see the above definition of feedback) should arise from comparison with other performers or self-exploration, rather than being imposed via norms and fixed criteria. This should make the display more versatile and easier to integrate in different teaching approaches.

## 2.3. Hodgson plots

The visual displays proposed in this paper are based on the cyclographs presented in Hodgson's book [10], and will therefore further be referred to as "Hodgson plots." In the current implementation Hodgson plots show basically the spatial trajectory followed by the bow frog during a chosen time span (typically 1 s). This provides a simple representation of bowing gestures with a direct relation to the actions of the right hand of the player.

The acquisition of 3D data using MoCap in combination with calibrated geometrical models of the bow and the violin allows for some important additional features. Firstly, the motion of the bow can be transformed to the reference frame of the violin, showing only the effective bowing gestures related to sound production. Thus, there is no need to constrain the movements of the player, allowing for natural playing conditions. Secondly, different projections can be chosen. This allows for example to show the bowing gestures from the perspective of the player to strengthen the association with his/her own actions. Finally, it is possible to visualize important landmarks on the violin, such as the bridge, the strings and the angles corresponding with string crossings, in order to provide a functional context for the displayed bowing gestures.

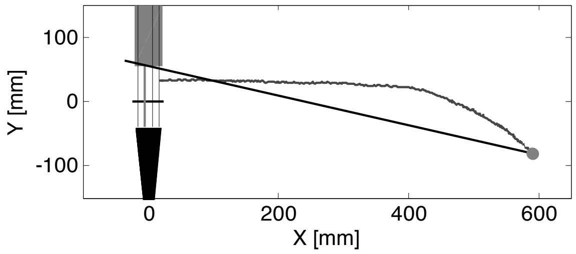



**Figure 2. Hodgson plot in orthographic top projection (xy-plane). The bow and the frog trajectory history are shown in a similar way as in Fig. 1. The context is formed by the 4 strings (vertical lines), the bridge (bold horizontal line), the fingerboard (gray rectangle) and the tailpiece (black shape), based on the specific measures of the instrument. To enhance the clarity, the string played at the "present" moment (i.e., the moment the bow is shown) is highlighted in red.**

Two types of Hodgson plots are proposed representing different orthographic projections, which together cover the main aspects of the motion of the bow. In the back projection (Fig. 1) the violin is more or less seen from the player's perspective. This projection is especially suited for showing complex bow coordination patterns involving bow changes and string crossings (see Hodgson [10] for an extensive overview of different types of patterns). The fragment shown in Fig. 1 is a selection of about three seconds of a performance of "Praeludium and Allegro" composed by F. Kreisler. It contains two clearly distinguishable coordination patterns: semi-quavers across two strings played détaché (circle-shaped pattern) and semi-quavers across three strings played spiccato (eight-shaped pattern). A wide variety of information can be obtained from the displayed patterns, for example about the bow distribution (bow position, amount of bow used), the regularity of the motion and the efficiency of the string crossings.

The top projection (Fig. 2) shows the violin from above. This projection gives a good sense of the bow-bridge distance and the skewness of the bow. The frog trajectories might also illuminate details of changes in bowing direction, which according to empirical findings follow curved rather than straight paths [10, 18, 19]. The example shown in Fig. 2 represents a long decrescendo note played down bow. It can be seen that at the end of the bow stroke the bowing was far from perpendicular to the string. This should, however, not be considered as a fault as it has been demonstrated that the skewness of the bow can be utilized to change the bow-bridge distance dynamically during the bow stroke [3]. In this particular case the skewness of the bow was used to drive the bow towards the fingerboard in order to accomplish a diminuendo note.

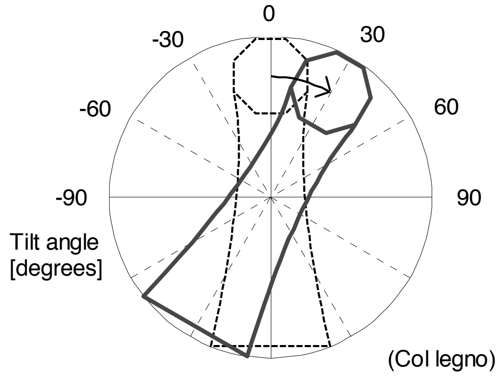



**Figure 3. Visual display for bow tilt. The keyhole-like shape represents the bow frog when looking at it along the direction of the stick. Tilt is shown as a rotation of the frog in a clock-like display. When the stick is turned away from the player (as during normal playing), this is shown as a clockwise rotation (30 degrees in this example). For col legno playing tilt angles of 90 degrees or more are employed, clockwise or anti-clockwise depending on the preference of the player.**

The above described projections allow for an effective visualization of the inclination and the skewness of the bow. For the visualization of bow tilt, another important bow control parameter, a third projection is added showing the rotation of the bow frog relative to the string played in a clock-like context (Fig. 3). During normal playing, the bow is often tilted so that the stick is turned slightly away from the player. This corresponds to a clockwise rotation in the tilt display. The tilt angle is easily quantifiable, realizing that an angle of 30 degrees corresponds with 5 minutes on the clock (in pp playing close to the frog bow tilt can reach up to about 45 degrees).

## 2.4. Additional displays and animations

The Hodgson plots described in the previous section provide a clear insight in the positioning and angling of the bow. This information is, however, not yet complete from an acoustical point of view, bearing in mind that tone production is mainly governed by bow speed, bow force and relative bow-bridge distance at the bow-string contact point. During the attack bow acceleration is also an important parameter. For a more

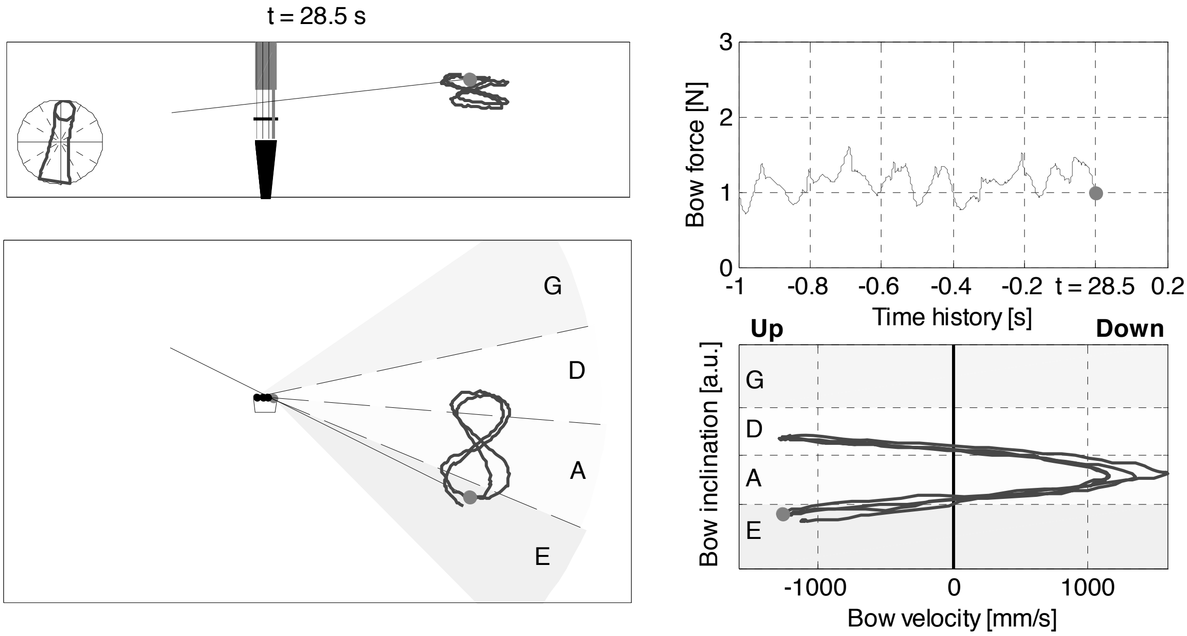



**Figure 4. Feedback display showing a combination of aspects of the use of the bow. The panels on the left side show the Hodgson plots in the two projections, as well as bow tilt. The two panels on the right side show additional information on the use of the bow. Depending on the purpose of the exercise different bowing parameters might be displayed here. In this example bow force versus time (present moment and history) is shown in the top-right panel, and a phase-like representation of bow inclination versus bow velocity is shown in the down-right panel. The background colors used in the latter are meant to strengthen the association with the Hodgson plot in the down-left panel.**

adequate feedback on tone production additional visualizations are needed to present this information to the player.

In Fig. 4 a total of five panels are combined for a more complete representation of the use of the bow. The example shows the beginning of the arpeggio part from “Preludio” of the third Partita for solo violin by J.S. Bach, played legato across three strings. As in the first example an eight-shaped bowing pattern can clearly be seen in the Hodgson plot (back projection). In the down-right panel the inclination of the bow is plotted versus bow velocity. This display contains more explicit information related to sound production, for example regarding the coordination between bow changes and string crossings. It can be seen that the bow speed in this fast passage reached rather high values of more than 1 m/s in order to obtain a loud sound. The bow force (upper-right panel) was rather constant, varying about 1 N.

Even if the static displays carry a lot of information, they do not yet provide a direct link with the sound. This can be achieved by animating the displays with synchronized sound. This was done making use of the QuickTime tools for Matlab by Slaney [20]. The resulting movies give players the possibility to analyze their bowing by repeatedly playing them back, paying attention to the different aspects of bowing. The movies also allow the players to scroll through their performances and search for different passages. Another advantage is that the movies can be played using a standard media player, which makes the prepared visualizations more accessible for players and teachers.

# 3. Discussion

The Hodgson plots, in combination with other types of displays, have the potential to provide informative feedback to the player. However, it should be realized that the way they are implemented in teaching and/or practicing is of vital importance for a successful pedagogical application. Further field studies are needed for the development of dedicated exercises and a database of reference performances, as well as an assessment of the usability. Moreover, little is known about expert bowing skills in musical performance as most quantitative studies of bowing are limited to relatively simple tasks. The details shown in this type of displays might be in conflict with popular beliefs in bowed-string pedagogy, as was for example the case with Hodgson's cyclographs [18].

Another interesting possibility would be to show the visualizations in real time serving as an enhanced mirror for the player as for example envisioned by Fober [21] and Ng et al. [13]. This might further strengthen the association with the player's own actions and allow for a more explorative use.

An important limitation is that an accurate measurement of bowing gestures is rather tedious and confined to the lab due to the need for expensive equipment. This forms an obstacle for a widespread use of this technique. However, the current state-of-the-art of the technology would allow for the implementation of this kind of technologies in a school environment as for example realized for piano pedagogy [22].

## 4. Conclusions

A set of displays is proposed for visualization of bowing gestures measured using motion capture techniques. The main displays shows the spatial trajectory of the frog in time, and are named after Percival Hodgson, who was the first to show photographic images of frog trajectories in violin playing. The proposed displays are mainly an extension of Hodgson's visualizations, showing the motion of the bow in the functional context of the violin. The presence of the context makes these plots both easier to understand and more informative to the player.

Different projections can be chosen to show different aspects of bowing. In the back projection (quasi player's perspective) the motion of the bow and the trajectory of the frog are shown in the context of the specific string crossing angles of the violin. This projection is especially suited for showing complex bow coordination patterns involving bow changes and string crossings. In the top projection bow-bridge distance and skewness of the bow can be clearly observed.

It is believed that Hodgson plots can provide informative feedback to violin and other bowed-string instrument players on their use of the bow. Especially when animated or shown in real time Hodgson plots - in combination with other visual displays of bowing parameters – can be used to illuminate the relationship between bowing gestures and the produced sound, allowing players to analyze their bowing technique and compare different strategies. The visualization methods could form an interesting tool for music education and could – due to their non-normative character – easily be adopted in different pedagogical approaches.

## 5. Acknowledgements

This work was partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC-SRO) and the Enactive European Network. Special thanks go to Matthias Demoucron for developing and sharing the bow force sensor, and Lambert Chen for his marvelous playing during the pilot experiment and the fruitful discussions.

## 6. References

[1] A. P. Baader, O. Kazennikov, and M. Wiesendanger, "Coordination of bowing and fingering in violin playing," *Cognitive Brain Research*, vol. 23, no. 2-3, pp. 436–443, 2005.

[2] H. Winold, E. Thelen, and B. D. Ulrich, "Coordination and control in the bow arm movements of highly skilled cellists," *Ecological Psychology*, vol. 6, no. 1, pp. 1–31, 1994.

[3] E. Schoonderwaldt, S. Sinclair, and M.Wanderley, "Why do we need 5-DOF force feedback? An analysis of violin bowing." in *Proceedings of Enactive07 Conference*, 2007.

[4] E. Maestre, J. Bonada, M. Blaauw, A. Pérez, and E. Guaus, "Acquisition of violin instrumental gestures using a commercial EMF tracking device," in *Proceedings of the 2007 International Computer Music Conference*, Vol. I. The International Computer Music Association, 2007, pp. 386–393.

[5] G. Shan and P. Visentin, "A quantitative three-dimensional analysis of arm kinematics in violin performance," *Medical problems of performing artists*, vol. 18, no. 1, pp. 3–10, 2003.

[6] L. Turner-Stoker and K. Reid, "Three-dimensional motion analysis of upper limb movement in the bowing arm of string-playing musicians," *Clinical Biomechanics*, vol. 14, pp. 426–433, 1999.

[7] P. Visentin and G. Shan, "The kinetic characteristics of the bow arm during violin performance: an examination of internal loads as a function of tempo," *Medical problems of performing artists*, vol. 18, no. 3, pp. 91–97, 2003.

[8] K. Guettler, "On the creation of the Helmholtz motion in bowed strings," *Acustica - Acta Acustica*, vol. 88, no. 6, pp. 970–985, 2002.

[9] J. C. Schelleng, “The bowed string and the player,” *J. Acoust. Soc. Am.*, vol. 53, no. 1, pp. 26–41, 1973.

[10] P. Hodgson, *Motion study and violin bowing*. American String Teachers Association, 1958.

[11] T. K. Ho, “A computer-assisted approach to the teaching of violin tone production,” *ACM SIGCUE Outlook*, vol. 21, no. 2, pp. 73–83, 1991.

[12] F. Rabbath and H. Sturm, “Art of the bow with Francois Rabbath,” Website.
[Online]. Available: http://www.artofthebow.com/

[13] K. Ng, O. Larkin, T. Koerselmans, B. Ong, D. Schwartz, and F. Bevilacqua, “The 3D augmented mirror: motion analysis for string practice training,” in *Proceedings of the 2007 International Computer Music Conference*, Vol. II. The International Computer Music Association, 2007, pp. 53–56.

[14] D. Young and A. Deshmane, “Bowstroke database: a web-accessible archive of violin bowing data,” in *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07)*, L. Crawford, Ed., New York, NY, USA, 2007, pp. 352–357.

[15] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, “The role of deliberate practice in the acquisition of expert performance,” *Psychological Review*, vol. 100, pp. 363–406, 1993.

[16] W. K. Balzer, M. E. Doherty, and R. O’Connor, “Effect of cognitive feedback on performance,” *Psychological Bulletin*, vol. 106, pp. 410–433, 1989.

[17] P. N. Juslin, J. Karlsson, E. Lindström, A. Friberg, and E. Schoonderwaldt, “Play it again with feeling: Computer feedback in musical communication of emotions,” *Journal of Experimental Psychology*, vol. 12, no. 2, pp. 79–95, 2006, journal of Experimental Psychology: Applied.

[18] P. Hodgson, “Motion study and violin bowing,” *The Musical Times*, vol. 76, no. 1106, pp. 347–348, apr 1935.

