Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis[*]

Thomas Fillon[†]
LAM,
UPMC Univ. Paris 06
UMR CNRS 7190
thomas@parisson.com

Joséphine Simonnot
CREM, LESC
UMR CNRS 7186)
MAE, Univ. Paris Ouest
Nanterre La Défense
Nanterre, France
josephine.simonnot@mae.u-paris10.fr

Marie-France Mifune
CNRS-MNHN-Université Paris
Diderot-Sorbonne cité
UMR 7206 Eco-anthropologie
et ethnobiologie
Paris, France
mifune@mnhn.fr

Stéphanie Khoury
CREM, LESC
UMR CNRS 7186)
MAE, Univ. Paris Ouest
Nanterre La Défense
Nanterre, France
stephanie.khoury@mae.u-paris10.fr

Guillaume Pellerin
PARISSON
16 rue Jacques Louvel-Tessier
Paris, France
guillaume@parisson.com

Maxime Le Coz
IRIT-UPS
118, route de Narbonne
31062 TOULOUSE CEDEX 9,
France
lecoz@irit.fr

Estelle Amy de la Bretèque
CREM, LESC
UMR CNRS 7186)
MAE, Univ. Paris Ouest
Nanterre La Défense
Nanterre, France
ebreteque@gmail.com

David Doukhan
Université Paris-Sud /
CNRS-LIMSI
Orsay, France
David.Doukhan@limsi.fr

Dominique Fourer
LaBRI - CNRS UMR 5800
Université Bordeaux 1
351, cours de la Libération,
33405 Talence, France
fourer@labri.fr

ABSTRACT

The audio archives of the CNRS-Musée de l'Homme are among the most important collections of ethnomusicological recordings in Europe. Yet, as the number of collections increase and as new audio technologies arise, questions linked to the preservation, the archiving and the availability of these audio materials have arisen. With this in mind, since 2007, ethnomusicologists and engineers have joined their efforts to develop a scalable and collaborative web platform for managing and increasing access to digitized sound archives. This web platform is based on Telemeta, an open-source web audio framework dedicated to digital sound archives. Since 2011, the Telemeta framework has been deployed to hold the platform of the CNRS-Musée de l'Homme's audio archives, which are managed by the Research Center for Ethnomusicology. This framework focuses on the enhanced and collaborative user experience in accessing audio items and their associated metadata. The architecture of Telemeta relies on TimeSide, an open-source audio processing framework written in Python and JavaScript languages, which provides decoding, encoding and streaming capabilities together with a smart embeddable HTML audio player. TimeSide can also produce various automatic annotation, segmentation and musicological analysis that have been developed in the interdisciplinary research project called DIADEMS. Furthermore it includes a set of audio analysis plug-ins and wraps several audio features extraction libraries. This paper introduces the Telemeta framework and discuss how, experimenting with this advanced database for ethnomusicology through the DIADEMS project, cutting-edge tools are being implemented to fit and encourage new ways to relate to sound libraries.

1. INTRODUCTION

In social sciences, as very large scientific databases become available and rapidly increase by both number and volume, their management thus rises new fundamental questions as well as new research challenges. In anthropology, ethnomusicology and linguistics, researchers work on multiple kinds of multimedia documents such as photos, videos and sound recordings. The need to preserve and to easily access, visualize and annotate such materials is problematic given their diverse formats, sources and the increasing quantity

[*]This work is partly supported by a grant from the french National Research Agency (ANR) with reference ANR-12-CORD-0022.

[†]This author is also affiliated to *PARISSON, 16 rue Jacques Louvel-Tessier, Paris, FRANCE*

of data. In the context of ethnomusicological research, the Research Center for Ethnomusicology(CREM) and Parisson, a company specialized in big music data projects, have been developing an innovative, collaborative and interdisciplinary open-source web-based multimedia platform since 2007. This platform, Telemeta is designed to fit the professional requirements from both sound archivists, researchers and musicians to work together on huge amounts of music data. The first prototype of this platform has been online since 2010 and is now fully operational and is used on a daily basis for ethnomusicological studies since 2011.

Recently, an open-source audio analysis framework, TimeSide, has been developed to bring automatic music analysis capabilities to the web platform and thus have turned Telemeta into a complete resource for Computational Ethnomusicology [20, 10]. The Section 3 focuses on this framework.

