



D3.2 First working prototypes of the measuring systems

Project acronym	TELMi
Grant Agreement nr	688269
Project title	Technology Enhanced Learning of Musical Instrument Performance
Call identifier	H2020-ICT-2015
Topic	ICT-20-2015
Type of action	Research and Innovation Action
Start date of project (dur.)	February 1st, 2016 (36 months)
Report availability	Pu-Public
Document due Date	September 30th, 2016
Actual date of delivery	September 30th, 2016
Leader	UPF (Internal review UNIGE)
Reply to	Oscar Mayor Soto (oscar.mayor@upf.edu)
Additional main contributors (author's name / partner acr.)	Alfonso Pérez (UPF), Zacharias Vamvakousis (UPF), Rafael Ramirez (UPF), Gualtiero Volpe (UNIGE)
Document reference	TELMi-D-WP3-UPF-20160915-D3.2-First_working_prototypes_of_the_measuring_systems
Deliverable type	DEM
Document status	Final



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

Executive Summary

This deliverable presents a technical description of the early prototypes developed during the initial months of the project that will be used to measure violin performances and to perform some feature extraction, aiming at obtaining high level descriptors from a performance. Additionally, the data synchronization methodology is explained as well as how the data is formatted and stored in the public database.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

Table of contents

[2 Prototypes for the Audio Measuring Systems](#)

[2.1 Real-Time play along prototype: ViolinRT](#)

[2.1.1 Context and Use cases](#)

[2.1.2 Goals / Questions](#)

[2.1.3 Description](#)

[2.1.4 Technology used](#)

[2.1.5 Next steps](#)

[3 Prototypes for the Video Measuring Systems](#)

[3.1 Web-camera marker-based tracking](#)

[3.1.1 Context and Use cases](#)

[3.1.2 Goals / Questions](#)

[3.1.3 Description](#)

[3.1.4 Technology used](#)

[3.1.5 Next steps](#)

[4 Prototypes for the Motion Capture Measuring Systems](#)

[4.1 High-performance motion capture system: Qualisys Cameras](#)

[4.1.1 Context and Use cases](#)

[4.1.2 Goals / Questions](#)

[4.1.3 Description](#)

[4.1.4 Technology used](#)

[4.1.5 Next steps](#)

[4.2 Alternative high-performance motion capture system: Polhemus EMF](#)

[4.2.1 Context and Use cases](#)

[4.2.2 Goals / Questions](#)

[4.2.3 Description](#)

[Figure 5. Polhemus EMF systems being tested and demonstrated at first public event in Barcelona.](#)

[4.2.4 Technology used](#)

[4.2.5 Next steps](#)

[4.3 Low-cost motion capture system: Kinect](#)

[4.3.1 Context and Use cases](#)

[4.3.2 Goals / Questions](#)

[4.3.3 Description](#)

[4.3.4 Technology used](#)

[4.3.5 Next steps](#)

[5 Synchronization and Data Storage](#)

[5.1 EyesWeb](#)

[5.2 Repovizz](#)

[5.2.1 Description](#)

[5.2.2 Data formatting, storage and sharing](#)

[6 Conclusions](#)

[7 References](#)



1 Introduction

For the first version of the TELMI prototypes (first 8 months of the project), individual software pieces are used for each modality acquired from a performance. In later versions of the prototypes, they will be further integrated with the final aim of having a TELMI platform prototype where all measuring systems are interconnected.

We are currently acquiring the following modalities:

- Audio (multi-channel)
- Video (multiple cameras)
- Motion capture (high performance, low-cost)

At the end of the document a section is dedicated to the synchronization methodology used to align data from different measuring sensors as well as a section to explain how the acquired data streams are formatted and stored in the public database (repovizz).

These prototypes are early versions of the measuring systems. They will be further developed and presented in future deliverables in the second year of the project *D3.3 Second version of the data acquisition systems prototypes (M20)* and at the end of the project *D3.4 Final platforms of the data acquisition systems (M33)*. We are willing to use other modalities in the future like EMG or EEG that will be explained in those deliverables.

2 Prototypes for the Audio Measuring Systems

2.1 Real-Time play along prototype: ViolinRT

This prototype is used to record a user performance and to carry out audio analysis . It extracts audio feature descriptors and visualizes them in real-time. It also does some preliminary performance analysis to rate the performance of the user in terms of tuning and timing. The target users are violin players that are willing to play following a score. After playing it, they can review the performance and get feedback from the system to improve the performance. Although it only uses audio as input, it will be adapted to other modalities in future versions of the prototype. The prototype can be used to record the audio input (from the performance) together with the extracted descriptors and review it again at some point in the future.

2.1.1 Context and Use cases

Audio recording, audio analysis, real-time visualization, music learning, music practicing, score visualization, musicXML rendering, performance feedback, violin performance.

2.1.2 Goals / Questions

- Recording of audio from violin performance.
- Analysis of audio to extract feature descriptors.
- Performance rating from audio analysis.
- Educational tool for violin performers.
- Real-time and off-line visualization of performance features.