[19] C. Williams, “Violin bowing skill analysis: the mechanics and acoustics of the change in bowing direction,” Ph.D. dissertation, University of Melbourne, 1985.

[20] M. Slaney, “MakeQTMovie - create QuickTime movies in Matlab,” Interval Technical Report 1999-066, Tech. Rep., 1999.
[Online]. Available: http://web.interval.com/papers/1999-066/

[21] D. Fober, “Miroirs technologiques pour la pratique instrumentale,” in *Le feedback dans la création musicale, rencontres musicales pluridisciplinaires*, 2006.

[22] G. Comeau, “Piano pedagogy research laboratory,” Website. [Online]. Available: http://www.piano.uottawa.ca

# The MAV Framework: Working with 3D Motion Data in Max MSP / Jitter

Thijs Koerselman, Oliver Larkin, Kia Ng
*ICSRiM - University of Leeds, School of Computing & School of Music, Leeds LS2 9JT, UK*
*info@i-maestro.org, www.i-maestro.org*

## Abstract

*This paper presents the Motion Analysis and Visualisation (MAV) Framework – an extension for Cycling 74's Max MSP / Jitter, designed to support working with 3D motion capture data both in real time and offline situations. We describe the technical implementation of the MAV framework and the application of this system in developing a teaching tool for string practice training as part of the i-Maestro EC-IST project.*

## 1. Introduction

Cycling 74's Max MSP / Jitter (www.cycling74.com) has become one of the most popular environments for developing musical research and multimedia performance applications. It offers a large number of objects designed for audiovisual processing and synthesis and an intuitive data-flow based visual programming model, which is accessible to many different types of users. 3D motion capture systems such as Vicon [7] are being used more and more in research into performance-arts [e.g. 2, 3]. As well as scientific analysis, these systems have artistic potential as an interface for controlling sound and visuals [1]. There are also many potential educational applications of this technology [4, 5]. As motion capture systems become more readily available to researchers and performers, we see a need to be able to interface these systems easily within applications such as Max MSP. This paper presents ongoing work on a library of Max / Jitter objects called the Motion Analysis and Visualisation (MAV) Framework, that facilitate the interface with a motion capture system and work with Motion Capture data in Max. Currently the MAV framework is tailored to our own specific application (see below) although it is designed in such a way that in the future it may be extended to provide a more generic solution for dealing with motion data in Max.

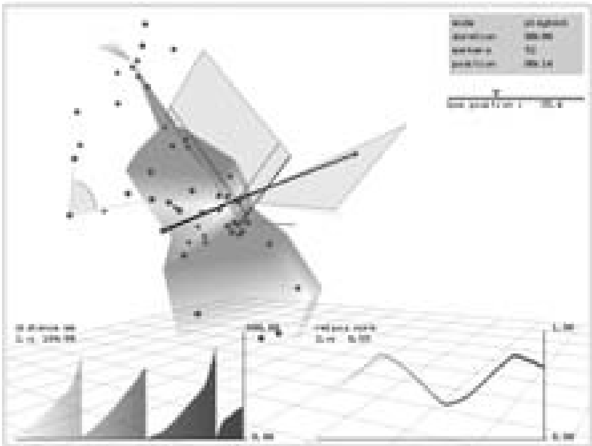


**Figure 1: AMIR Screenshot**

## 2. Our Application

The i-Maestro project (http://www.i-maestro.org) aims to explore novel solutions for music training with a focus on bowed string instruments. One part of the project involves using motion capture technology to develop a tool to support string practice training. We aim to provide feedback about the performer's bowing gesture and body posture which can be used by a teacher to illustrate/identify certain techniques/issues or by a student to study their performance. We call this tool the "3D Augmented Mirror" (or AMIR for short) since we see its use as similar to the traditional use of a mirror in instrumental training. On a basic level the performance is visualised in 3D with synchronised audio and video data. The software allows the user to manipulate the environment in order to study the performance from different perspectives. At a more advanced level the motion data is analysed in various ways to study characteristics of the performance in detail. These analyses can be performed in real time or in an offline context (for example after the user has recorded the performance). The output of the analyses can be visualised with graphs or they can be linked to sonification algorithms. For more information see [5].

## 4. The MAV Framework

The MAV framework consists of a cross-platform C/C++ library and a suite of Max objects for performing various motion data handling, analysis and visualisation tasks. The design uses low-level data processing and dynamic binding to realise a highly efficient and flexible system. The objects are based on the Jitter API which offers greater flexibility over the standard Max API, providing functionality for inter-object communication, attributes and runtime scripting. The applications of these features within the framework are described in Sections 6 and 7.

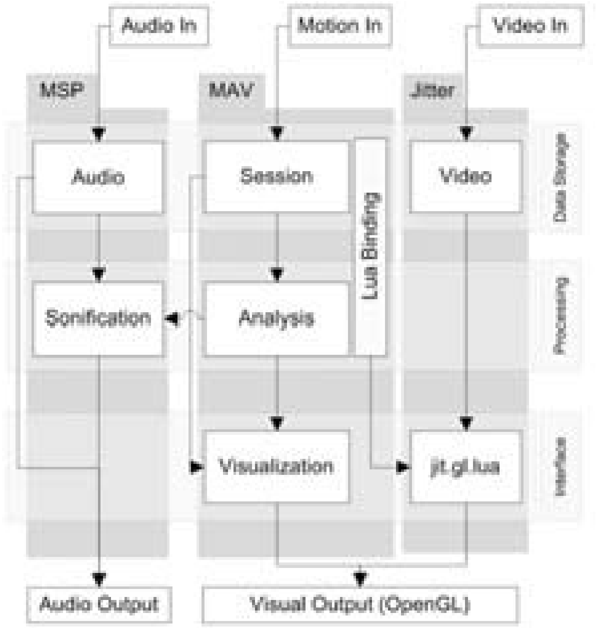



**Figure 2: MAV Framework Overview**

Figure 2 shows a high level overview of the MAV framework structure and how it integrates with Jitter and MSP objects within the AMIR application.

The framework can be divided into three processing layers. The first layer takes care of all data management (e.g. playback, recording and file i/o). All motion data is handled by a "session" object, which serves as a common data resource for all other MAV objects connecting to that specific session data. An arbitrary number of session objects can co-exist within the same system for playing back and analysing multiple datasets simultaneously. Standard Jitter and MSP objects are used to record and playback audio and video data. The three different types of media contained within the data layer are synchronised by a clock signal.

The second layer contains the MAV analysis and statistics objects described in Sections 8 and 9. Their output can be displayed as feedback in the visualisation layer, used for sonification, or used as an input source for other MAV objects.

The third layer contains the graphical objects for displaying user feedback. The MAV framework currently includes several objects based on the Jitter OB3D / OpenGL API for applications that involve a substantial amount of processing and data access, such as "running" graphs and motion trails (Section 10). All other drawing available in AMIR is managed by low-level OpenGL calls in Lua (Section 6).

## 5. Interfacing with Vicon

To acquire the motion data we use a Vicon 8i motion capture system. To send the data from the Vicon real-time processing engine into Max, we developed a bridge application which requests a data stream from the system and forwards the relevant information using the TCP/IP protocol. Figure 3 shows an overview of the Vicon bridge application.

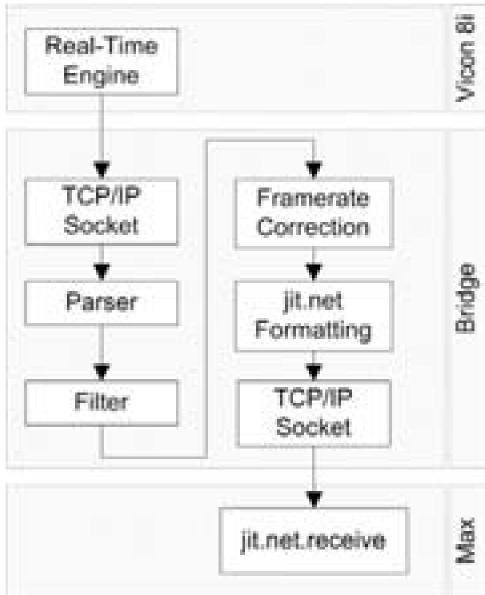



**Figure 3: Vicon / Jitter Bridge**

In the first stage the incoming data is parsed and filtered to provide the raw marker positions and labels, omitting any modelling or session information from the Vicon setup.

The second stage contains an interpolation algorithm to account for any possible dropped frames. We have found that the Vicon 8i real-time engine is unable to deliver a reliable data stream at a fixed high frame rate. In a typical motion capture situation this is not an issue, since the Vicon software provides extensive post processing to fix discontinuities in the recorded data. For our application it is necessary to work with the motion data input in real-time and therefore it is essential to acquire a reliable and solid data input stream. Using the frame-rate correction we output the data at the maximum rate allowed by the system, which is 200 Hz with our current hardware setup.

The MAV framework is by no means limited to the use of this particular motion capture system. Any system capable of producing 3D position data should be compatible. However a custom bridge application might be required to interface Max with other hardware systems besides the Vicon 8i.

## 6. Dynamic Linking and Scripting

Every MAV object is able to register itself under a unique ID. This ID is used by other objects to “find” their link target object with a given ID and retrieve a pointer to its structure allowing objects to call the targets member functions, attributes and access internal data structures. The MAV library makes use of this technique to share motion and analysis data amongst different objects and process layers.

Since the MAV objects are based on the Jitter API, they are compatible with the scripting facilities within Max. Scripting allows for object compositions to be instantiated and modified in real-time and outside of the traditional Max programming paradigm of the “patcher” window. The main advantage of scripting is that it enables the application to be dynamically adjusted for different scenarios and setups.

The use of run-time linking and object scripting require facilities to ensure data integrity. In a dynamic environment, the order in which objects are created and their lifespan are both uncertain. Every MAV object requiring inter-object data sharing makes use of a special link module which utilises Max’s Pattr SDK for client-server notifications. The link module also uses a technique common to Jitter objects called “lazy registration” in which the actual link to the target object is resolved upon the next process call. In a scenario where a server object changes ID or decides to free its resources the clients using those resources are notified of the fact that the data they are accessing is no longer valid and the link will break.

Once objects are scripted instead of being contained within in a Max patcher, they loose their traditional connections to output data to a patcher. In the case of AMIR (Section 2) where we use the output of analysis modules as the input for sonification algorithms, we need the MAV objects to be able to send their output to standard Max/MSP objects. For these kinds of situations all MAV objects are able to link to a named outlet in a Max patcher, allowing them to connect their processing output to any standard object.

## 7. Lua Bindings

Lua [8] is a light-weight, fast, and extensible scripting language which can interface easily with C/C++ libraries and applications. Lua support for Jitter is provided by the jit.gl.lua external developed by Wesley Smith [6]. This object contains a Lua interpreter with additional bindings for Jitter and low level OpenGL based on the LuaGL library. These bindings make it possible to script and control Jitter objects as well as perform OpenGL function calls on the available rendering context.

The role of Lua within the AMIR application can be divided into two parts: 1) Scripting and controlling MAV objects. 2) Displaying basic user feedback and interface related graphics.

Bindings were added to the MAV library in order for the Lua scripts to access data from the different objects in a more direct and convenient way.

## 8. Motion Analysis

The current set of MAV analysis objects is able to extract the following basic features from the motion data:

- speed
- acceleration
- distance traveled
- vector angles
- rotation

The first three of these are easily extracted from the raw position data of any individual marker. Vector angles are calculated by specifying two vectors and calculating the angles between them using dot-product calculations. These vectors could for instance be the upper and lower arm of the body, or the head in relation to the spine. It is possible to measure any angle providing the markers are placed at the correct positions on the body. Rotation can be determined for any set of three markers, by extracting a transformation matrix using cross-product vector calculations, given that they share a common plane in 3D space.

Since our system is aimed at the analysis of string instrument performance, we have developed a dedicated object to segment and study bowing movements. Initially a local coordinate transformation matrix is extracted from three specified markers positioned on the instrument body, typically placed in a way that the origin can be fixed on the bridge with the Y-axis aligned along the neck of the instrument (Figure 4). This first step enables us to analyse the bowing movements in relation to the instrument as opposed to the world coordinates, so that the performer can move their instrument and change position while playing without affecting the analysis.

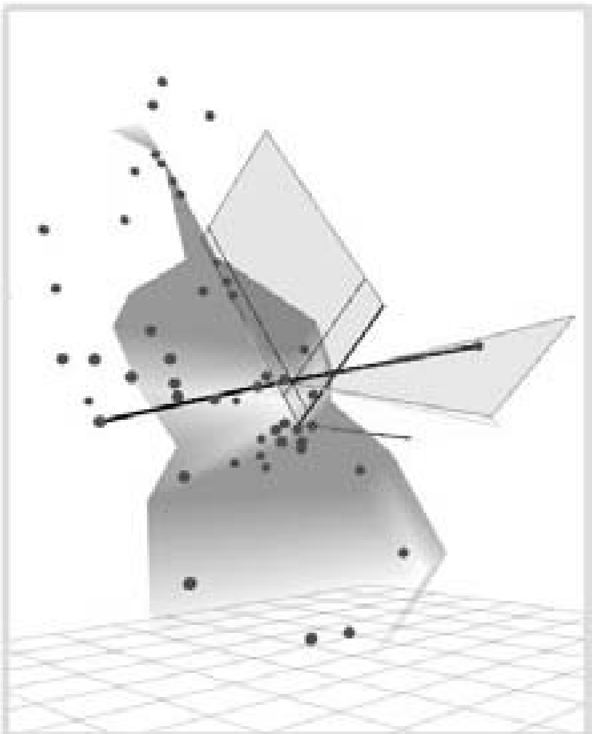


**Figure 4: Local coordinate system on the bridge of the Cello**

This transformation matrix is then used to extract the point where the bow is crossing the centre-plane of the instrument as shown using the thin line crosshair on the grey plane in Figure 4. Once this point is calculated we are able to segment the movement of the bow in three categories:

- up bow
- down bow
- not on strings

Using this local coordinate system the following features can be extracted:

- angles between bridge and bow
- distance between bridge and bow
- part of the bow being used
- speed of the bow movement over strings

The angles between the bow and the bridge in the XY and XZ planes (Y-axis along the neck of the instrument) are illustrated in Figure 4 by the triangles on the right-hand side of the picture.

## 9. Data Correlation and Clustering

The next step in analysis is to extract high-level information that might tell us something useful about the character of the movements rather than the low-level values which were described in the previous section. Data correlation and feature clustering can be used to analyse the similarities between movements and classify them according to what the system is trained to recognise.

To compare movements the motion data first needs to be segmented appropriately, depending on the type of motion input and the movements that one wishes to analyse, after which a number of relevant features can be extracted from every segment.

In AMIR we use the bowing segmentation algorithm to separate each bow stroke and process each segment in order to extract features such as average speed, bow length, and bow centre position. These features can help us to classify the type of bowing technique that’s being used, and how consistent the individual movements are.

The MAV framework features an object which is designed to perform statistics analysis analyses both offline and online. This object reuses other objects already available without disturbing their operation. Since the objects are already configured, reusing those objects provides an elegant and convenient way of extending the running analysis with statistics and correlation in a modular way without duplicating settings and algorithms.

The online statistics mode listens to the output of the bowing segmentation object and triggers the statistics processing of the last segment as soon as it is finished. To allow the statistics object to briefly use the algorithms without disturbing the real time process, a facility is created to store and restore every objects internal algorithm state, using a snapshot of all relevant data for each individual object.

Figure 5 shows a simplified sequence diagram of a single segment being processed by the statistics object in online mode whilst motion data playback, recording and/or analysis are running synchronously.