The benefits of this collaborative platform for humanities and social sciences research apply to numerous aspects of the field of ethnomusicology, ranging from musical analysis to comparative history and anthropology of music, as well as to the fields of anthropology, linguistics and acoustics. Some of these benefits have been mentionned in several ethnomusicological publications [17, 11, 18]. The current and potential applications of such a platform thus raises the needs and the underlying challenges of implementing an online digital tool to support contemporary scientific research.

2. THE TELEMETA PLATFORM

2.1 Web audio content management features and architecture

The primary purpose of the project is to provide the communities of researchers working on audio materials with a scalable system to access, preserve and share sound items along with associated metadata that contains key information on the context and significance of the recording. Telemeta[1], as a free and open source software[2], is a unique scalable web audio platform for backuping, indexing, transcoding, analyzing, sharing and visualizing any digital audio or video file in accordance with open web standards. The time-based nature of such audio-visual materials and some associated metadata as annotation raises issues of access and visualization at a large scale. Easy and on-demand access to these data, while listening to the recording, represents a significant improvement. An overview of the Telemeta's web interface is illustrated in Figure 1a. Its flexible and streaming safe architecture is represented in Figure 2. The main features of Telemeta are:

- Pure HTML5 web user interface including dynamic forms
- On-the-fly audio analyzing, transcoding and metadata embedding in various multimedia formats
- Social editing with semantic ontologies, smart workflows, realtime tools, human or automatic annotations and segmentations
- User management with individual desk, playlists, profiles and group access rights
- High level search engine (geolocation, instruments, ethnic groups, etc...)
- Data providers : DublinCore, OAI-PMH, RSS, XML, JSON and other
- Multi-language support (currently english and french)

Beside database management, the audio support is mainly provided through an external component, TimeSide, which is described in Section 3.

2.2 Metadata

In addition to the audio data, an efficient and dynamic management of the associated metadata is also necessary. Consulting metadata grants both an exhaustive access to valuable information about the source of the data and to the related work of peer researchers. Dynamically handling metadata in a collaborative manner optimizes the continuous process of knowledge gathering and the enrichment of the materials in the database. One of the major challenges is the standardization of audio and metadata formats with the aim of long-term preservation and usage of the different materials. The compatibility with other systems is facilitated by the integration of the metadata standards protocols Dublin Core[3] and *OAI-PMH* (Open Archives Initiative Protocol for Metadata Harvesting)[4]. The metadata includes two different kinds of information about the audio item: contextual information and analytical information of the audio content.

2.2.1 *Contextual Information*

In an ethnomusicological framework, contextual information may include details about the location where the recording has been made, the instruments, the population, the title of the musical piece, the cultural elements related to the musical item, the depositor, the collector, the year of the recording and the year of the publication. Moreover, through the platform, diverse materials related to the archives can be stored, such as iconographies (digitalized pictures, scans of booklet and field notes, and so on), hyperlinks and biographical information about the collector.

2.2.2 *Descriptive and analytical information on the audio content*

The second type of metadata consists in information about the audio content itself. This metadata can relate to the global content of the audio item or provide temporally-indexed information. It should also be noted that such information can be produced either by a human expert or by an automatic computational audio analysis (see Section 3 below).

Visual representation and segmentation.