2.1.3 Description

The ViolinRT prototype aims at serving as an educational tool for musicians that are learning to play the violin. The prototype can import a musicXML file containing a musical score and renders it (see figure 1), so that the user can play while looking at the prototype screen to follow the score notes. Before starting to play, the user clicks the play or record button and a visual and audible metronome can be selected to follow the adjusted tempo. While playing or recording, the current score note at each time is highlighted so the user knows where she is. Also a reference synthetic sound can be triggered to hear the score notes before playing them, as a reference for beginners. The user can navigate through the score clicking in the desired bar or using the transport widget and buttons so that the practicing can be started from any part of the score, there is no need to repeat the whole piece over and over.

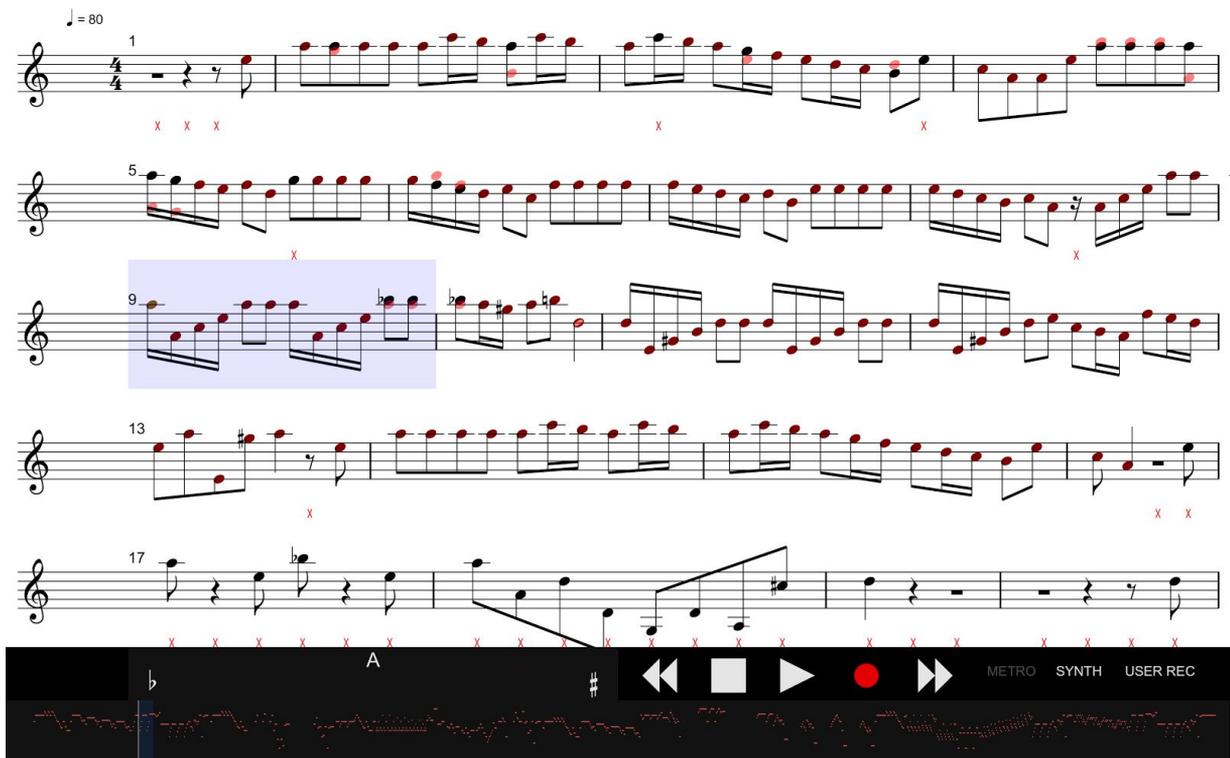


Figure 1: ViolinRT Score centred interface for musicians

While playing, the prototype analyzes the user's audio (violin sound) in real-time and extracts some audio features that are used by the *performance rating algorithm* to rate the performance and give indicators about whether the user has played the notes in tune and at the correct timing. A preliminary performance rating algorithm has been implemented to compare the pitch curve extracted from each played note with the reference notes in the musical score. For each note a linear regression of the pitch is performed in order to guess the played note, which is compared with the target note. An indication of mistakes and pitch deviations are displayed over the score in red (see figure 2). Also if a note in the score is not played, it is visually marked in the prototype gui as an x below the note. These indicators will be improved in terms of UX in future versions of the prototype.

The user can switch to other more advanced visualizations (less musical oriented and more scientific) where the extracted audio descriptors can be visualized giving further useful information about the performance, like pitch curve (note tuning) over the note (in green), dynamic curve (note intensity) (in red),



piano roll representation of the score notes, and performance rating probabilities (see figure 2). These probabilities are not very informative for the user, but they are very useful for us to refine the performance rating algorithms.