The "session data" object represented in this diagram is merely a motion data output buffer contained within the actual session object, and not the complete performance dataset. The use of this buffer is outside of the scope of this paper, but it is important to notice that the recorded data stays untouched by this process.

Depending on the nature and the amount of algorithms used by the statistics object, this process is very likely to happen in a fraction of the screen refresh rate, leaving the user interface running smoothly.

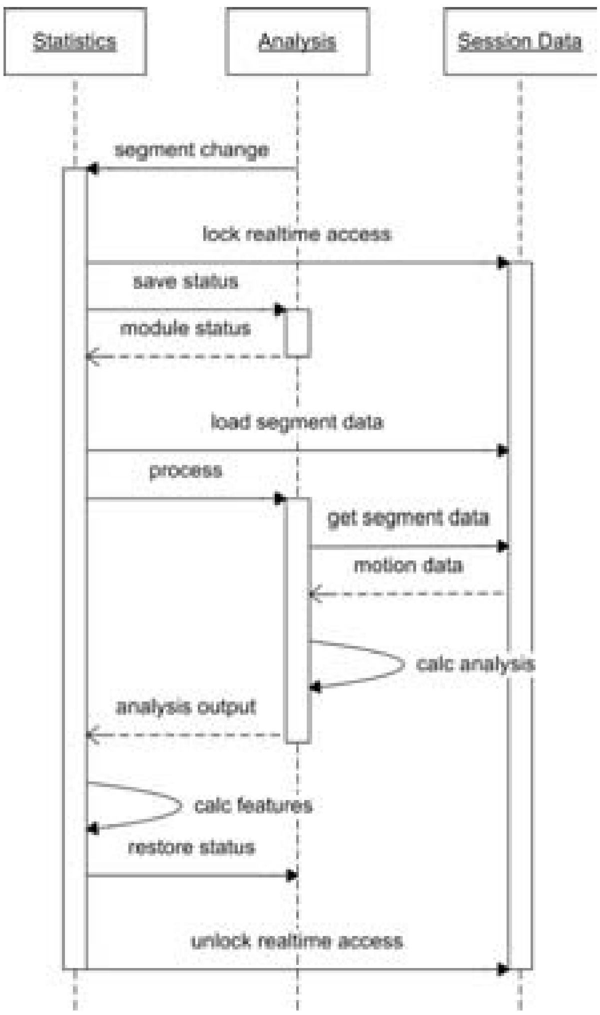



**Figure 5: Statistics operation in online modus**

## 10. Visualisation

The MAV library currently only contains native visual objects for drawing routines that are either inefficient or inconvenient to do otherwise. All other drawing used to build the AMIR user interface are either realised through LuaGL or standard Jitter objects.

A 2D running graph object can display floating point output from any of the analysis objects. The graph contains a data buffer for providing a variable time window and is able to display multiple data streams at once in different layers and in different visual modes.

Another object is dedicated to motion trails (Figure 6), which are drawn from a variable length motion data buffer that can be scaled according to the preferred time window. The data cache can be specified for an arbitrary number of points allowing individual trails to be drawn simultaneously at runtime. We are currently improving the trail drawing routines to appear as true volumetric shapes.

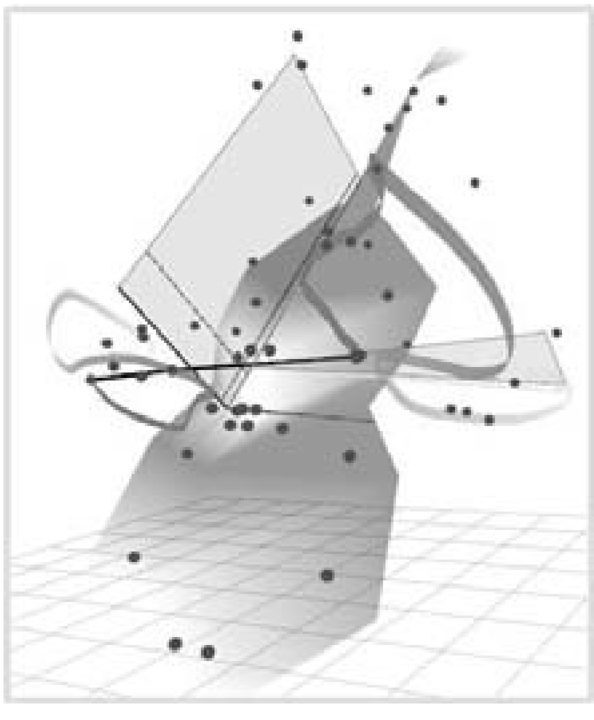


**Figure 6: Motion Trails**

Another object is dedicated to the visualisation of the feature clustering described in Section 9. We are developing multidimensional graphs to plot the feature clusters onto the screen by combining 3D geometry and colour to represent multiple features using a single object.

## 11. Conclusion and Future Work

We described the technical implementation of the MAV framework - a library for recording, analysing and visualising motion capture data within the Max MSP / Jitter environment. We discussed our application of the framework in a tool to support string practice training.

Future work on MAV will include further development of the framework including more analysis and visualisation objects. We aim to move towards a generic open-source framework which can be used and extended by 3rd parties. We are especially interested in the possibilities of using the processed motion data as a controller for creative applications such as multimedia performance and composition.

## 12. Acknowledgements

The i-Maestro project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i-maestro.org). Thanks to all i-Maestro project partners and participants, for their interest, contributions and collaboration.

## 13. References

[1] Dobrian C., and Bevilacqua, F. "Gestural Control of Music Using the Vicon 8 Motion Capture System". In Proceedings of the International Conference on New Interfaces for Musical Expression Conference (NIME03), Montréal, Canada. 2003

[2] Wanderley M.M. "Quantitative Analysis of Non-obvious Performer Gestures". In I.Wachsmuth and T. Sowa (eds.) *Gesture and Sign Language in Human-Computer Interaction*. Springer Verlag, pp. 241-253. 2002

[3] Neagle, R.J., Ng, K.C., Ruddle, R.A., Notation and 3D Animation of Dance Movement, in Proceedings of the International Computer Music Conference (ICMC2002), Göteborg, Sweden, pp. 459 – 462, 2002.

[4] Mora, J., Won-Sook, L., Comeau, G., Shirmohammadi, S., El Saddik., A. "Assisted Piano Pedagogy through 3D Visualization of Piano Playing", Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and Their Applications (HAVE' 2006), Ottawa, Canada, 2006.

[5] Ng, K., Larkin, O., Koerselman, T., Ong, B., Schwarz, D. Bevilaqua., F. "The 3D Augmented Mirror: Motion Analysis For String Practice Training", in Proceedings of the International Computer Music Conference (ICMC2007), Copenhagen, Denmark, 2007.

[6] Wakefield, G., Smith, W. "Using Lua for Multimedia Composition", Proceedings of the International Computer Music Conference (ICMC2007), Copenhagen, Denmark, 2007.

[7] VICON, http://www.vicon.com

[8] Lua, http://www.lua.org

# Pedagogical experimentation using the Sound and Gesture Lab

Fabrice Guédy, Nicolas Leroy, Frederic Bevilacqua, Tobias Grosshauser, Norbert Schnell
*Real Time Musical Interactions*
*CNRS STMS UMR 9912*
*Ircam Centre Pompidou*
*1 place Igor Stravinsky, 75004 Paris – France*
*+ 33 1 44 78 48 31*
*{Fabrice.Guedy, Nicolas.Leroy, Frederic.Bevilacqua, Tobias.Grosshauser, Norbert.Schnell}@ircam.fr*

## Abstract

*This article describes the "Sound and Gesture Lab", a prototype application developed by Ircam in the context of the European project "i-Maestro". This project is focused on technology enhanced music pedagogy, and is more specifically devoted to string instruments teaching.*

## 1. Introduction

The Sound and Gesture Lab is a pedagogical prototype application supporting a teacher and/or a student on specific aspects of music lessons and scenarios. By saving different configurations of this application, various tools may be generated and used, focused on specific needs. A "snapshot" of the state of any tool may be taken at any time, allowing to resume the work on a given topic, at another time.

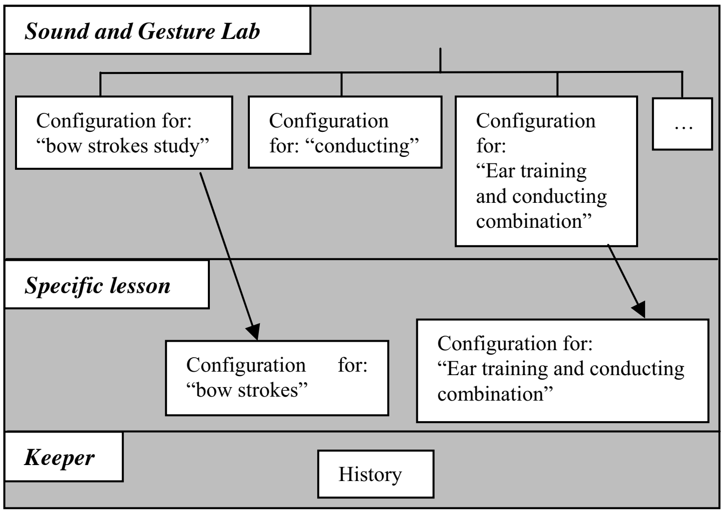
## 2. The pedagogical foundation

The Sound and Gesture Lab has been built following a set of pedagogical prescriptions centered on the following areas:

- Connection between theory and practice
    - Connection between audio and music (extract musical descriptors from audio)
    - Link interpretation with technical issues (for instance locate the cadences on a musical phrase, and relate them with sensor data showing a ritenuto).
- Representation (audio and visual)
    - Of gesture phenomena
    - Of audio phenomena
- Magnification of representations
- Access to previous work
    - History of use of a given tool
    - Searching and browsing inside this history
- Build creative projects
- Learn acoustics
    - See and alter fundamental sound qualities (pitch, timbre, dynamic…)
    - Understand sound synthesis issues, relate them to composition
- Play mixed music (instrument / computer)
    - Pieces
    - Etudes
- Control sound or musical processes with gesture
    - Conducting experiments
- Control sound with sound
    - Extract a specific quality such as brilliance, and make the variation of this quality control another sound or an automated process
- Cooperative work in the context of :
    - creative projects with several students
    - mixed music (with instrument / computer interaction)

This list of pedagogical strategies and goals has been made with the help of a user group of teachers. The most recurring prescription was to avoid excessive intrusion into their normal work and pedagogical habits. This is why the Sound and Gesture Lab's interface allows fast access to pedagogical tools based on the above functionalities, without having to follow a given path or method.

Different tools may be built by teachers to address a given issue in different ways, or “flavors”. Below is a chart showing this “use-path” :




*Use architecture of the Sound and Gesture Lab*

## 3. Implementation

The Sound and Gesture Lab integrates the following technologies in order to be able to match the pedagogical needs listed in the above chapter :

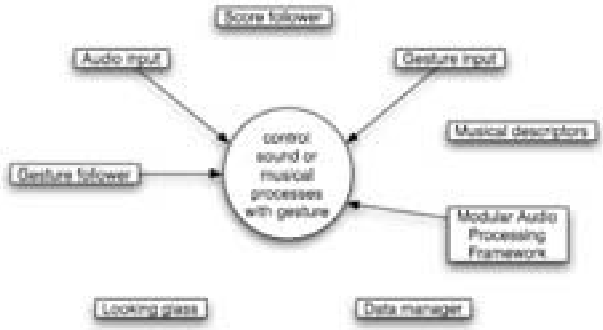
- Audio and gesture input
- Audio analysis
- Gesture-following
- Score-following
- Looking glass
- Data management
- Audio rendering

*Gesture and score following* allow synchronization of a musical representation of music with direct audio and / or gesture input. *Audio analysis* can extract musical qualities in real time from audio input, such as brillance or effort. The *looking glass* functionality allows audio or visual exploration and magnification of such phenomena. *Gesture input* allows capture of “raw” sensor data, the *Data manager* handles file handling functionalities and representation of any data. The *Modular Audio Processing Framework* is a rendering engine allowing sonification, synthesis, and sound processing, that can be controlled by any of the above components.

These technologies may be combined, exchange data, control each other and save their state for later completion.
Each pedagogical tool generated by the Sound and Gesture Lab is a particular combination and use of these technical functionalities. Here are a few examples tools and their relationship with technology. In the following charts, the pedagogical tool in the center is surrounded by the needed technological functionalities written in the rectangles.

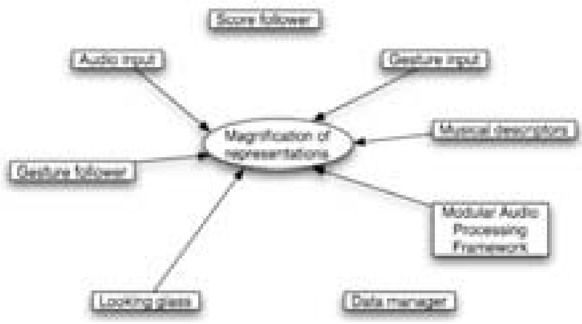
**Examples**

The following tool, for instance, allows the control of a synthesis parameter of the rendering engine (MAPF), by a gesture recognized by the gesture follower. As a variation, the processed sound may come in real time from the audio input. In this last example, while a student play his sound may be processed by the gestures of another student :




*A tool allowing fine subtle sound control with potential benefits in pedagogy*

The following chart shows a second example tool that can align and magnify gesture data with sound quality, in the context of an exercise. In such case, the score follower is not needed:




*A tool for making obvious certain relationships between gesture and sound quality*

## 4. Embedding the Sound and Gesture Lab into a pedagogical workflow

**Temporal issues**

*Short scale and granularity*

The use of such tools into a standard pedagogical context requires identification by the teacher of some common key moments in the flow of a standard lesson, such as:

- the need to emphasize the effect of a given musical gesture on a specific sound quality
- the need to overcorrect a defect

These categories of needs are local and last rarely more than a few minutes in a common lesson.

*Large scale and granularity*

On the other side, the Sound and Gesture Lab may be used in a class during a whole course centered on a topic such as :

- Gestural aspects of conducting
- Acoustics
- Creative projects
- Ear training

**Experimentation**

The following two pictures have been taken from a course where a student recorded an ear training exercise, then recorded the conducting movements synchronized with the recorded sound. Then the student played the ear training exercise to the others, only by conducting it.

| | |
|---|---|
|  |  |
| Recording with the Sound and Gesture Lab | Recording a conducting gesture synchronized with sound |

| | |
|---|---|
|  | The students writing the ear training exercise as the violin student plays it by conducting it measure by measure. |

The benefits of this application of the Sound and Gesture Lab can be viewed from two sides :

*The violin student's benefit.*

In a regular training exercise, the same measure is played at least twice, which allows the player, the second time, to correct an eventual imperfection or approximation of the first time. In the case of that exercise, the student recorded with greater attention, knowing that the only variation when she will conduct it will be temporal.

*The rest of the class benefit*

The other students of the class were particularly attentive to the temporal variation of the violin student since they knew that an eventual lack of smoothness would lead to a rhythm alteration. They tried to detect such effects on sound restitution, thus learning about this aspect of playing.

## 5. The pedagogical "braid"

The student's progresses may be viewed as a complex braid where each wire represents a given skill. At some point of the braid (i.e. at certain moments of the academic year), some skills may seem to stagnate because :

• The student concentrates on a given parameter which he improves, preventing him to make progresses on the remaing aspects of his playing skills, which are in "standby"
• The student make progresses, but they are not yet visible; they will show up at a later time (hidden progresses)

In such cases, the Sound and Gesture Lab may support pedagogical strategies and assist both the teacher and student:

*Support for "standby" progresses*

While working on a given sound quality such as timbre or intonation, the student may use the Sound and Gesture Lab functionality that allow sonification of a single parameter. By hearing only the pitch intonation applied to a sequence where every other parameter is "flattened", the student can keep his concentration on this very parameter without being stressed by the weakness of other playing aspects such as regularity or dynamics.