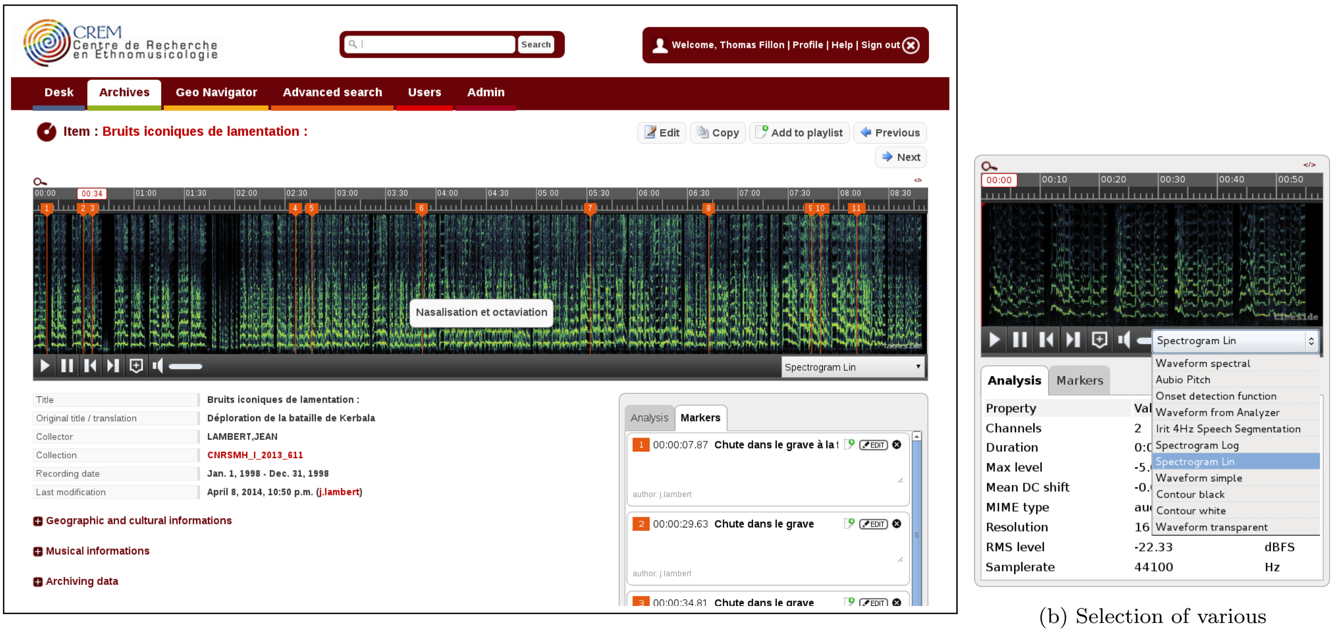
[1] http://telemeta.org

[2] Telemeta code is available under the CeCILL Free Software License Agreement

[3] Dublin Core Metadata Initiative, http://dublincore.org/

[4] http://www.openarchives.org/pmh/

Digital Libraries for Musicology workshop (DLfM 2014) London, UK
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$15.00.



Item : Bruits iconiques de lamentation :

Title	Bruits iconiques de lamentation :
Original title / translation	Déploration de la bataille de Kerbala
Collector	LAMBERT,JEAN
Collection	CNRSMH_I_2013_611
Recording date	Jan. 1, 1998 - Dec. 31, 1998
Last modification	April 8, 2014, 10:50 p.m. (j.lambert)

Geographic and cultural informations

Musical informations

Archiving data

1 00:00:07.87 Chute dans le grave à la f

author: j.lambert

2 00:00:29.63 Chute dans le grave

author: j.lambert

3 00:00:34.81 Chute dans le grave

Property	Val	
Channels	2	
Duration	0:0	
Max level	-5.	
Mean DC shift	-0.	
MIME type	au	
Resolution	16	
RMS level	-22.33	dBFS
Samplerate	44100	Hz

(a) Screenshot excerpt

(b) Selection of various sound representations

Figure 1: Telemeta web interface

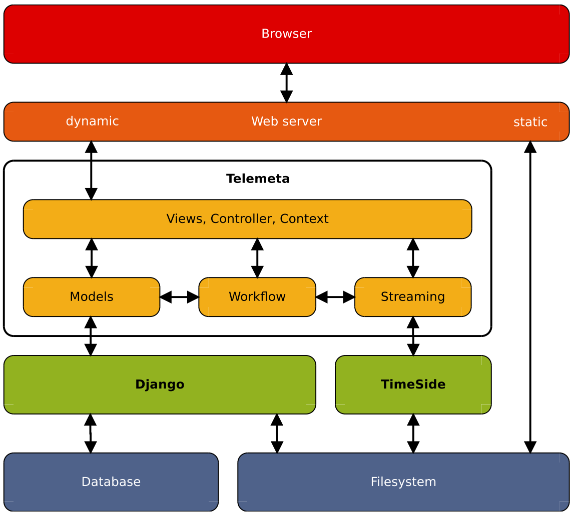


Figure 2: Telemeta architecture

As illustrated in Figure 1b, the TimeSide audio player embedded in the Telemeta web page view of a sound item allows for a selection of various visual representations of the sound (e.g. waveforms and spectrograms, see section 3 for details) and some representations of computational analysis.