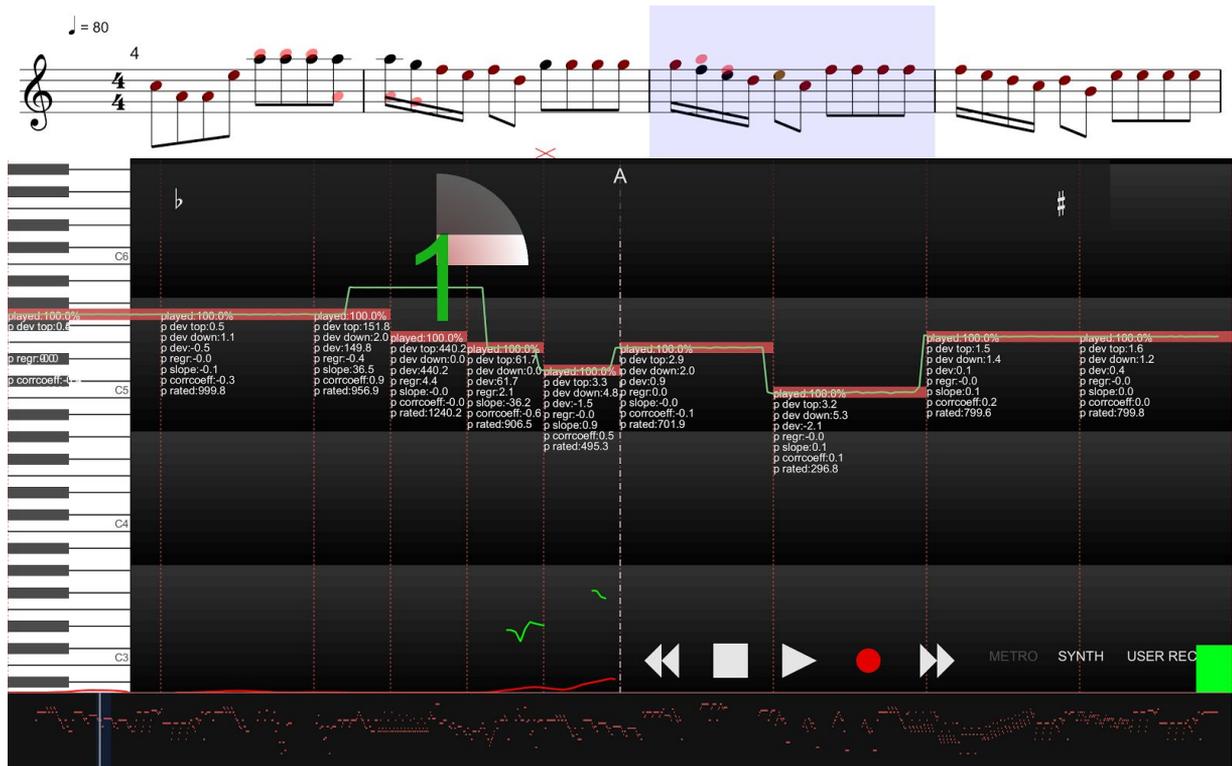


Figure 2: ViolinRT Audio centred interface for research

Also a tuning widget and a volume widget have been implemented that allow the user to see the instantaneous pitch and volume respectively.

Currently the user can control the prototype with a set of transport buttons (play/pause/stop/record) and also using the keyboard keys and mouse actions for some features. After a recording is performed, it can be reviewed immediately, and it can also be saved to the local storage together with the extracted descriptors to be loaded at any future time. In newer versions of the prototype, upload of performance recordings to the public remote database (repovizz) will be enabled, as well as download of reference recordings from the database to the prototype in order to be compared with the user performance.

A video of the current system can be found at:

<https://vimeo.com/176355821>

2.1.4 Technology used

- Microphone or pickup (to capture the violin sound)
- Audio feature extraction (UPF)
- Pitch analysis and tuner (UPF)
- Metronome (UPF)
- Performance rating (UPF)
- Score rendering from MusicXML (UPF)
- Real-time descriptors visualization widgets (UPF)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

- The software application has been developed in C++ using the GLFW OpenGL framework for the GUI.

2.1.5 Next steps

In future releases the prototype will allow to:

- Make a better rating.
- Practice loops.
- Upload performance recordings to the public database.
- Retrieve performance recordings from the public database.
- Compare a user recording with the reference expert performance.
- Visualize real-time 3D mocap data and time series from other sensors (received via OSC).

3 Prototypes for the Video Measuring Systems

3.1 Web-camera marker-based tracking

The lowest-cost system we developed is based on the RGB video captured by a laptop web camera. In this approach the ArUco library has been utilized for estimating a 3D pose of the violin. The ARUCO library [4] (Garrido-Jurado et al., 2014) provides real-time marker based 3D pose estimation using markers. For tracking the bow, two spherical colored markers are attached at the edges of the bow. The main advantage of this approach is its low cost, and the low usage of computational resources. The disadvantage is that the detection accuracy of the system is much lower than the one offered by more expensive alternatives.

3.1.1 Context and Use cases

The purpose of this prototype is to research the amount of information that can be extracted through low-cost camera and microphones, available on laptops and/or mobile devices.