*Support for hidden progresses*

A usual type of skills involved in such kind of progresses is related to rhythm, especially regularity. While a student working with a metronome may still have problems with regularity, the Sound and Gesture Lab allows him to work on an indirect cause of this defect, such as tenseness, and then align the data related to posture relaxation with a visual representation of his rhythmic accuracy, to observe their relationship and correlation.

This data is not intended to be viewed directly “as is” by the student, but may help the teacher or may be sonified in a given way.

## 6. Strict academic work embedded into creative work

*Ability to play contemporary pieces using computer*

While most European conservatories tend to separate the classes working on classical, baroque and romantic music from contemporary music classes, the Sound and Gesture Lab is a tool intended to be used in these two contexts. While supporting various aspects of musical gesture, the same tool can be used to play “mixed music’ (i.e. music mixing acoustic instruments and computers). This important repertoire is actually very difficult to study and play due to the lack of technical support, which the Sound and Gesture Lab can do.

*Creative projects*

The Sound and Gesture Lab supports various creative project such as:

- *Cross-classes project* involving different fields such as music and literature. These projects are actually becoming more and more important in pedagogy for allowing low level students to gain autonomy, as well as because they move the traditional boundaries of these domains of knowledge.
- *Performances* using sensors
- *Installations* in collaboration of fine arts students.

## 7. Acknowledgements

The i-Maestro project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Program for R&D (IST-026883, www.i-maestro.org). We thank all i-Maestro project partners and participants, for their interests, contributions and collaborations.

## 8. Bibliography

Agon, C. *OpenMusic : Un langage visuel pour la composition musicale assistée par ordinateur*, Thèse de Doctorat, Université Pierre et Marie Curie (Paris 6), 1998.

Agon, C., Assayag, G. et Bresson, J. *The OM Composer's Book .1*, Ircam – Editions Delatour France, 2006.

Bevilacqua F., Rasamimanana N., Fléty E., Lemouton S., Baschet F., “The augmented violin project: research, composition and performance report”, Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06), Paris, 2006.

Bevilacqua, F., Guédy, F., Fléty, E., Leroy, N., Schnell, N. « Wireless sensor interface and gesture-follower for music pedagogy », International Conference on New Interfaces for Musical Expression, New York, 2007

Bresson, J., Guédy, F. et Assayag, G. "Musique Lab Maquette : approche interactive des processus compositionnels pour la pédagogie musicale", *Revue STICEF (Sciences et Techniques de l'Information et de la Communication pour l'Éducation et la Formation*, Volume 13, 2006.

Bresson, J. "Processus compositionnels et opérateurs musicaux dans ML-Maquette – Les opérateurs de traitement du signal", soumis aux *Journées d'Informatique Musicale JIM-07*, Lyon, 2007.

The Collected Works of John Dewey, 1882-1953, edited by Jo Ann Boydston (Carbondale and Edwardsville: Southern Illinois University Press, 1969-1991)

Ferguson, S., Vande Moere, A., and Cabrera, D. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In Proceedings of the Tenth Symposium on Information Visualisation, London, UK, 2005. IEEE.

Fléty E., Leroy, Ravarini N. J., Bevilacqua F., “Versatile sensor acquisition system utilizing Network Technology”, Proc. of the International Conference on New Interfaces for Musical Expression (NIME 04), Hamamatsu, 2004

Fober, D., Letz, S., Orlarey Y., L'environnement de composition musicale ElodyActes des Journées d'Informatique Musicale - JIM'97, Lyon, 1997.

Guédy, F. « La Rousserolle Effarvatte », Actes des 10èmes Journées d'Informatique Musicale, Montbéliard, 2003

Guédy, F. « Musique Lab 2, un environnement pour la pédagogie musicale ». L'Inouï, Vol. 2, Ircam – Editions Léo Scheer.

Guédy, F. « Le traitement du son en pédagogie musicale ». L'Inouï, Vol. 2, Ircam – Editions Léo Scheer.

Iazzetta, F. (2000). "Meaning in Music Gesture". In: Trends in gestural Control of Music, Marc Battier & Marcelo M. Wanderley (Eds.) Paris-IRCAM - Centre Pompidou - CD-ROM.

Puig V., Guédy F., Fingerhut M., Serrière F., Bresson J., Zeller O. (2005). Musique Lab 2: A Three Level Ferguson, S. Learning Musical Instrument Skills Through Interactive Sonification. NIME 2006: 384-389

Approach for Music Education at School. Procceedings of the International Computer Music Conference, Barcelona, Spain.

Schnell, N., Schwarz, D. "Gabor, Multi-Representation Real-Time Analysis/Synthesis", COST-G6 Conference on Digital Audio Effects (DAFx), Madrid, 2005.

Schnell, N., Borghesi, R. Schwarz, D., Bevilacqua, F., Müller, R. « FTM — Complex data structures for Max », International Computer Music Conference (ICMC), Barcelona, 2005

Wang, M.C., Haertel, G.D., & Walberg, H.J. (1994). What helps students learn? Educational Leadership, 51(4), 74-79.

Orlarey, Y. Fober, D., Letz, S.. Elody : a Java+MidiShare based Music Composition Environment Proceedings of the ICMC 97, Thessaloniki, 1997.

# 1st VARIAZIONI Workshop on Technologies for Content Enrichment in the Web2.0 era

*Organised by*

VARIAZIONI EC e-Content Plus Project
http://www.variazioniproject.org/

*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 

# Collaborative content enrichment in multilingual Europe, European Schoolnet approach on educational resources

Riina Vuorikari
European Schoolnet
61, Rue de Treves
1050 Brussels, Belgium
+32 2 790 75757
Riina.Vuorikari@eun.org

Frans Van Assche
European Schoolnet
61, Rue de Treves
1050 Brussels, Belgium
+32 2 790 75757
Frans.Van.Assche@eun.org

## ABSTRACT
In this paper, we describe the first steps of the content enrichment approach that European Schoolnet, with its partners in the MELT project, has taken to allow users collaboratively annotate the existing learning resources. The paper will briefly outline the MELT strategy for co-existence of structured and unstructured learning resources metadata (LOM), then it will outline the envisaged enrichment services for multilingual folksonomic approach. Moreover, we give the first evaluations on multilingual user-given free-keywords on learning resources which leads us to outline our approach to manage tags in multiple languages within the MELT federation of learning resource repositories.

## Categories and Subject Descriptors
H.3.3 Information Search and Retrieval: Search process, H.3.5 Online Information Services: Web-based services, H.5.2 User Interfaces: User-centred design: User-centred design.

## Keywords
Learning objects, learning repositories, federation of learning repositories, learning object metadata, annotation, user-given annotations, collaborative tagging, social bookmarks, multilingualism

## 1. INTRODUCTION

The use of social, collaborative classification systems has gone through a continuous growth in the latest years [1, 6]. An example of this is a multitude of sites that provide some type of social annotation of digital artefacts and a social navigation system (Flickr, del.icio.us, CiteULike, Last.fm, among others). Social tagging, i.e. allowing individuals to apply free text keywords to digital objects, potentially offers advantages in terms of personal knowledge management, serendipitous access to objects through tags, and enhanced possibilities to share content with emerging social networks.

This type of user-given content annotation is also seeing its emergence in the context of digital learning resources. There are already learning resources repositories that allows users-given annotation of their resources (e.g. KlasCement), and some allow users to create their own collections of learning resources (e.g. Merlot). Both of these are element of social, collaborative bookmarking and annotation, the type of content enrichment that this paper focuses on.

We first outline the MELT strategy for co-existence of structured and unstructured learning resources metadata (LOM). Section three describes the architecture and section four the envisaged enrichment services for multilingual folksonomic approach (e.g. tagging, rating, pedagogical annotations). Section five gives the first evaluations on multilingual user-given free-keywords on learning resources, which leads to section six that describes how tags in multiple languages are managed. Finally we will conclude with the future work.

## 2. METADATA ECOLOGY FOR LEARNING TECHNOLOGIES

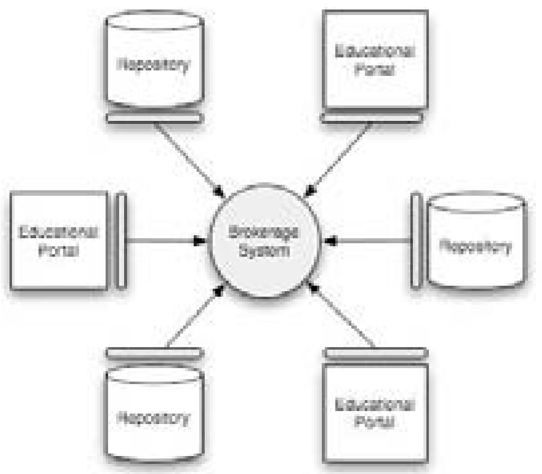
Since 1999 European Schoolnet has worked towards a goal of facilitating multi-lingual access to learning resources repositories and their catalogues of content throughout its network of national and regional educational stakeholders [4]. European education, especially that of K-12, being inherently multi-lingual and multi-cultural challenges traditional information retrieval methods based on simple metadata in one language. Approaches allowing end-users to use their native languages to search and receive resources from throughout Europe is a goal that promotes cross-national and cross-lingual use of digital educational content. Controlled vocabularies, such as multilingual LRE Thesaurus and LRE LOM Application Profile, have been developed by European-wide community of experts to overcome some hurdles of semantic interoperability.

However, the gap between the terms used by experts and practitioners in the field has proved problematic. In [7] the evaluators found that one third of teachers held negative views regarding the relevance of the search terms based on controlled multi-lingual vocabularies. That has lead our current interest to look into co-existence of taxonomies with end-user given keywords, i.e. tags [http://www.melt-project.eu] that could facilitate the discovery and access of resources that reside in different repositories throughout a federation of repositories.

In this section we gave reasoning for semantic interoperability when accessing resources on a scale of multi-lingual Europe. In the following we shortly describe the federated architecture.

## 3. FEDERATED ARCHITECTURE FOR EDUCATIONL CONTENT ENRICHMENT

The MELT federation of repositories builds on earlier work of the CELEBRATE project [10]. The backbone of the federation is a brokerage system, to which repositories of learning resources and educational portals connect using a client java library (depicted as a grey bar) that encapsulates the different networking protocols behind standard application programming interfaces (APIs).




The MELT approach extends the tagging practice by allowing social tagging on learning resources coming from the entire federation and not just a single repository. Moreover users can discover learning resources using tag clouds that span learning resources from the entire federation.

In addition however, a federation of learning resource repositories in a multilingual context needs to support multiple languages at the system level in order to support each repository and its national user-base, but at the same time, there is a need to allow people (i.e. user information and preferences), resources and tags to “travel” across national and linguistic borders.

## 4. MULTILINGUAL FOLKSONOMIC ENRICHMENT

In this section we describe European Schoolnet's collaborative content enrichment approach that compliments the expert indexers who use the LRE Application Profile and its agreed taxonomies to catalogue and describe educational content. This content enrichment approach is based on user interacting with the LRE portal, which acts as a gateway to the federation of repositories. It offers a service on top of the federation, as explained previously. From the user perspective, it comprises of three main parts; social bookmarking and users given keywords (tagging), ratings of usefulness and pedagogical annotations. Lastly, we also discuss the levels of user engagement when interacting with the system.

### 4.1 Social bookmarking and tagging

Social tagging and bookmarking commonly refer to a Web-based service to share a pointer to a digital item that people want to keep track of and go back to. Similarly, when users find learning resources (item) of interest through the LRE portal, they have a possibility to add this item into their list of bookmarks (hereafter referred as favourites) to access it at ease at any later stage. In the other words, users are able to create their own sub-collections of resources, that are also sharable with other users.

A user can add an item to favourites when they are viewing items in the search result list (hereafter SRL) or when they are viewing an interesting item in someone else's favourites area. At the same time, the user can associate as many free keywords with the item as they wish (hereafter called tags), as well as a comment, which can be private or public.

Users can manage (view/edit/delete) their own tags in favourites. The tags are made available to other users in diverse ways, e.g. in the SRL in relation to other metadata about the item, in a tag cloud, which is a representation of all popular tags in the system, or tags can also be used as a search term.

The ideas of tags is that they can be helpful for the user when she wants to get back to her interesting items and remember, for example, the content easily and how she intended to use it with pupils.

Users are invited to use their preferred languages for these tags. It is expected that users can use multiple languages while adding tags to resources. On the system level, we intend to work towards obtaining and validating a reliable translation for tags. It is intended that tags can be translated using a custom created multilingual dictionary, an automated validation will be required by a user before making this translation public.

### 4.2 Rating

Users have a possibility to manifest their subjective relevance judgements regarding resource by rating its usefulness. Users can rate items that they find in the SRL and in their own favourites area. On the other hand, these ratings are made available to other users when they view items in the SRL.

The users are asked to rate the usefulness of the resource on a scale from 1 to 5 (of no use to very useful), as well as to estimate the suitable age range of users for this item. In this rating form the users are also invited to comment on their judgement. The average of these subjective ratings is made available to other users in the SRL when they review items. By clicking on the rating, users are able to see all the individual ratings and textual comments. As for the individual ratings, users manage (view/edit/delete) their own ratings in the favourites area.

### 4.3 Pedagogical annotations

In the favourites area users are invited to add/edit/delete pedagogical annotations to items that they have selected. These are related to how users intent to use the item or have already used it within their teaching. A pedagogical annotation is considered as a metadata element that can be repeatable.

The pedagogical annotation has a number of sub-elements. Firstly, users are asked to use the 8LEM vocabulary [5] to classify the learning event for which the item will be used for/was used. Secondly, there is a description of the learning event organisation with the item. Moreover, users are asked whether any modifications were required and learners' reactions regarding the resource (vocabularies). Also, users can indicate whether they are available for further questions from other users regarding this annotation.

These pedagogical annotations are made available to other users in the SRL with other metadata, as well as in the pedagogical section of the LRE portal. It is also envisaged that other users can annotate these (e.g. “worked for me”).

### 4.4 Levels of user engagement

Apart from the three described content enrichment services, user interactions with the portal will be recorded on the server side logging. We are interested in better understanding on how users interact with the portal, with other users and with the content found. It is intended that this observation can lead to the development of a scale of user-engagement similar to Yahoo!'s STAR [8]. These can later be used to design a recommender system or an algorithm to be applied for LOR.

## 5. FIRST EVALUATIONS OF MULTILINGUAL USER-GIVEN TAGS

Several studies have been undertaken to better understand the behaviour and evolution of social tagging systems. A prevailing aspect among current studies concerning tagging is that they assume that tags are represented in a common language [2], understandable by all the members of the user community.

An early evaluation on a multi-lingual educational community is presented in [11] who used a tagging system across country and language borders. The data for this analysis is from a period of about three months (January 24 to April 21 2007). There were 77 teachers who made 459 bookmarks with 585 multilingual tags on 320 different learning resources (items). It was found that there was an average of 1.92 tags per item. One third of these tags were in Hungarian, another third in German and Polish and 26% in English, even though none of the users were native English speakers.

A further attempt was introduced to categorise the tags using three categories; factual, subjective and personal [9]. Interestingly, it was observed that about 13% of tags contain a general term, a name, place, e.g. EU, Euroopa, Europa, europe, geograafia, Pythagoras, etc. [11] furthermore hypothesises that this type of “travel well” tags, even if not translated, could be found useful for other users for their close similarity in spelling in many languages. These tags could be useful in scenarios where we are not sure about the preferred languages of the user or in cases where no tags in some given language exist yet.