Among those automatic analysis some can produce a list of of time-segments associated with labels. Those labels have been specified by the partners of the DIADEMS project (see Section 5 to be relevant for ethnomusicological studies (e.g. detection of spoken voice versus sang one, chorus, musical instrument categories, and so on).

Annotations.

As illustrated in Figure 1a, the embedded audio player also enables to annotate the audio content through time-coded markers. Such annotations consist in a title and a free text field associated with a given time position.

Ethnomusicologists, archivists, as well as anthropologists, linguists and acousticians working on sound documents can create their own annotations and share them with colleagues. These annotations are accessible from the sound archive item web page and are indexed through the database.

It should be noted that the possibility for experts to annotate time-segments over a zoomable representation of the sound is currently under development in order to improve the accuracy and the quality of time-segment base annotations.

3. TIMESIDE, AN AUDIO ANALYSIS FRAMEWORK

As illustrated in Figure 2, one specificity of the Telemeta architecture is to rely on an external component, TimeSide[5], that offers audio player web integration together with audio signal processing analysis capabilities. TimeSide is an audio analysis and visualization framework based on both Python and JavaScript languages to provide state-of-the-art signal processing and machine learning algorithms together with web audio capabilities for display and streaming. Figure 3 illustrates the overall architecture of TimeSide together with the data flow between TimeSide and the Telemeta webserver.

3.1 Audio management

TimeSide provides the following main features:

- Secure archiving, editing and publishing of audio files over internet
- Smart audio player with enhanced visualization (waveform, spectrogram)

[5] `https://github.com/yomguy/TimeSide`

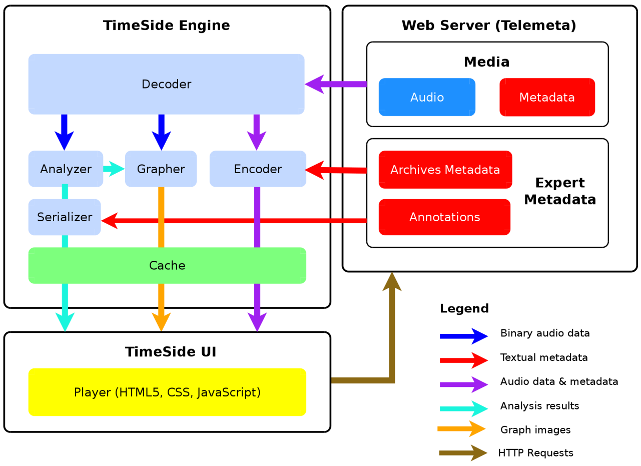


Figure 3: TimeSide engine architecture and data flow with Telemeta web-server

- Multi-format support: decodes the vast majority of audio and video formats through Gstreamer and transcodes them with smart streaming and caching methods
- On-the-fly audio analyzing, transcoding and metadata embedding based on an easy plugin architecture

3.2 Audio features extraction

In order to implement Music Information Retrieval (MIR) analysis methods to be carried out over a large corpus for ethnomusicological studies, TimeSide incorporates some state-of-the-art audio feature extraction libraries such as Aubio[6] [4], Yaafe[7] [13] and Vamp plugins[8]. Given the extracted features, every sound item in a given collection can be automatically analyzed. The results of this analysis can be stored in a scientific file format (e.g. NumPy format and HDF5), exported to sound visualization and annotation softwares like sonic visualizer [5], or serialized to the web browser through common markup languages: XML, JSON and YAML.

As an open-source framework and given its architecture and the flexibility provided by Python, the implementation of many audio and music analysis algorithms can be considered and then apply to audio archives through Telemeta. So, it is a very interesting platform for researchers in computational musicology to develop and evaluate their algorithms.

4. SOUND ARCHIVES OF THE CNRS - MUSÉE DE L'HOMME

Since June 2011, the Telemeta platform is used by the Sound archives of the CNRS - Musée de l'Homme[9] and managed by the CREM. According to the CREM specific aims, the Telemeta platform makes these archives available for researchers, students and, when copyright allows it, to a broader audience. Through this platform, these archives can be shared, discussed and worked on.

The Telemeta platform has also been deployed for the sound archives of the *String instruments - Acoustic - Music* team of the "Jean Le Rond d'Alembert Institute"[10].