3.1.2 Goals / Questions

In the current version, the prototype is able to perform the following operations:

- 3D pose estimation of an aruco marker attached to the violin.
- Audio recording of violin performance
- Estimation of basic bowing controls (bow inclination and speed)
- Visualization of 3D models of the markers

3.1.3 Description

The colored markers are made out of styrofoam in order to be light and do not add extra weight on the bow. As it is illustrated in figure 3, the markers are yellow spheres and are placed on both extremes of the bow, so they can be easily tracked from any point of view and lighting conditions. The ArUco fiducial marker is used for estimating the camera pose with respect to the marker. In other words, it allows us to detect the translation and orientation of the fiducial marker according to the point of view of the camera. Regarding our application setup, we employ a single fiducial marker that is placed at the bottom right of the top plate of the violin body, under the fingerboard where the neck starts. This position minimizes the possibility of sight overlapping by the performer's hand and bowing movements while recording the video.



Both video and audio acquisition are performed through a laptop's built in microphone and camera. In the current version of the system the pose estimation is visualized online. The speed and direction of the bow are computed offline. A demo of the prototype can be found at: <https://youtu.be/fmHFrymPjyg>



Figure 3. Styrofoam color markers for detecting the bow, in addition to the ArUco marker for estimating the 3D pose of the violin.

3.1.4 Technology used

The prototype currently runs on a macbook pro laptop, utilizing its built in microphone and camera. The system is developed on top of the OpenCV¹ software library. OpenCV is a computer vision and machine learning software library.

The ArUco library is used from the violin pose estimation. The ArUco library was developed on top of OpenCV and is considered as an integral part of the OpenCV package, hence inheriting its core functions and capabilities.

3.1.5 Next steps

Future improvements include:

- Create a 3D model of the violin (current version only performs 3D pose estimation of the marker)
- Compute more bowing control features, such as distance of the bow from the bridge. Utilizing machine learning and audio features, more features can be approximated, such as the bow force.

4 Prototypes for the Motion Capture Measuring Systems

4.1 High-performance motion capture system: Qualisys Cameras

A working prototype has been developed based on the Qualisys motion capture system presented in D3.1, and on the EyesWeb XML platform. The prototype can track, record, play back, and analyse the 3D positions of a set of reflective markers located on (i) the violin, (ii) the bow, and (iii) the body of the performer. The prototype also enables to define and track virtual markers. Starting from such raw data, it is possible to compute the bow control and the body movement features defined in Section 3 and in Section 4 of D4.1. Concerning the main advantages of this prototype, it provides highly accurate data and it does not need to put any hardware or any wire on the performer and on the violin, besides the markers.

¹ <http://opencv.org/>



It also does not require a very long calibration phase. The main disadvantages are that it needs accurate (and possibly time consuming) positioning of markers and post-processing of data. It also is rather expensive, so that its cost cannot be afforded by a single user nor by a conservatory, possibly. Its main use is in lab environment, e.g., for creating the reference archive of performances the TELMI project envisages. Mechanisms were developed to synchronize motion capture data with other multimodal streams (video, audio, IMUs, and possible biometric signals). Figure 4 shows the bow and violin position being computed in real-time during a violin performance.

4.1.1 Context and Use cases

This prototype allows for highly accurate tracking of 3D marker positions, enabling extraction, analysis, and visualization of bow control and body movement features. This enables recording data for the envisaged TELMI reference archive and characterizing performances with a collection of descriptors to be used as ground-truth to assess students' performances. The system is going to be used mainly in laboratory settings and environment.

4.1.2 Goals / Questions

In the current version the prototype is able to perform the following:

- Tracking 3D positions of reflecting markers.
- Recording the tracked positions.
- Playing back the recorded data.
- Computing a subset of bowing control and body movement features (D4.1).
- Visualizing the computed features.
- Synchronizing with audio, video, and sensors.

4.1.3 Description

The current prototype consists of a 13-cameras Qualysis motion capture system, synchronized with two professional JVC video cameras, one Kinect v.2 device, one pick-up on the violin and two microphone for capturing audio in the environment, and one or two myo sensors (www.myo.com) providing both acceleration and EMG data. In order to receive and record data from myo sensors and to keep them synchronized with motion capture, video, and audio streams, the EyesWeb XMI platform was extended to fully support the myo devices. The performer wears a set of markers and of rigidbodies composed of a fixed number markers (see D3.1 for more details), and one or two myo sensors on her arms. Tracking of the violin and of the bow is performed with both real and virtual markers. EyesWeb XMI applications are included in the prototype to extract features from the captured data (at the moment a subset of the bow control and body movement features listed in D4.1).