The same paper also describes a focus group study on the usefulness and acceptance of multilingual tags in both languages languages that users are and are not familiar with. It was found that multilingual tags divide users, 50% found tags in multiple languages useful whereas the other 50% found them rather confusing. Moreover, the focus group found tags slightly less useful than keywords given by experts (4/10 top two most useful keywords were tags), which is an encouraging outlook for producing added value for users with no overlay from the part of the repository.

## 6. SYSTEM APPROACH TO MANAGE TAGS IN MULTIPLE LANGUAGES

The above described evaluations lead us to define an approach towards the better management of tags in multiple languages. This approach will have the following requirements. It a) allows a good recognition of the language of each tag, b) it leverages on the “travel well” tags and c) it allows automated validation of translated tags. All these have consequences for both when the user adds tags to the system and when the user uses tags for retrieval of learning resources, e.g. views a tag could, navigates tags and other users favourites and when used to enhance the match to the search query.

## 7. FUTURE WORK

However, there are challenges and research questions that need further attention. As it becomes clear that some tags are useful for some users, the design challenge becomes “hiding all but the right tags”. This implies for both entering and viewing the tags, e.g. what tags and in what languages to show/recommend to users when they are about to add a tag and what kind of tags to show for retrieval and social navigation.

Additionally the future work will consist of designing and validation of a system that supports translation of tags. As it is not yet entirely clear how end-users will perceive the value of translated tags, further attention through end-user studies are needed. Both, automated translation of tags (with the validation of the accuracy of this translation on an automated base) or crowd-sourcing the translation (e.g. user will do this) will be investigated to validate the most useful way to obtain a reliable level of multi-linguality when using end-user generated annotations. ..

## 8. ACKNOWLEDGMENTS

The authors which to thank Jim Ayre, Thomas Maier, David Massart, Sylvia Hartinger, Quentin Tremerie, and all members of the social tagging workgroup of the MELT and CALIBRATE projects.

# 9. REFERENCES


[1] Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), pp. 198—208. (2006).

[2] Guy, M., Tonkin, E.: Tidying up tags. D-Lib Magazine 12(1) (2006)

[3] Hammond, T., Hannay, T., Lund, B., Scott, J.: Social bookmarking tools (i). D-Lib Magazine 11(4) (2005)

[4] http://lre.eun.org

[5] Leclercq, D. & Poumay, M. (2005) The 8 Learning Events Model and its principles. Release 2005-1. LabSET.University of Liège, available at http://www.labset.net/media/prod/8LEM.pdf

[6] Marlow, C., Naaman, M., Boyd, D., Davis, M.: Position paper, tagging, taxonomy, flickr, article, toread. In: Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland. (2006).

[7] McCormick, R., Scrimshaw, P, Li, N., Carmel, C., Celebrate Evaluation, Deliverable 7.2, European Schoolnet. (2004)

[8] Ramakrishnan, R., Tomkins, A., Toward a PeopleWeb, Computer, vol. 40, no. 8, pp. 63-72, Aug., 2007

[9] Sen, S., Lam, S.K., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., Riedl, J.: tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. pp. 181–190 (2006)

[10] Van Assche F., Massart D., *Federation and Brokerage of Learning Objects and Their Metadata*, Proceedings of the I-CALT 2004

[11] Vuorikari, R., Ochoa, X., Duval, El Analysis of User Behavior on MultilingualTagging of Learning Resources. In Workshop proceedings of the EC-TEL conference: SIRTEL07 (EC-TEL '07) (Crete, Greece, September 17-20, 2007)

# Collaborative and Semantic Enrichment of Musical Libraries in the VARIAZIONI project


Carlos A. Iglesias
*Germinus XXI (Grupo Gesfor)*
*cif@germinus.com*

Francesco Spadoni
*Rigel Engineering*
*spadoni@rigel.it*

Jens Grivolla
*Music Technology Group*
*Univ. Pompeu Fabra*
*jgrivolla@iua.upf.edu*

Joachim Neumann
*Music Technology Group*
*Univ.t Pompeu Fabra*
*jneumann@iua.upf.edu*



## Abstract

*This article introduces the content enrichment techniques that have been explored and combined in the eContentPlus VARIAZIONI project[1]. The project has proposed a content metadata model for musical assets based on FRBR, which has been integrated with a standard content management system. Wizards for manual metadata feeding have been developed within the content management system. In addition, it supports social annotation, and includes automatic enrichment based on the sound properties of the contents and on smart clipping from available Internet resources, such as Wikipedia or Yahoo.*


## 1. Introduction

The web 2.0 phenomenon and its social approach is reaching new application domains, such as the enterprise systems (so called *Enterprise 2.0*) [La06,Cr06] and, more recently, is approaching the scope of Digital Libraries, so called *Library 2.0* [Ca06]. The main peculiarity of this approach is its user orientation. The Library 2.0 approach proposes that the user is not longer a pure passive entity that consults a catalogue. Instead, the user acquires greater protagonism, and contributes with his opinion about the items and an active role in their cataloguing.

One interesting notion of the Web 2.0 is "the Long Tail". In the context of the Library 2.0, while most bookshops or even libraries can only host the most demanded books, due to space and budget restrictions, the Internet has shown the value of the specialisation. Niche content providers and business can find a profitable business thanks to the attraction of minorities, which in Internet terms, constitute an attractive commercial target. In the same way, these specialised communities create social networks and can be even organised into movements such as the Open Source Movement.

This article presents the content enrichment techniques of the project Variazioni in order to apply the emerging Library 2.0 approach to the Musical Digital Libraries.

The rest of the article is organised as follows. Section 2 introduces the context of this research, the eContentPlus project Variazioni. Section 3 describes the content enrichment techniques that are being explored in the project. Section 4 describes the basics of the Variazioni Content Model and its user orientation, as well as the workflows of content enrichment that have been identified within the project. Section 5 introduces the role of social annotation in content enrichment. Section 6 and 7 describes briefly automatic content enrichment through sound analysis and smart Internet clipping, respectively. Finally, section 8 draws conclusions and presents future works.

## 2. The Variazioni project

The Variazioni Project is an eContentPlus Project funding as Content Enrichment Project with a duration of 30 months, starting on September 2007. The project is being coordinated by the musical private institution Fundación Albéniz and counts with several additional musical institutions (Lithuanian Academy of Music and Theatre, Koninklijk Conservatorium Brussels, Escolal Superior de Música e Artes do Espectáculo do


[1] This research has been co-funded by the European Community under the programme eContentPlus. The authors are solely responsible for this article and it does not represent the opinion of the European Community. The European Community is not responsible for any use that might be made of information contain within.

co-funded by the Community programme eContent*plus*

Porto, Sibelius Academy, and Association Europeenne of Conservatoires, Academies de Musique et Musikhochschulen) and technical partners (Germinus XXI, Rigel Engineering, Exitech, Universitat Pompeu Fabra and Università degli Studi di Firenze).
The purpose of Variazioni is to provide a Content Enrichment Portal where users and musical institutions can publish, annotate and access musical contents, including its protection. In order to validate its approach, the project will provide a minimum of 700 audiovisual hours, 1000 audio hours and 2000 written documents.

## 3. Content Enrichment Techniques

In the context of Variazioni project, content enrichment is defined as the process of adding new metadata to contents. In order to bring the Library 2.0 concept into the musical assets, the project has analyzed the requirements of users and musical content providers in order to publish and enrich their assets. Some of these requirements are (i) *easiness*, content should be easy to add and enrich; (ii) *specialization,* specific metadata should be defined in order to provide efficient retrieval and accurate cataloguing; (iii) *security*, since some of the contents are copyright protected securing mechanism should be provided in order to make them available online. Several content enrichment techniques have been identified: manual enrichment, automatic enrichment, social enrichment, and repurposing contents into different contexts.

*Manual enrichment* is the process of adding metadata by users according to a predefined metadata model. This is the traditional way of cataloguing items by librarians, with cataloguing standards and rules, such as AACR [AacrURL], MARC [MarcURL] or MODS [ModsURL],

*Automatic enrichment* adds new metadata based on the content characteristics (textual or audiovisual) or pre-existing metadata.

*Social enrichment* consists of adding annotations and comments by the users. The annotations given by the users constitute a free text taxonomy that is called folksonomy (also known as collaborative tagging, social classification, social indexing or social tagging). There are several important differences between this approach and the manual one. Firstly, this is a bottom-up cataloguing process, since there is no predefined taxonomy, but it emerges from the individual free annotations. In addition, social enrichment metadata is generated not only by experts, but also by content creators and consumers. Moreover, since the same item is annotated per different users, it is possible to improve the annotation of contents using the most popular annotations.

*Repurposing contents* consists of reusing a content in a different context, and adding metadata in this process. In the traditional libraries, items are classified in an aseptic way. Social tagging has brought contextual tagging, since users can add tags to a content depending of their current interests.

Variazioni combines these approaches taking in mind a user orientation approach, and are described below: manual enrichment (section 4), social enrichment (section 5), automatic enrichment, which is carried out in two different ways: based on audio properties of the contents (section 6) and on pre-existing metadata (section 7).; and repurposing, thanks to the usage of a standard content management system, where users can reuse contents for building new contents, such as articles, critics, etc.

## 4. Manual Content Enrichment

Variazioni proposes that manual content enrichment can be done in a collaborative way, and both content providers and content consumers can help in the cataloguing process.

### 4.1. Variazioni Content Metadata Model

Variazioni Content Model is based on FRBR (*Functional Requirements for Bibliographic Records*) [IF98]. FRBR is a conceptual entity relationship model developed by the International Federation of Library Associations and Institutions (IFLA), which has supposed a paradigm shift in cataloguing, since it considers the requirements of the users for searching, identifying, selecting and obtaining details of the bibliographic record.

The FRBR model is structured in three groups with different entities for each group. The first group comprises the core entities that represent the products of the artistic or intellectual model: work, expression, manifestation and digital item. A *work* is a distinct intellectual or artistic creation, for example a composition. An *expression* is the intellectual or artistic realization of a work, for example, the interpretation of a composition in a concert. A *manifestation* is the physical embodiment of a work, for example a CD production with the recording of the concert. A *digital item* is a single exemplar of a manifestation, for example, one CD bought at a shop with a serial number.

Some of the advantages of this model is that it is easy to establish relationships between different digital items. One can catalogue that different manifestations (i. e. video, book and audio recording) correspond to the same expression (i.e. musical event), or define that one digital item complements another one, for example. In addition, FRBR metadata has been defined taking into account the users' needs.

In order to apply this model in Variazioni, several assumptions have been taken. Firstly, since Variazioni is a digital library, the notion of digital item has not been included. Secondly, since our scope is a musical library, the notion of work has been mapped onto musical composition. Users only catalogue manifestations (called musical contents), but the model makes the difference between expression and manifestation. In this way, if for example, a user has catalogued a video of a concert, he/she can add an audio of the same concert, without having to introduce again all the metadata of the musical event. The reason of this is to simplify the way uses introduce the contents.

In order to adapt the general model to musical contents, several expressions (content types) have been defined with their associated metadata (concert, score, libretto, class) and several manifestations with their associated metadata (video, audio, paper and image). The user interface makes this differentiation transparent. In addition, metadata has been organized in aspects, in order to promote its reuse for different content types.

### 4.2. Content Enrichment Workflows

Variazioni has defined several workflows in order to assist in the enrichment process, which is described below.

The first workflow is *Content Enrichment Review*. When a user different from the content owner enriches a content, there is the possibility that the content owner do not agree with this contribution. It is needed some quality assurance mechanism in order to ensure metadata quality. Several strategies could be applied in this case. An approval of changes could be defined, and each content owner would validate the content contributions. Another alternative is notification of changes, the contributions are accepted automatically but the user receives notification of these changes. The latter is the one selected currently for Variazioni, taking into account mainly the overhead that this approval could mean for content providers.

The second workflow is *Content Protection*. When content owners select that the content should be protected, there is a workflow which starts the protection of the media file with Axmedis P2P Network and synchronizes all the metadata between the content management system and the Axmedis MPEG21 database. Once the media is protected, the media is available through the portal.

*Translation workflow*. It would be feasible to include a translation workshop, inside a musical institution, that assigns translation tasks to translators when a content is created. Nevertheless, translation is treated as a standard enrichment workflow, since there are no special requirements.

Finally, there are workflows for *automatic enrichment*. Depending on the nature of the content, a workflow can be started in order to produce a thumbnail of the content, or invoke automatic enrichment as described in section 6 and 7.

## 5. Social Annotation

VARIAZIONI adopts and refines emerging collaborative practices for content enrichment based on web2.0 concepts, leveraging the use of folksonomies , focusing on user participation, and exploiting the architecture of the web as a platform.

In particular, it developed tools and facilities for exploiting socially derived taxonomies (i.e. folksonomies). This classification schema has proven to be very accurate when communities tag the same resources, and in addition it provides feedback for improving quality of tagged resources.

Such tools enable the participation of user communities in the classification of existing content they are interested in, as well as content they create/integrate. Hence, the tools developed support the creation of user generated tags (in form of folksonomies) which can complement and enrich the existing metadata. The tools consist of a rich multifaceted user interface (UI) adapted to the specific VARIAZIONI content and metadata, wizards for tags selection and insertion, powerful functions for simultaneous tagging operations on multiple content objects, support for quality assurance mechanisms (support for evaluation, revisions and modifications of tags).

VARIAZIONI tools allow users to tag content objects with a descriptive word, expressing a characteristic of the content or associated meaning. They represent folksonomies as a tag cloud, which displays the most popular tags;

When a user creates a tag for a specific content object, the tag is stored in the database and associated with that object (by its ID). The system keeps track of all the tags that users have entered and the number of times that they have entered the same tag.

Each tag is visualized with a font size based on the popularity of that tag. This allows users to browse the content by way of a user-driven categorization of that content. When a user clicks on one of the tags in the tag cloud, the application retrieves a list of the associated contents. searching by tags (folksonomies) or by metadata.

Like most advanced digital users, VARIAZIONI users are increasingly interested in accessing all the aspects of a digital content, like user-generated video, photos, podcasts, music, games and more. They want access to all available data, all in real-time. VARIAZIONI is leveraging the users themselves to help organize the content and make it accessible and searcheable, using the "wisdom of the crowds". As VARIAZIONI CEP support the use of tags, it is able to assemble collections of social media based on the interests of VARIAZIONI users.

Significant effort went in making the user interface simpler, clean and more intuitive, doing user testing, performing validation sessions and listening to VARIAZIONI users, collecting and prioritizing what they wanted, liked and disliked.

## 6. Audio-based Content Enrichment

In order to automatically describe musical content by analyzing audio material, software has been developed and customized for the Variazioni project. In this software, music is described according to different meaningful facets:

1. Timbre, related to its instrumentation.
2. Tonality, related to its harmony and melody.
3. Rhythm and structure, related to the temporal location of events.
4. Dynamics, related to loudness and expressivity.

The result of this automatic tagging is integrated into the metadata description scheme defined in WP2, more specifically in D2.2 VARIAZIONI Musical Metadata Definition.

For this task, existing technology related to audio description of popular music has been adapted to the user profile of music professional and music lovers and specialized in classical music and genres included in the VARIAZIONI collection, which are mainly classical and folk music.

The software is based on an analysis of different levels of abstraction:

- Low-level features are closely related to the audio signal.
- Mid-level features use statistics and machine learning to create descriptors that are semantically meaningful.
- High-level descriptors provide relevant meaning to human users. They often require a modeling process.

## 7. Smart Clipping Content Enrichment

In order to supply some metadata without user intervention, existing sources of information are used to produce an initial set of descriptive tags for the given content.

In particular, Wikipedia serves as a basis for selecting an initial set of potential labels, taken from the Wikipedia page for an artist/composer or a specific work. A special algorithm is used to robustly match the available metadata (composer name, title of the work) to the pages contained in Wikipedia, allowing for variations in spelling, etc.

The candidate labels are then weighted, ranked, and filtered based on a measure of mutual information with the work or composer in question. The Yahoo Search web services are used to obtain the necessary statistics about the co-occurrence of different terms on the web, which we take to be representative of topical correlation.

The enrichment component also leverages existing folksonomies by connecting the content added to the Variazioni library to annotations provided by users in music related communities on the web. These annotations are filtered to reduce the noise typically found in uncontrolled open sources of user supplied information, and then included in Variazioni.

The use of tags extracted purely from web searches was also investigated (e.g. using the Yahoo term extraction APIs), but has proven to not be feasible with sufficient accuracy.

## 8. Conclusions and Future Work

This paper has shown different strategies for content enrichment in the project VARIAZIONI. Currently these strategies are being testing with real users. A real time monitoring system has been developed in order to adapt these strategies to users' interests and improve its effectiveness. In particular, next phases of the project will explore the ability to create user communities for attracting users to the system and improving the metadata of digital items according to their interests and peculiarities. The usage of web2.0 strategies for content enrichment can help to provide accurate cataloguing with a sustainable financing model.

## 9. References

[AxmURL] Axmedis Project Official Web Site, available at http://www.axmedis.org.
[AacrURL] Anglo-American Cataloguing Rules (AACR) official web site, available at http://www.aacr2.org/.
[Ca06] Casey, Michael and Savastinuk, Laura. Library 2.0. Service for the next-generation library. Library Journal, January 2006

[Cr06] Crawford, Walt. "Library 2.0 and Library 2.0". Cites & Insight: Crawford at Large, 6, number 2, 2006. Available at: http://citesandinsights.info/v6i2a.htm

[IF98] International Federation of Library Associations and Institutions, IFLA, Functional Requirements for Bibliographic Records, 1998. Available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf

[La06] Martin LaMonica, "Web 2.0 meets the enterprise", ZDNet News, April, 2006.

[MarcURL] MAchine Readable Cataloguing, developed by the Library of Congress, official web site available at http://www.loc.gov/marc/.

[ModsURL] Metadata Object Description Schema (MODS), official web site. Available at http://www.loc.gov/standards/mods/

[WiURL] Entry Folkonomy at Wikipedia, available at http://en.wikipedia.org/wiki/Folksonomy

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, "Article Title", *Journal*, Publisher, Location, Date, pp. 1-10.

[2] Jones, C.D., A.B. Smith, and E.F. Roberts, *Book Title*, Publisher, Location, Date.

# Enriching E-Learning Contents for Architecture in the MACE Project — Activities and Outlook


| Stefan Apelt | Christian R. Prause | Till Nagel | Martin Wolpers |
|---|---|---|---|
| Fraunhofer FIT | Fraunhofer FIT | FH Potsdam | KU Leuven |

Markus Eisenhauer
Fraunhofer FIT



## Abstract

*Today, learning object repositories for architectural contents are distributed over several European countries. Having been designed and implemented independently, sophisticated interchange of and access to learning objects is impossible. The valuable information contained therein is broken into pieces, time-consuming and expensive to obtain, and cannot be used to its full potential.*

*MACE overcomes these issues by building a metadata knowledge base on top of the various repositories. It links all the learning objects' metadata on a semantical level and provides premium search and browsing interfaces to its users. Learning objects can then be easily located and accessed in a uniform way; hence existing rifts between contents from different repositories are eliminated.*


## 1. Introduction

The goal of the MACE (*Metadata for Architectural Contents in Europe*) project[1] is to unify and enable access to huge amounts of architectural learning objects — which we will refer to as *contents* — scattered across heterogeneous and unaligned repositories throughout Europe. Typical architectural contents include such diverse matters as photographs and blue prints of buildings, texts about architects, questionnaires, local building codes, and material characteristics.

Content providing repositories like DYNAMO[2], ICONDA[3], ARIADNE[4] and WINDS[5] form the foundation of MACE. They are the outcome of former projects. Further repositories can be included as well, irrespective of their educational, professional or commercial background. MACE aims at being an open and flexible infrastructure.

Our approach to consolidate architectural contents in MACE implies only minimal modification costs for the affected repositories themselves. Instead we are building an infrastructure for a metadata knowledge base on top of existing repositories that will provide uniform access to contents. The basis for this knowledge base will be various kinds of metadata which we will attach to contents. The only preconditions that a repository has to fulfill are a standardised way for accessing its contents and standard protocol implementation for harvesting existing metadata.

Additionally, we will use information from non-architectural repositories for enhancing our metadata. Such repositories include geo-information systems like Geonames[6] for determining influence of surrounding landscape and culture on architecture, natural hazard databases for revealing similarities between buildings of similar risk groups, DBpedia[7] for additional information on buildings, and others, which will provide an information gain for the end user and allow us to better link the original repositories.

Besides utilising and combining existing contents and metadata we offer possibilities to enhance this knowledge base according to architectural needs. Thus, we create diverse MACE tools and user interfaces which support centralised enriching in an automatic, semi-automatic, or manual way.

## 2. MACE Approach

The MACE project is structured in three project cycles focussing on different activities for reaching the overall project objectives. The user-centred development process in MACE has followed the principles of the standard ISO 13407 (see [7]). During the first cycle, the user-oriented

[1] `http://www.mace-project.eu`
[2] `http://dynamo.asro.kuleuven.be`
[3] `http://www.irbdirekt.de/iconda/`
[4] `http://www.ariadne-eu.org`
[5] `http://winds.fit.fraunhofer.de`
[6] `http://geonames.org/`
[7] `http://dbpedia.org/`

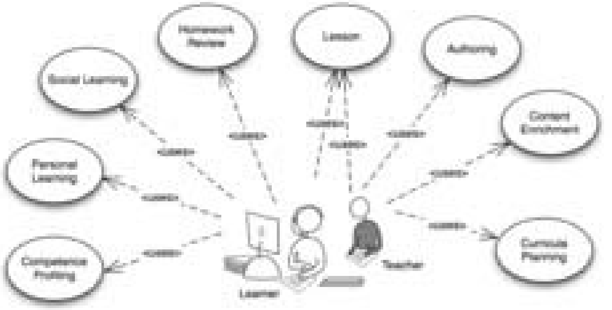


**Figure 1. MACE use cases with learning and teaching activities**

process has contributed to the specification of first prototypes, which will become available for evaluation during the end of the year. This is an ongoing and iterative process of user involvement, requirements analysis, and user evaluation. The standard does not prescribe specific methods to achieve these goals; they are to be chosen according to current state of the art and what is appropriate under individual project circumstances. Based on practical experiences from other projects, we have devised a scenario-based approach, combined with user interviews and expert analysis. We started by identifying user groups and community profiles and proceeding with a description of best practices in scenarios and use cases, determining principal activities and information needs.

Scenarios are a step-by-step description of typical user behaviour, of the interaction between the system and the user, and of the required knowledge processing. They describe main features of the application domain and thus are the basis for development of MACE system specification guidelines. As MACE focusses on educational practices, project stakeholders are either learners (students/professionals) or teachers. Use cases involving them describe the application domain through several types of learning and teaching activities with the stakeholders (see figure 1).

## 3. Metadata

We use four types: *Content and domain* metadata, *context* metadata, *competence and process* metadata, and *usage related and social* metadata; for further information see [9].

MACE has developed an application profile (*MACE-AP*) in order to harmonise the metadata descriptions and unlock the repositories. The application profile is based on the Learning Object Metadata standard (LOM) [6] with adaptations and extensions based on relevant classifications for architectural contents and learning objects. The content and domain metadata schema is based on the analysis of learning object providers.

In the following we explain how these are used to allow full-fledged access to the learning objects.

### 3.1 Content and Domain Metadata

Each learning object in any of the MACE repositories has a set of metadata already attached to it. This metadata, however, often follows a proprietary information schema. Therefore we identified abstractions, which are either common to all information schemas or which are worthwhile to implement for each repository. Repositories can then enrich their metadata to achieve compliance with the MACE-AP.

To extent our profile beyond the LOM standard, we use the *Classification* category in LOM and include additional attributes from architectural standards and classification systems. Through these extensions, specific taxonomy values can be added to contents not only from each repository, but from experts using the MACE Enrichment tool.

### 3.2 Context Metadata

*Context information* characterises the situation of a person, place or object, and is relevant to the interaction between an user and a computer.[2] Context offers a great opportunity to make contents better and more easily accessible, because it allows to create relations between unlinked digital contents that arise against a similar contextual background. In MACE, we take this one step further and connect objects and contents that, at first glance, have little in common. We distinguish between three kinds of *MACE entities* with essentially different natures and their own contexts: *Real world objects* with a relation to architecture, *Users* and *Digital contents* describing either of them.

It is important to note that the *state* part of contextual information is changing over time. Therefore we do not attach contextual information to objects like it is done with content and domain metadata. Instead we store relations between MACE entities, which in turn have attributes themselves.[14] Consider this example: “A user visits a building” — the relation would be “visits” with the start node being the user and the end node being the building. The relation would have a name (“visits”) and a “time” attribute containing the date of the visit.

This idea of storing data at relations allows for a very flexible approach in connecting digital contents like learning objects with geo-information systems, historical databases and other, seemingly unrelated data.

### 3.3 Competence and Process Metadata

*Competence metadata* describes competencies needed to interpret a learning resource or gives qualities a certain person has obtained. On this note, competencies can be described in various ways.[1] In coordination with the TEN-Competence consortium, MACE will interpret competencies as all factors for an actor to perform in an ecological niche.

*Process metadata* is used to describe the learning processes. In architectural learning three kinds of design methods are most prominent: *problem based learning*, *case based instruction* and *discourse based learning*.

Competence metadata allows searching for content related to a specific competence. To construct a problem aimed at a group of learners with certain competencies a teacher locates suitable learning content by using the competence metadata. Moreover, process metadata makes the reuse of teaching constructs and existing learning objects possible. Teachers use existing instructional designs and fit them to their classes. Additionally, specific structures of learning content can be stored, exchanged and found for reuse in more than one learning design.

Learning processes can be modelled in reusable designs using the IMS Learning Design (IMS-LD) specification. For competencies we use a competence card metaphor, which is derived as an extension to competence definitions currently available. The elements of the competence card will be based on two of the competence standards available: the IMS RCDEO[5] and the HR XML[4] standard. The competencies on the competence cards will be taken from existing descriptions of architectural qualifications.

### 3.4 Usage Related and Social Metadata

*Usage related and social metadata* is obtained from the content providers as well as from the MACE tools. In the case of usage, metadata captured from front-end tools and widgets can be saved to complement the user profile. The usage information is unified according to the *Contextualised Attention Metadata schema* (CAMs). By correlating usage data from different sources we obtain new knowledge about the usage of learning objects[12] Once captured, CAM does not change. Instead, CAM represents a continuous stream of new instances.

## 4. Using MACE

Imagine a teacher who wants to prepare a course in architecture using MACE: She starts her preparations by searching contents for an author or some generic keyword, thus accessing the content and domain metadata. Soon she realises, that her search matches many contents, but also that there has been much interest in the same topic by other users. So she decides to primarily browse the hot topics, because she expects that these have a higher importance for her work. She can do so, because MACE provides her the required usage metadata. While browsing the contents, the teacher recognises that most of them are from the same climatic region. Having found enough contents for a specific architect, the teacher wants to know if the architect is representative for that region or culture and inspects other contents from the very same region; thus accessing the context metadata. Finally she notices that there are some gaps between the provided competences of early learning objects in her course and the requirements of later learning objects. Again MACE can help by providing learning objects that exactly fill this competence gap.

## 5. Technical Infrastructure

The main goal of the MACE infrastructure is to create a framework for integration of multiple sources, for content enrichment of different types of metadata, and for allowing improved content access via a wide range of user interfaces and visualizations. Thus we implemented a service oriented architecture to combine services and databases flexibly. We came up with a hybrid combination of harvesting metadata from content repositories and federating searches to existing metadata repositories. Note, only the metadata describing the learning objects are transferred. The learning objects themselves stay in the repository and thus in control of their owner without changing the respective IPR.

The metadata is harvested through interfaces at each content repository which implements the OAI-PHM Protocol.[8] In turn, the central metadata repository will also offer an OAI-PMH interface so that content providers can eventually retrieve metadata suitable for their learning objects.

Moreover, MACE aims to provide personalised services, such as personalised search taking into account the search terms, user's context and usage behaviour. Such advanced services require vast amounts of information about the user, her context and the available learning resources. Therefore this information is captured in a number of federated stores: the metadata store (describing the learning objects), the contextual metadata store (describing the context in which learning objects exist), the competency server (providing competency mappings) and the usage data store (which describes usage of learning objects).

The federated search queries metadata stored in the central store to find suitable learning objects, eventually taking competency, usage and contextual metadata into account. It is enabled through the Simple Query Interface [3], which allows the federation of queries and the agglomeration of query results over repository boundaries. SQI is query language neutral and thus can be combined with any query language. Within MACE, we use the ProLearn Query Language (PLQL[8]) to query harvested content and domain

[8] http://ariadne.cs.kuleuven.be/lomi/index.php/

metadata through SQI. This is also used in the GLOBE consortium to federate queries over the global network of learning repositories.[10] Through the GLOBE portal, all MACE repositories can be searched as well.

As CAM (see section 3.4) is a continuous stream of data, we use the lightweight RSS protocol [11] to harvest metadata from the providing repositories to the central metadata store. In case of content providers, we suggest using secure web services following the OASIS SAML specification[9] to ensure privacy and security of the exchanged data.

MACE creates an infrastructure providing services to unify and ease access to contents in different repositories. Specialised metadata services simplify usage and guarantee proper reading and writing different types of metadata. Using the Simple Publication Interface[10], specific data is submitted into its respective metadata repository directly from the MACE infrastructure. Furthermore, we also implemented services to provide user authentication and authorisation, event logging and aggregation of raw information and reasoning on top of it. Hence, every service can utilise and build upon these common methods. The services are accessible via Simple Object Access Protocol interfaces [13], thus enabling tools, and applications as well as other services to employ existing operations. By combining multiple services, MACE can provide new ways to retrieve and enrich learning object metadata.

## 6. Conclusion

So far, the MACE project has been very successful in its first year. We have created a metadata taxonomy (the MACE-AP) for the domain of architecture which incorporates many other standards and taxonomies. Moreover, we implemented a technical infrastructure for enriching digital contents in a comfortable manner, while keeping the metadata consistent by validating against said taxonomy. We also started to use this technical solution to enrich contents and first evaluation results look promising.

In the near future, we will concentrate on enriching a larger piece of our core contents and evaluate our infrastructure. We will continue development of metadata type-specific solutions and tools, and start evaluating them.

In parallel to that, we will focus on opening up to the public, making our tools available to non-expert users and increase the load on our infrastructure. We will also update our metadata taxonomy to reflect shifting current state and what users deem helpful.

QueryLanguages_v1.0

[9] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security

[10] http://ariadne.cs.kuleuven.be/lomi/index.php/SimplePublishingInterface

## Acknowledgment

The findings presented in this paper are based on the hard work of the whole MACE consortium. The authors would like to thank the European Commission, for partially supporting this research with the project number: ECP-2005-EDU-038098.

## References

[1] G. Cheetham and G. E. Chivers. *Professions, Competence And Informal Learning*. Edward Elgar, 2005.