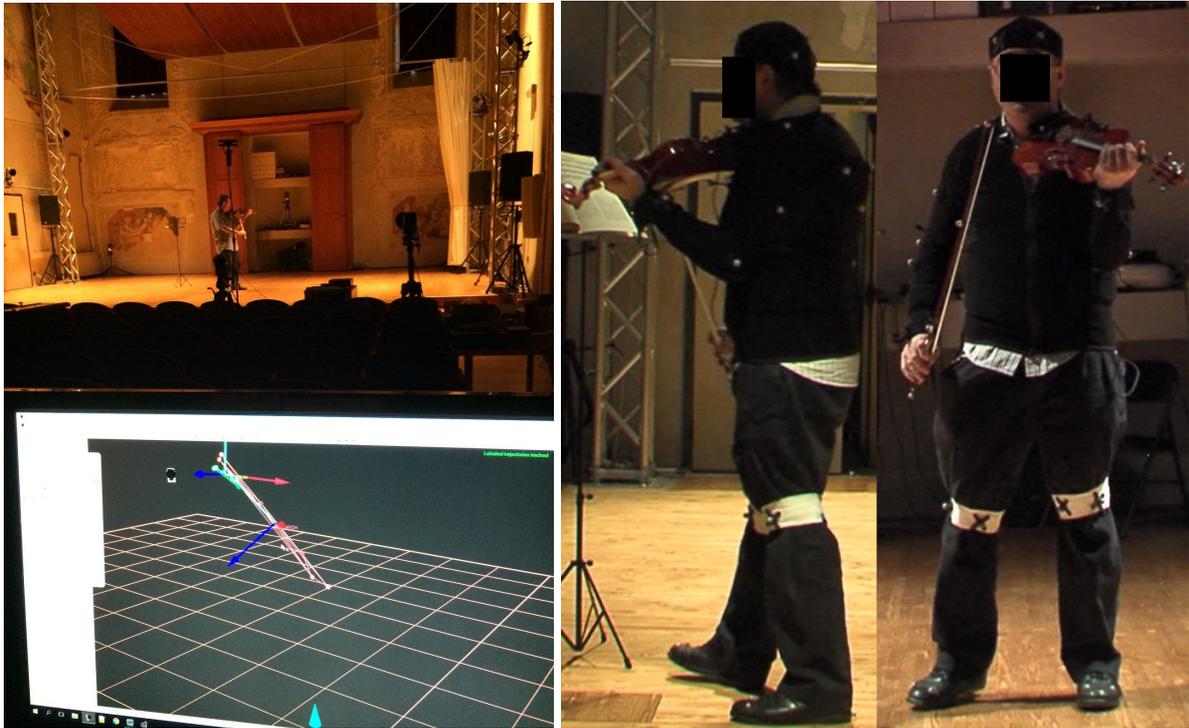


Figure 4. On the left, bow and violin positions being extracted in real-time from the information provided by the Qualysis motion capture systems installed at UNIGE's premises at Casa Paganini, Genova. On the right, a detailed view of the locations of the markers on the body of the violin player.

A video of the working system can be found at: <https://vimeo.com/184812886>

An extension of the EyesWeb XMI platform, including the developments carried out in the framework of the TELMI project (especially the support of myo sensors) is available for download here: ftp://ftp.infomus.org/pub/TELMI/EywTELMI/EyesWeb_XMI_TELMI_setup_5.6.0.0.exe (it requires to install EyesWeb before installing it).

4.1.4 Technology used

The main hardware used is the Qualysis motion capture system. Software was developed in C++ as modules for the EyesWeb XMI platform (UNIGE).

4.1.5 Next steps

Future improvements include:

- Investigating and developing algorithms to extract the whole collection of bow and body movement features described in D4.1.
- Investigating and developing algorithms to assess the technical quality of a violin performance, based on the recorded data and features.
- Research and design of musically and pedagogically effective visualizations of recorded data and extracted features, e.g., how to visualize in a meaningful way the bow and body movement features extracted from the TELMI reference archive.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

4.2 Alternative high-performance motion capture system: Polhemus EMF

A working prototype has been developed based on the Polhemus EMF system presented in D3.1 that is able to compute the bowing parameters defined in Section 3 of D4.1. The main advantages of this system are the capability of real-time computation of the bowing parameters, its high accuracy and the simple and fast procedure to calibrate and use. The main disadvantages are that it is a wired system, therefore a bit intrusive for the performer, especially in the bow. Moreover, it requires special equipment and space not available to every user or conservatory. The system presents a signal input for synchronization with other devices, which is not standard, so specific algorithms have been developed to make synchronization with audio streams possible.

4.2.1 Context and Use cases

This prototype allows for the real-time computation and visualization of highly accurate bowing controls. The importance of such a prototype is that it will be possible (which has never been done before) to design visualizations of bowing parameters (and audio features) that can be used for real-time monitoring and exploration, as well as retrospective analysis in an off-line scenario.

4.2.2 Goals / Questions

In the current version, this prototype is able to perform the following operations:

- Recording of audio from violin performance.
- Recording of violin and bow positions
- Recording and computation of bowing controls (D4.1)
- Synchronization with audio
- Visualization of 3D models of violin and bow
- Visualization of bowing control features

4.2.3 Description

As explained in Section 4 of deliverable D3.1, bowing parameters are measured with two motion sensors that track, respectively, bow and violin 6DoF motion.

A VST-Plugin (https://en.wikipedia.org/wiki/Virtual_Studio_Technology) software has been developed in C++, which is able to connect and read data from the Polhemus system, save raw data in files, compute the bowing descriptors and show the visualizations. The advantage of using a VST-Plugin at this stage, is that this is a recognised standard, that can be plugged to with any compliant DAW (Digital Audio Workstation) and facilitates to handle synchronization with audio. Synchronization is done through the VST interface by sending a clock signal to the different devices. The plugin visualizes a 3D representation of the position of the violin and bow in real time, as well as some performer actions curves, which are useful to follow the performance and detect possible errors during the calibration or the recordings. The system has been tested with great success at the first Public Event in Barcelona, where we also had feedback from musicians. Some pictures taken in that event are shown in figure 5.





Figure 5. Polhemus EMF systems being tested and demonstrated at first public event in Barcelona.

Videos of the working system can be found at:

<https://vimeo.com/176355823>

<https://vimeo.com/176355816>

4.2.4 Technology used

The main hardware used is the Polhemus System, a sound card, a computer, and a microphone. Software was developed in C++ using the SDK provided by Polhemus and following the VST standard.

4.2.5 Next steps

Future improvements include:

- Research and design of musically and pedagogically effective visualization of bowing parameters. One of the main concerns of the users that have tested the prototype is the need for visualizations adapted to specific training skills and exercises.
- Divide the application into two pieces of software, the first one responsible for the data acquisition and recording, the second one responsible for the visualizations. Both parts will communicate via OSC protocol messages. The idea to separate both parts is to allow visualizations to be represented in a device with less computational power such as tablets or mobile phones.
- Allow for loading for replay previously recorded excerpts (off-line mode).



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

4.3 Low-cost motion capture system: Kinect

An initial working prototype, based on the Kinect v.2 range imaging sensor, has been developed for analysis of bow control parameters. The final goal is to investigate which features can be reliably extracted, which is their accuracy, and at which extent they can be used in the analysis. The prototype has been developed as an application for the EyesWeb XMI platform. The current prototype can track the 3D positions of the tip and of the frog of the bow and can compute kinematic features from their trajectories. The main advantage of this prototype is its low cost. Moreover, it neither requires a calibration phase nor it needs to put any hardware or any wire on the performer and on the violin, besides two small markers on the bow. Its robustness and accuracy is, however, much lower than those reached with motion capture systems. The synchronization mechanisms embedded in EyesWeb XMI enable simultaneous recording of data from this prototype and from high-performance motion capture systems in order to allow comparison.

4.3.1 Context and Use cases

The prototype enables real-time tracking of the 3D positions of the bow tip and frog, enabling extraction, analysis, and visualization of a selected subset of bow control features. Being low-cost, it is particularly suited for being used at home or at education institutions, which are not endowed of any other more expensive and high-performance prototype.

4.3.2 Goals / Questions

In the current version the prototype is able to perform the following operations:

- Tracking 3D positions of reflecting markers on the tip and on the frog of the bow, using Kinect v.2 as input device.
- Recording the tracked positions.
- Playing back the recorded data.
- Computing a subset of bowing control features (D4.1), mainly kinematics.
- Visualizing the computed features.
- Synchronizing with audio, video, and sensors.

4.3.3 Description

In the current prototype, the tip and the frog of the bow were marked with some special tape reflecting the infrared light Kinect emits (an infrared reflexive marker such as those used with the Qualysis motion capture system also suits). Such a tape on the one hand is not particularly intrusive and does not affect the music performance and, on the other hand, provides quite robust markers for motion tracking. The velocity of the bow and its distance from the Kinect device are then computed (see also D4.1). Figure 6 shows the low-cost motion capture system, based on Kinect.

Videos of the working system can be found at: <https://vimeo.com/176358282>

An extension of the EyesWeb XMI platform, including the developments carried out in the framework of the TELMI project (especially the EyesWeb XMI application for real-time tracking of the bow tip and frog using Kinect v.2) is available for download here:

ftp://ftp.infomus.org/pub/TELMi/EywTELMi/EyesWeb_XMI_TELMI_setup_5.6.1.0.exe
(it requires to have previously installed EyesWeb 5.6.1.0).



4.3.4 Technology used

The main hardware used is the Kinect v.2 device. Software was developed as an application for the EyesWeb XMI platform (UNIGE).

4.3.5 Next steps

Future improvements include:

- Investigating and developing algorithms to extract a broader collection of bow control features, among those described in D4.1.
- Investigating and developing algorithms to extract a collection of body movement features, among those described in D4.1.
- Assessing robustness and accuracy with respect to high-performance systems.
- Research and design of musically and pedagogically effective visualizations of recorded data and extracted features.



Figure 6: the low-cost motion capture system, based on Kinect v.2, being used and demonstrated at the final concert of the TELMI Workshop in Barcelona (July 1st, 2016). The system tracked the bow movement of the second violin of the string quartet and computed kinematic measures on the trajectories of the of the bow tip and frog. Measures were displayed on the screen in the background.

5 Synchronization and Data Storage

In this section we include two technologies that although they are not part of the measuring system prototypes they are somehow related with them so it makes sense to include a brief description of them in this deliverable. On the one hand, EyesWeb is an open platform to support the design and development of real-time multimodal systems and interfaces and it is used for synchronization of multimodal data streams between different devices. On the other hand, repovizz is an online platform for remote storage, browsing, annotation and sharing of multimodal data and it is used to store the recordings done during the TELMI project and to make them available publicly online.