[2] A. K. Dey. Understanding and using context. *Personal Ubiquitous Computing*, 5(1):4–7, 2001.

[3] European Committee for Standardization (CEN). Workshop agreement on a simple query interface specification for learning repositories, 2005.

[4] HR XML Consortium. Competencies (measurable characteristics). online, August 2004. URL: http://ns.hr-xml.org/2_3/HR-XML-2_3/CPO/Competencies.html.

[5] IMS Global Learning Consortium. IMS reusable definition of competency or educational objective specification (RCDEO). online, October 2002. URL: http://www.imsglobal.org/competencies/.

[6] Institute of Electrical and Electronics Engineers Learning Technology Standards Committee. IEEE 1484.12.1 standard for learning object metadata (draft), 2002.

[7] International Organization for Standardization. ISO 13407: Human-centred design processes for interactive systems, 1999.

[8] Open Archives Initiative. Protocol for metadata harvesting 2.0. online, June 2002. URL: http://www.openarchives.org/OAI/openarchivesprotocol.htm.

[9] C. R. Prause, S. Ternier, T. de Jong, S. Apelt, M. Scholten, M. Eisenhauer, B. Vandeputte, M. Specht, and E. Duval. Unifying learning object repositories in MACE. 2007.

[10] S. Ternier, D. Olmedilla, and E. Duval. Peer-to-Peer versus Federated Search: towards more Interoperable Learning Object Repositories. In *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications*, pages 1421–1428. AACE, AACE, 2005.

[11] UserLand Software. Rss 2.0 specification. online, August 2002. URL: http://www.rssboard.org/rss-2-0.

[12] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking Actual Usage: the Attention Metadata Approach. *International journal Educational Technology and Society*, 7(7):1176–3647, 2007.

[13] WorldWideWeb Consortium. Simple object access protocol (SOAP) 1.1, May 2000.

[14] A. Zimmermann and A. Lorenz. An operational definition of context. In *CONTEXT 07*, 2007.

# Applying Semantic Web Technologies for Enriching Master Classes

Paloma de Juan
*Depto. Ing. Sistemas Telemáticos*
*Univ.Politécnica de Madrid*
*paloko@gsi.dit.upm.es*

José C. González
*Depto. Ing. Sistemas Telemáticos*
*Univ. Politécnica de Madrid*
*jcg@gsi.dit.upm.es*

Carlos A. Iglesias
*Germinus XXI (Grupo Gesfor)*
*cif@germinus.com*

## Abstract

*This article presents how semantic web technologies have been applied for enriching existing contents within the SEMUSICI project. The SEMUSICI project has the goal of researching on how semantic web technologies can be applied to digital libraries, and how this can improve searchability and accessibility. The project takes the results from the eContent project HARMOS, which defined a musical taxonomy for cataloguing master classes, and proposes a methodology for evolving this taxonomy into an ontology, and migrating the contents accordingly.*

## 1. Introduction

Cataloguing standards, such as MODS [1], MARC [13] or Dublin Core [14] define metadata following a flat property value orientation, which provides textual search capabilities. In some contexts, such as the musical digital libraries, this approach is too narrow, since some of the metadata are entities themselves, such as Compositions of Composers. In the Harmos project (section 2) an object oriented taxonomy was defined, where some of the values, such as compositions, movements or composers were modeled as entity objects, and an advanced search system based on these properties was developed and is available at [17]. This article presents an evolution of this approach, where semantic technology is used for modeling the relationships of the domain model. The main advantage of this approach is its powerful retrieval and inferential capabilities.

The rest of the article is organized as follows. Section 2 and 3 give an overview of the projects Harmos and Semusici, respectively, which constitute the context of this research. Section 4 describes a generic methodology for transforming taxonomies into ontologies. This is the main contribution of SEMUSICI project for content enrichment. Finally, section 7 draws out the main conclusions of the article and the future work.

## 2. The HARMOS project

The European eContent HARMOS project [16] had the aim of providing access through Internet to videos of master classes from big maestros. HARMOS has produced a collection of audiovisual contents that belong to the musical heritage, where education was the principal focus and the project's main objective.

Harmos defined a pedagogical taxonomy [15] which aims to cover the whole spectrum of musical practice and teaching, focusing on pedagogical aspects. The potential semantic descriptors of this taxonomy where structured around three main concepts, the music, the musician and the musical expression, con more than 400 descriptors as detailed at [15] and more than 700 audiovisual hours of recorded master classes have been catalogued according to this taxonomy.

## 3. The SEMUSICI Project

The SEMUSICI project [18] aims to evolve the results of the Harmos project by introducing semantic web technologies. The Harmos system provides several retrieval facilities, which allow finding a master based on the previous selections of the user, such as a composer, a composition, a movement, a teacher that has explained this composition, etc. as described in [15]. The introduction of new retrieval possibilities required to extend the database model and huge investment in development the new consults, which should be tuned and optimised given the big volume of the database. The usage of semantic web technology, which allows an easy extension of properties and relationships with new predicates, is expected to make this feasible. In addition, semantic web technology can contribute to improve the quality of metadata, since semantic web technologies can help in checking the consistency of the cataloguing.

The inclusion of semantic web technology points out several challenges. Firstly, it is needed to define an ontology that contains the concepts of the Harmos taxonomy. Secondly, since musical analysts should not be aware of the usage of semantic web technology for cataloguing, it is needed to develop easy interfaces in order to catalogue semantically. Thirdly, it is needed migrating all the Harmos catalogued multimedia collection to the new semantic schema. Finally, it is needed to evaluate the current status of semantic web technology in terms of throughput and performance, given the size of the multimedia collection. This article will cover the first objective.

## 4. Methodology for transforming taxonomies into ontologies

The central aim of Semusici is to provide a semantic structure that fits the former concepts taxonomy. The purpose of this new approach is to gather information about the relationships between disjoint leaves and build a new representation of both the concepts and these relationships. This leads to a richer representation of the knowledge that is really associated to the digital audiovisual items. This implies a deep understanding of the subject domain. There is no definite methodology for this task, but a general process together with best practices has been proposed.

Once the problem has been well defined and all the requirements have been identified, a suitable structure has to be chosen. As we are looking to take advantage of the technologies of the semantic web, the most appropriate structure is an ontology. The reason why we have chosen this structure is that it provides a formal way to represent roles and their corresponding relations in a specific domain. By placing a concept in such a structure, we are stating that it has certain properties and satisfies some restrictions about his meaning. In other words, each leaf of an ontology represents the definition of a certain resource.

The main difference between an ontology and a taxonomy is the kind of structure in which each of them is based. A taxonomy can be represented as a tree where each leaf is a class. No connections are allowed between disjoint branches. Relations between classes can only be established between a concept and its direct children. So an instance of a certain class can be defined as ”a kind of” its parent class. An ontology is a graph in which richer definitions can be expressed through a more extensive set of relations. This means that any class can be defined in terms of any other resource that is connected to it, not necessarily being its parent or child. Therefore ontologies can store more semantic information than taxonomies, allowing us to infer undeclared knowledge by studying the relations and restrictions of a certain class.

### 5.1. First step: choosing the appropriate tools

There is a wide variety of tools available to create, edit, browse and store ontologies. There are also many inference engines or reasoners, which are very important to obtain knowledge from the ontology. Several tools have been examined in order to choose the most suitable framework for our purposes. Some of these were Protégé [2], RacerPro [3], Sesame [4], SWOOP [5], WebODE [6], etc. A survey was carried out in order to find distinctive features. Therefore eleven parameters were chosen and thirteen tools were evaluated according to these key features. Some of these parameters were the supported languages, consistency check support, availability, maintenance, etc. As a result, Protégé and Sesame were chosen.

All these tools support a number of languages. Choosing the right language to implement an ontology is probably the most important step in the process. This depends on how thorough the ontology is intended to be. For Semusici, our initial choice was RDFS as it is the main language in Sesame. It proved to be complete enough to allow the building of a basic version of the ontology. Later we decided to include some restrictions to enforce the definition of the elements that we have already defined. These restrictions were also intended to help us perform consistency checks when adding new contents. For that purpose, new OWL statements were added.

### 5.2. Semusici knowledge base

There are two distinct parts in the knowledge base that is to be represented by the ontology. One is intended to capture all the information that is not directly related to the collection and can be useful to locate a recording. The aim of this is to answer any query that is not directly related to the contents of the recording itself. For instance, “give me all the recordings related to composers born in the 18$^{th}$ century”.

The other part of the knowledge base is the concepts taxonomy. The features of this structure have already been discussed. This taxonomy contains over 200 pedagogical concepts that are used as tags to describe the recordings. In the process of cataloguing the content, these recordings are to be labelled according to semantic descriptors that are part of this taxonomy.

The semantic descriptors were defined according to a tree diagram of concepts. This was based on three large branches that served as a starting point: the musician, music and musical expression. Each one of the divisions that structured the tree diagram of concepts was joined to one of these large branches. The smaller branches were then organized according to a series of categories, reaching, in the end, a didactic concept.

### 5.3. Building the ontology

This first ontology had to be built from scratch, as most of the concepts it should represent were new. Following a methodology is strongly recommended for this task. The goal of using a methodology is trying not to miss information in the process of transferring knowledge between the different actors that take part in the process. It also provides a set of steps to follow in order to avoid inconsistency, which would lead to undesirable rework. The quality of the ontology will be strongly affected by the choice of an appropriate methodology [7].

There is no single generic ontology-design methodology [8] that covers all the kinds of applications. This means that there is no standard way to build an ontology [9] neither a standard mechanism to evaluate a methodology. However, all published methodologies have proven to be useful, as they all have been applied to some process at least once. The key to finding the best guidelines for a certain application is to analyze the purpose those methodologies were used for and find similarities between that purpose and our application. This could be viewed as a way of reusing knowledge. Reuse is a very common practice in ontology engineering.

There are some steps that are common to almost every methodology. The first step is to identify the purpose and scope of the ontology. Both of them have already been mentioned in this document. Next, one must find out which questions is the ontology supposed to answer. These are called competency questions [10]. We gathered a list of over 50 questions and identified keywords that later would become part of the terminology of the ontology.

Next step was to decide which ones of these keywords should be represented as classes, attributes and instances. The most important thing to consider at this point is how specific we want our ontology to be. Thus we chose those concepts which we found they need a precise definition and separated them from those which constituted the most specific level of the ontology. We also considered reusing some published ontology but finally decided to define our own vocabulary.

### 5.4. From the concepts taxonomy to an ontology

The first step to turn this taxonomy into an ontology was to create a root class called *Concept*. Every instance of this class is assigned a concept name. This name is the same as the corresponding tag used to classify the digital recordings. Although the original taxonomy was divided into three main branches, we decided to create a first level of more specific classes. We intended to group concepts that had basic semantic features in common in order to make it easy to define relations between different classes.

The original classification grouped most of the concepts according to the instrument they were referred to. For instance, every technique that is related to a string instrument is placed in the subcategory Strings technique, child of Strings. We decided to create a main category, called Technique, to group all the specific technique related concepts, given that we can not consider that a technique "is-a" String. Thus we could establish that every instance of a subclass of Technique should be related to some type of instrument.

We followed these same criteria to create the main categories and build the first level of our ontology. We also defined some properties, such as "relatedTo", "partOf" and "elementOf". The first one was defined as a symmetric property and was meant to connect concepts that could be interesting to the same users. For instance, if a user searches for a lesson about hammers, he will probably be interested in videos about keyboards too.

Both "partOf" and "elementOf" are transitive properties. This means that if a first concept is part/element of a second one and this one is part/element of a third one, we can state that the first concept is also part/element of the last one. The difference between them is that if concept A is part of concept B, every instance of B has A (i.e. the frog is part of the bow, because every bow has a part called frog). However, if concept A is element of concept B, that means that only some instances of B have A (i.e. the reed is element of the embouchure, because there are wind instruments that have no reed in their embouchure). Considering this difference, we can state that if concept A is part of concept B and this concept is related to concept C, A is related to C. This is not true if A is element of B though.

Some of these semantic relations were established between concepts that stood under disjoint classes, in order to help the system make future recommendations. Finally, we also used restrictions to enforce the definition of the classes to make it easy to preserve the consistency of the ontology when expanded. Most of the important decisions were taken as a consequence of a thorough analysis of the distribution of the concepts. This analysis led to follow a bottom-up strategy, in order to find the most natural way of classifying the elements of the original taxonomy.

## 6. Conclusions and future work

As a result of the conceptualization of the subject domain, a list of classes and properties was elaborated. The formalization was carried out using Protégé. This tool provides all the means to code the ontology and visualize some of its elements. The resulting ontology was tested with Sesame. A set of custom rule was arranged in order to support some OWL reasoning.

Almost 1,500 statements were generated as a result of the codification. This is only the ontology, as the knowledge base has not been integrated yet. This includes more than 150 classes and almost 50 properties.

We are currently working on interlinking our ontology with some other data sources in order to improve searching. We would like to incorporate information from the CIA Factbook [11] to perform geographical reasoning. We would also like to add biographical information about the composers. We are testing some datasets from DBpedia [12].

Our second line of work is that of developing a consistency check system. Our purpose is to provide our ontology with a means to preserve consistency and coherence in case that there are several annotators working on the same dataset.

## 7. References

[1] MODS (Metadata Object Description Schema). Available at

[2] *The Protégé Ontology Editor and Knowledge Acquisition System*, http://protege.stanford.edu.

[3] *Racer Systems GmbH & Co. KG*, http://www.racer-systems.com.

[4] *Sesame: RDF Schema Querying and Storage*, http://www.openrdf.org.

[5] *SWOOP - Hypermedia-based OWL Ontology Browser and Editor*, http://www.mindswap.org/2004/SWOOP.

[6] *WebODE Ontology Engineering Platform*, http://webode.dia.fi.upm.es/WebODEWeb/index.html.

[7] S. Hakkarainen, D. Strasunskas, L. Hella, and S. Tuxen, "Choosing appropriate method guidelines for web-ontology building," in *Proceedings of the 24th Conference on Conceptual Modelling (ER 2005), ser. LNCS 3716*. Springer-Verlag, November 2005, pp. 270–287.

[8] N. F. Noy, and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology", *Tech. Rep. SMI-2001-0880*, Stanford University School of Medicine, 2001.

[9] M. Uschold and M. Grüninger, "Ontologies: principles, methods, and applications", *Knowledge Engineering Review*, vol. 11, no. 2, pp. 93–155, 1996.

[10] M. Grüninger and M. S. Fox, "Methodology for the design and evaluation of ontologies," in *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*, April 13, 1995.

[11] *CIA Factbook*, http://www4.wiwiss.fu-berlin.de/factbook

[12] *DBpedia. Querying Wikipedia like a Database*. http://wiki.dbpedia.org

[13] MARC Standards by the Library of Congress, available at http://www.loc.gov/marc/.

[14] Dublin Core Metadata Initiative, available at http://dublincore.org/

[15] C. Iglesias, M. Sánchez et al. A Multilingual Web based Educational System for Professional Musicians, M-ICTE 06, Current Developments in Assisted Education, 2006.