5.1 EyesWeb

EyesWeb (www.casapaganini.org) is a modular software platform, where a user can assemble single modules in an application by means of a visual programming language. EyesWeb supports its users in designing and developing interactive multimodal systems in several ways, such as for example (i) by providing built-in input/output capabilities for a broad range of sensor and capture systems, (ii) by enabling



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

to easily define and customize how data is processed and feedback is generated, and (iii) by offering tools for creating a wide palette of interfaces for different classes of users.

In order to fully support synchronised recordings and analysis of multimodal data, EyesWeb XMI is endowed of a synchronisation mechanism, based on SMPTE timestamps. SMPTE (Society of Motion Picture and Television Engineers, from the name of the society who defined this standard) is a protocol that is commonly used to synchronise video and audio content. Each multimodal data is stored with an SMPTE timestamp as the data is saved to a file. In such a way, EyesWeb makes it possible to operate a synchronisation procedure to acquire different signals from different systems working at different sample rates, with different internal clock, and with possible drifts and delays in data acquisition. Such synchronisation can be achieved in real-time, but a more robust synchronisation is achieved after the recording, based on the SMPTE synchronisation streams. SMPTE is used both during the recording and during the data analysis and playback. During the recording phase, SMPTE is used as a hardware synchronisation signal for devices supporting it. If the hardware device is able to receive SMPTE as a synch signal, the device is locked to such a clock, i.e., no bias occurs among devices locked in this way. As an example, the Qualisys Motion Capture System, the Fireface 800 RME audio card, and, partially, the JVC GY-HD251 video cameras support this type of synchronisation. The multimodal recordings planned in the TELMI project make use of these devices. If the device does not accept a clock input signal, a software synchronisation can be achieved by means of the EyesWeb XMI platform. EyesWeb receives both the SMPTE signal and the data to be recorded, and adds a timestamp to each sample. As an example, low-cost RGB-D sensors (e.g., Kinect) and biometric sensors (e.g, the myo sensors) fall in this category of devices: they can be soft-synchronised by means of the EyesWeb XMI platform. The SMPTE information is recorded together with the time-stamped data; in the case of audio/video files, the SMPTE information is recorded as an additional audio track; in the case of MoCap data or analog sensors data (including biometric sensors), SMPTE is recorded as an additional time value. EyesWeb XMI is publicly available at www.casapaganini.org.

5.2 Repovizz

5.2.1 Description

For the public database, the repovizz platform [1] will be used. Repovizz (<http://repovizz.upf.edu>) is an integrated online system capable of structural formatting and remote storage, browsing, exchange, annotation and visualization of synchronous multi-modal, time-aligned data. Motivated by a growing need for data-driven collaborative research, repoVizz aims to resolve commonly encountered difficulties in sharing or browsing large collections of multi-modal data. At its current state, repovizz is designed to hold time-aligned streams of heterogeneous data: audio, video, motion capture, physiological signals, extracted descriptors, annotations, and so on . Most popular formats for audio and video are supported, while Broadcast WAVE or CSV formats are adopted for streams other than audio or video (e.g., motion capture or physiological signals). The data itself are structured via customized XML files, allowing the user to (re-)organize multi-modal data in any hierarchical manner, as the XML structure only holds metadata and pointers to data files. Datasets are stored in an online database, allowing the user to interact with the data remotely through a powerful HTML5 visual interface accessible from any standard web browser; this feature can be considered a key aspect of repovizz since data can be explored, annotated or visualized from any location or device. Data exchange and upload/download is made easy and secure via a number of data conversion tools and a user/permission management system.

5.2.2 Data formatting, storage and sharing

Once the different data streams from different modalities are recorded using the above described prototypes they need to be time synchronized as explained in the previous section either using EyesWeb or manually for some low cost sensors that cannot be integrated in the above mentioned method. After



synchronization, the data is formatted accordingly to be compatible with formats accepted in repovizz. The following formats are used for each type of data:

- Audio: any common audio format that can be decoded by ffmpeg (wav, mp3, ogg, flac, aac, etc). Once uploaded to repovizz original audio streams are kept in the server and are additionally converted to wav files at a sampling rate of 44.1Khz and 16 bits for audio feature extraction. Web friendly mp3 and ogg files are generated.
- Video: any common video format that can be decoded by ffmpeg (mp4, avi, mkv, mov, webm, etc). Once uploaded to repovizz original video streams are kept in the server and are additionally converted to webm and mp4 at a resolution of 720p to make them compatible with standard html5 browsers.
- Time varying Signals / Descriptors: csv containing a header line as defined in the repovizz tutorial [2]
- Musical Scores: music xml (compatible with the musescore open source software)
- Mocap Data: multiple csv files for each marker coordinate as defined in the repovizz tutorial [2]
- Annotations: txt files containing lines with time and label information as defined in the repovizz tutorial [2]

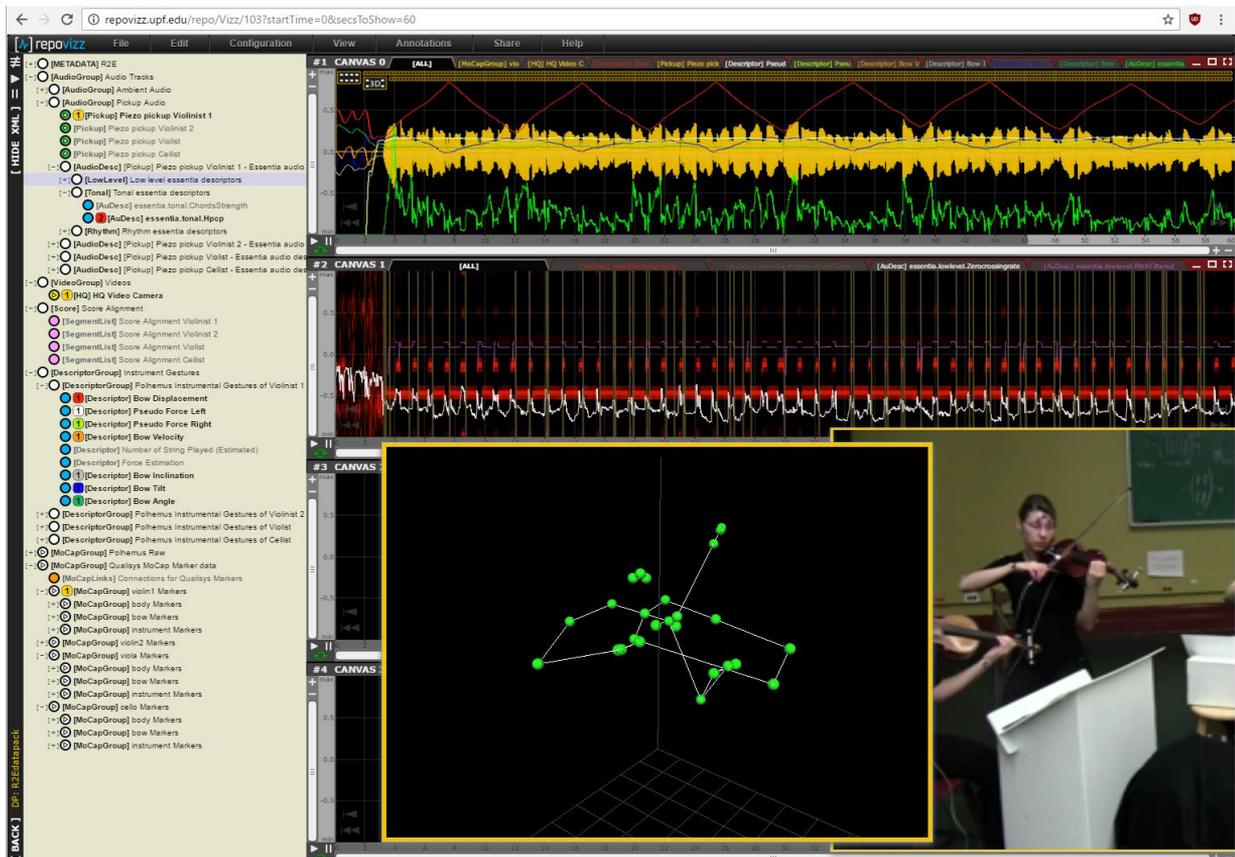


Figure 7: a violin performance recording in repovizz including audio, video, mocap and audio descriptors.

The upload of data streams can be done directly from the prototypes using the repovizz RESTful API [3]. This feature is not implemented yet for any of the abovementioned prototypes but the repovizz API has



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688269

already this functionality. The API also allows users to access and use all data stored in repovizz programmatically. Using the API users can browse, search, list, and download datasets and individual streams contained in the repository.

The public database stored in the repovizz infrastructure will serve as a sharing and visualization platform, allowing third parties to download data as well as visualize it in a user friendly way just opening a url in a browser (see figure 7).

6 Conclusions

In conclusion, the current development of the first version of the prototypes of the TELMI measuring systems addresses the expected results as stated in the DoW by providing robust prototypes for data acquisition as well as computation of preliminary features that will serve as a base for the TELMI research objectives. There are some overlaps between different prototypes in terms of features and further integration should be explored. Our approach was to let emerge common goals/questions that could arise from different modalities, technologies or use cases. This convergence of questions will be the foundation of the development of the second version and final version of TELMI prototypes.

The next steps concern the refinement of the current prototypes, their evaluation and incorporation of selected technologies from these prototypes in the TELMI platform.

7 References

[1] Mayor, O., Llop J., & Maestre E. RepoVizz: A multimodal on-line database and browsing tool for music performance research. 12th International Society for Music Information Retrieval Conference (ISMIR 2011)

[2] Repovizz Tutorial (<http://repovizz.upf.edu/misc/repoVizzTutorial.pdf>)

[3] Repovizz RESTful API documentation (<http://repovizz.upf.edu/repo/api-doc/index.html>)

[4] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognit. 47, 2280–2292. doi:10.1016/j.patcog.2014.01.005.