[16] Harmos project official web site available at http://www.harmosproject.com

[17] Magister Musicae web site, available at http://www.magistermusicae.com

[18] The SEMUSICI project official web site, available at http://semusici.germinus.com

# Enriching Pictorial Cultural Content with 3D Models

Manolis Lourakis†, Francesco Spadoni‡ and Piero Alcamo‡

† *Institute of Computer Science*
*Foundation for Research & Technology - Hellas*
*Vassilika Vouton, 70013 Heraklion – Greece*
*lourakis@ics.forth.gr*

‡ *Rigel Engineering S.r.l.*
*Via Spagna, 10*
*57010 Guasticce – Italy*
*{spadoni|alcamo}@rigel.li.it*

## Abstract

*This paper presents an approach for enriching culture-related pictorial content with 3D models obtained through semi-automatic reconstruction techniques. Starting with a single perspective picture, these techniques rely on a limited amount of interactive user input to recover a 3D textured graphical model corresponding to the depicted scene. Such 3D models constitute geometric scene descriptions and can serve as the digital content for interactive multimedia applications which improve the accessibility and visibility of cultural resources. Moreover, they can be reused in applications such as virtual reality, video games, 3D photography, digital visualization, visual metrology, art history study, etc.*

## 1. Introduction

Widespread deployment of information and telecommunications technologies has triggered an ever increasing demand for digital content to support a variety of applications. In particular, advances in three-dimensional (3D) model rendering and visualization have emphasized the need for 3D digital content to be employed in computer graphics, mixed reality and communication. This in turn, has created a tremendous potential for techniques capable of producing digital 3D models corresponding to scenes and objects, transforming the relevant research into a hot topic for several years [3]. Equally important to the production of digital content is the capture of the content's semantics with appropriate metadata, aiming to improve content reuse, personalization, searchability, interchange and management [6].

Being non-intrusive and cheap in terms of the required equipment, approaches that are based on the processing of images provide a particularly attractive paradigm for achieving 3D reconstruction and at the same time adding geometric semantics to images. RECOVER (full title "Photorealistic 3D Reconstruction of Perspective Paintings and Pictures")[1] is an EU-funded FP6 co-operative research project that focuses on the development of a system for the semi-automatic extraction of 3D graphical models corresponding to scenes depicted primarily in Renaissance perspective paintings but also in sketches, gravures, postcards and photographs.

To infer the 3D scene structure, single-view reconstruction (SVR) computer vision techniques are employed, aiming to "invert" the process of perspective image formation that lays down the geometric rules followed by artists when drawing. SVR is approached in an uncalibrated framework, in which there is no need for the camera pose or internal parameters to be known beforehand. To disambiguate among the infinitely many 3D reconstructions that are compatible with a given 2D image, simple geometric knowledge about the imaged scene should be available. This knowledge is supplied by a user based on his/her interpretation of the scene and concerns constraints such as coplanarity, parallelism, perpendicularity, etc. For this reason, single view reconstruction necessitates some manual intervention and concerns paintings rich in geometric regularity. The resulting 3D information is refined and enhanced with the aid of interactive editing tools, yielding a photorealistic 3D model of the depicted scene.

The remainder of the paper is organized as follows. Section 2 provides the motivation for pursuing this work, followed by a brief overview of the use of perspective in painting in section 3. Section 4 provides some background knowledge on linear perspective and section 5 offers a brief overview of the vision techniques for reconstructing a set of points and associated planes. Subsequent section 6 concerns the texturing of the recovered reconstructions and their storage in VRML format. Sample reconstruction results are presented in section 7, followed by a conclusion in section 8.

[1]RECOVER's website is at http://www.ics.forth.gr/recover/.

## 2. Motivation

According to the current state-of-practice, fully manual reconstruction techniques based on the use of CAD and 3D modeling tools for reconstructing paintings are quite tedious and labor-intensive, therefore time-consuming and costly. Laser scanning techniques cannot be applied due to the fact that the canvas used for painting is 2D. Conventional photogrammetric approaches and multi-view geometry vision techniques are also inapplicable due to their need for several images acquired from different viewpoints. Our approach, on the other hand, capitalizes on recent research results in order to bridge the gap between the research state-of-the-art and the state-of-practice in the construction of 3D models from 2D paintings.

Textured 3D models constitute a new and exciting way for perceiving and appreciating paintings. Their viewer can experience a feeling of immersion; paintings are no longer perceived as static artifacts from a long-gone past but as living, vibrant entities. With the aid of appropriate software, the viewer can literally dive into the painting, interacting with it and observing it from various viewpoints in impressive walk-throughs and inspiring fly-bys. This enables non-specialists to step into history and experience the scene in the space and time frame perceived by the artist. Ultimately, the viewing of paintings becomes a more appealing, exploratory endeavor, arousing the public's interest in fine art and cultural heritage in general.

It has already been mentioned that multimedia content such as images, is often annotated with some form of metadata that describe it. In the case of images, annotation typically refers to the task of describing their semantic content with a set of keywords or a caption [4]. Annotations of this sort are primarily used for image retrieval in large databases through keyword-based search. A 3D model reconstructed from an image can be considered as an alternative means of annotating the latter. Such an enriched image, accompanied by metadata in the form of a 3D model reconstructed from it and possibly additional graphical elements, can support several visualization types for the imaged scene. Furthermore, the 3D model can be reused in a wide spectrum of applications such as virtual reality, video games, 3D photography, visual metrology, etc.

## 3. Perspective in painting

Until the beginning of the 15$^{th}$ century, artists lacked the knowledge of creating an illusion of the third dimension in their works, which essentially look “flat” and fail to represent volume. Objects and characters were typically drawn depending on their importance rather than their distance from the observer. Such drawing practices were abolished during the Renaissance. The Italian painters of the time were the first to be interested in naturalism and studied the geometry of image formation in order to rationalize the representation of space by reproducing the perspective effects in the images of the world that they were creating. Giotto di Bondone was the first painter to treat a painting as a window into space, being concerned with the third dimension, the proportions and the natural appearance of surfaces. However, it was not until the writings of Florentine architects Filippo Brunelleschi and Leon Battista Alberti that linear perspective was formalized as an artistic technique aimed at creating a systematic illusion of space behind the canvas. The comprehension of the relations of perspective to perceptual aspects of depth and space, allowed painters to take advantage of the impressive ability of the human visual system to infer 3D properties of shape from a single 2D image. Hence, the use of perspective revolutionized the art of painting and raised it to a prestigious level among the fine arts. Renaissance masters such as Masaccio, della Francesca, da Vinci and Dürer pushed theory to a considerably sophisticated stage, paving the way for its complete mathematical formulation.

## 4. Elements of linear perspective

Intuitively, the basis of perspective image formation involves rays of light that travel from scene objects and through the imaging plane to a viewer's eye or a camera. A perspective image corresponds to the intersections of the light rays with the image plane and is formed by a pinhole camera, a device that performs central projection of points in space onto a plane [3].

One of the more striking features of perspective projection is that the images of infinite objects can have finite extends. For instance, an infinitely long scene line projects to an image line that terminates in a finite point. This point is known as the *vanishing point* and depends only on the 3D line's direction and not on its position. Thus, parallel 3D lines share the same vanishing points. In a similar manner, the vanishing points of sets of non-parallel, coplanar 3D lines lie on the same image line, which is known as the *vanishing line* of the underlying plane. The vanishing line of a ground plane is often referred to as the *horizon*. Parallel planes share the same vanishing line. After identifying the image projections of at least two parallel 3D lines, their corresponding vanishing point can be detected as their point of intersection. Knowledge of a length ratio defined by three collinear points forms the basis for an alternative scheme for vanishing point detection. The vanishing line of a plane can be detected from at least two vanishing points corresponding to different directions that are parallel to the plane in question. Alternatively, a vanishing line can be directly determined from the images of three parallel 3D lines with known ratios of distances among them [3].

A 3D plane that is viewed on a planar image under perspective projection induces a general plane-to-plane projective transformation that is known as a *homography*. Homographies also encode the transformation between different images of the same 3D plane. A particularly useful kind of planar homography is that referred to as a *metric rectification homography* [3]. Such a homography maps the image of a plane to another one so that it removes the effects of projective distortion (i.e., spatial foreshortening). A metric rectification homography allows metric properties of the imaged plane, such as angles, length and area ratios, to be directly measured from its perspective image. Furthermore, a metric homography is of utmost importance in texture mapping, since it allows the synthesis of a distortion-free texture map for a non-frontoparallel (i.e. slanted) plane. The most straightforward way to estimate a metric rectification homography is through identifying a scene rectangle with known aspect ratio (i.e. height over width ratio) and associating its four corners with those of the quadrangle corresponding to its image projection. Alternatively, a metric rectification homography can be estimated from the vanishing line of its underlying plane along with at least two constraints arising from combinations of line segments with known angles or length ratios [3].

## 5. Reconstruction of points and planes

In this work, objects are modeled using surface rather than volume representations. Thus, planar faces are reconstructed as opposed to polyhedral primitive solids such as prisms, parallelepipeds and pyramids. This is because solid primitives are often not fully visible in a single image due to factors such as occlusions and field of view limitations, therefore their reconstruction is not possible without considerable, arbitrary generalization. Reconstructed points are represented by their Euclidean coordinates while planes are represented with their normal vectors and distances from the origin.

Roughly, the workflow for obtaining a reconstruction from an image involves three steps. First, the image has to be calibrated in order to determine the optical properties of the device that acquired it, be it a camera or a painter's eye. Then, a preliminary reconstruction of a set of planes and associated points is recovered. Finally, this reconstruction is refined in order to adhere to user-supplied geometric constraints. More technical details concerning the approach can be found in [5] and references therein. Apart from the 3D geometry, the reconstruction permits the estimation of the viewpoint of the employed camera. In the case of a painting, this is equivalent to its *vantage point*, i.e. the location from which the observer experiences the liveliest three-dimensional illusion regarding the painted scene.

## 6. Texture mapping and VRML export

To increase the realism of a reconstructed 3D model, textures automatically extracted from the corresponding image are mapped on the model surface. These textures are thus photorealistic and are saved after being compensated for perspective distortion effects using their corresponding rectification homographies. On the one hand, this last choice renders easier the editing of extracted textures using ordinary image editing software and on the other hand, facilitates the generation of extended textures with the aid of texture synthesis or texture transfer algorithms: One of the main shortcomings of SVR is its inherent inability to cope with occlusions that result in "holes" in the reconstruction. To fill in the missing information, occlusion filling techniques can be employed. More specifically, non-parametric texture inpainting and synthesis algorithms are incorporated, which are capable of masking out certain image regions that correspond to unwanted objects or enlarging small patches by synthesizing stochastic textures based on their structural content [1, 2].

Recovered reconstructions are saved in the VRML (Virtual Reality Modeling Language) text-based scene description language, which is an open, ubiquitous standard for 3D graphics on the Web. In the context of SVR, VRML/X3D is very convenient for visualizing the reconstructed 3D models and importing them to a wide variety of 3D graphics software for further use. Another useful feature offered by VRML is the support for various sensors that monitor a viewer's actions and can trigger events. Such events can be used for loading web pages, displaying billboards and triggering animations which combined with the reconstructed virtual world, considerably augment a viewer's interaction with an image.

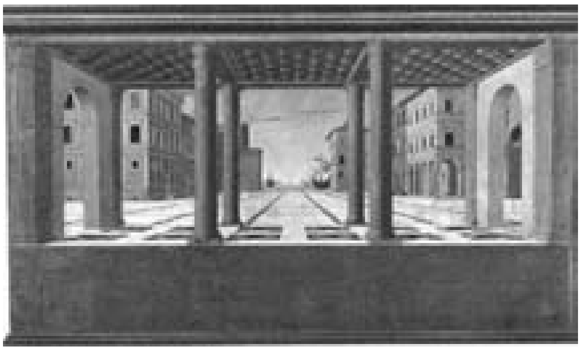


Figure 1: "Città Ideale", F. di Giorgio Martini, 1470s.

## 7. Sample results

Due to space limitations, only one reconstruction experiment that was carried out with the aid of the 674x400 image shown in Figure 1 is reported here. This image is a 15th century painting titled "Città Ideale" and illustrates a typical example of Renaissance architecture and urban planning. The painting was executed using *one point perspective*, under which the sides of buildings recede towards the vanishing point, while all vertical and horizontal lines are drawn face on. Camera calibration was based on the homography of a three by two rectangle formed by floor tiles. The sole finite vanishing point was estimated from the intersection of inwards oriented parallel lines provided by the user and, since the horizon is horizontal, sufficed to estimate the latter. The outlines of planes to be reconstructed were then interactively marked on the painting and plane parallelism/perpendicularity relationships were specified by the user. Following this, the reconstruction was carried out automatically, producing a textured VRML model two views of which are illustrated in Figure 2. Note that by exploiting regularity, it has been possible to synthesize the texture of floor areas that have been occluded by the pillars in the original painting.

## 8. Conclusion

This paper has presented an overview of our efforts to investigate cultural content enrichment through semi-automatic single view reconstruction from a perspective image. Reconstruction from a single image can trigger a new breed of presentation/communication methods for paintings, targeted to the education, entertainment and tourism industries. Additionally, flexible SVR techniques can find practical applications in several other domains [5].

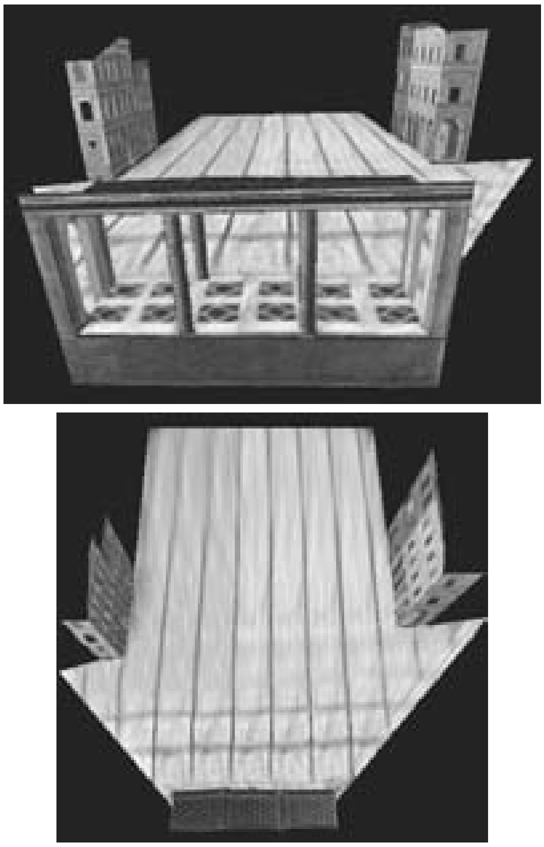


Figure 2: Side and top view of the reconstructed model.

## 9. Acknowledgements

The RECOVER project has been supported by the Commission of the EU through contract COOP-CT-2005-017405. M. Lourakis was also partially supported by the EU FP6-507752 NoE MUSCLE.

## 10. References

[1] A. Criminisi, P. Perez, and K. Toyama, "Region Filling and Object Removal by Exemplar-Based Image Inpainting", IEEE Trans. on IP 13(9): 1200-1212, 2004.
[2] A. Efros and W. Freeman, "Image Quilting for Texture Synthesis and Transfer", Proc. of SIGGRAPH '01, pp. 341-346. ACM Press, New York, NY, USA , 2001.
[3] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2003.
[4] M.S. Lew et al, "Content-based Multimedia Information Retrieval: State of the Art and Challenges", ACM Trans. Multimedia Comput. Commun. Appl., 2(1): 1-19, 2006.
[5] M. Lourakis, P. Alongi, D. Delouis, F. Lippi and F. Spadoni, "RECOVER: Photorealistic 3D Reconstruction of Perspective Paintings and Pictures", Proc. of EVA London 2007, London, UK , 2007.
[6] K. Nagao, "Digital Content Annotation and Transcoding", Artech House, 2003.

## Author Index




*AXMEDIS 2007: proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*
edited by Jaime Delgado, Kia Ng, Paolo Nesi, Pierfrancesco Bellini
ISBN 13: 978-88-8453-678-5 © Firenze university press, 2007