4.1 Archiving research materials

The Sound archives of the CNRS - Musée de l'Homme is one of the most important in Europe and gather commercial as well as unpublished recordings of music and oral traditions from around the world, collected by researchers attached to numerous research institutions across the world, among which some prominent figures of the field of ethnomusicology (among which Brailoiu, Lomax, Shaeffner, Rouget and Elkin).

The platform offers access to records collection (nearly 3700 hours, e.g. more than 5000 discs, many of which are very rare) and to 4000 hours of unpublished recordings, as early expeditions (e.g. Dakar-Djibouti (1932), Ogooué-Congo (1946)). Most of the recordings come from the fieldwork of researchers in all the continents. More than 110 years of the world's oral culture are now available online, from the 1900 Universal Exhibition of Paris up to the recent digital recordings. The sharing of data allows several people to collaborate to the enrichment of the database. Today, 47,200 items are in the database, and more than 26,000 sound files have been included (12,000 sounds on free access since May 2014). Recently, the CREM has decided to give full access to the records published by the CNRS-Musée de l'Homme (Chant du Monde/Harmonia Mundi)[11] which distribution stopped ten years ago. As a web platform, this tool is also a way to cross borders, to get local populations involved in their own cultural heritage and to offer resources to researchers from all over the world.

4.2 Uses and users of digital sound archives

Through the few years since the sound archive platform has been released, it appears to support three main activities: archive, research and education (academic or not). These usages are those of archivists, researchers (ethnomusicologists, anthropologists and linguists), students and professors of these disciplines. Nonetheless, a qualitative survey showed that other disciplines (such as art history) found some use of the platform to foster and/or deepen individual research. The unexpectedly broad uses of the sound archives once digitized and accessible emphasize the necessity and the benefits of such database. From the standpoint of archive development, the long-term preservation of the archives is ensured while, thanks to the collaborative nature of the platform, users can cooperate to continuously enrich metadata associated with a sound document and submit their own archives to protect them. Furthermore, it allows fulfilling the ethical task of returning the recorded music to the communities who produced it. Researchers from different institutions can work together on specific audio materials as well as conduct individual research in both synchronic and diachronic perspective, on their own material, material of others or both. When use for education, the platform provides a wide array of teaching materials to illustrate the work of students as well as support teaching *curricula*.

[6] http://aubio.org/

[7] https://github.com/Yaafe/Yaafe

[8] http://www.vamp-plugins.org

[9] http://archives.crem-cnrs.fr

[10] http://telemeta.lam-ida.upmc.fr/. Online since 2012, these archives consist in recordings of a wide range of musical instruments, mostly including solo recording of traditional instruments and illustrating various playing techniques and are used as materials for research in acoustics.

[11] http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Editions/

5. EXPANDING DEVELOPMENT: THE DIADEMS PROJECT

The goals and expectations of the platform are of many kinds and expand through time, as users experience new ways to work with the archives database and request new tools to broaden the scope of their research activities linked to it. The reflexion collectively engaged by engineers and researchers on the use of the sound archives database led us to set up a large scale project called DIADEMS (*Description, Indexation, Access to Ethnomusicological and Sound Documents*)[12]. Started in January 2013, the French national research program DIADEMS is a multi-disciplinary project whose consortium includes research laboratories from *Science and Technology of Information and Communication*[13] (IT) domain, *Musicology and Ethnomusicology*[14] domain and Parisson, a company involved in the development of Telemeta.

The goal of the DIADEMS project is to develop computer tools to automatically index the recording content directly from the audio signal in order to improve the access to and the indexation of this vast ethnomusicological archive. Numerous ethnomusicological recordings contain speech and other types of sound that we categorized as sounds from the environment (such as rain, biological sounds, engine noise and so on) and sounds generated by the recording (such as sound produced by the wind on the microphone or sounds resulting from the defect of the recording medium). The innovation of this project is to automatize the indexation of the audio recordings directly from the recorded sound itself. Ongoing works consist in implementing advanced classification, indexation, segmentation and similarity analysis methods dedicated to ethnomusicological sound archives. Besides music analysis, such automatic tools also deal with speech and other types of sounds classification and segmentation to enable a more exhaustive annotation of the audio materials.

The automatic analysis of ethnomusicological sound archives is considered as a challenging task. Field recordings generally contain more sound sources, noise, and recording artefacts than those obtained in studio conditions, so the automatic analysis of these recordings requires methods having a stronger robustness. Preliminary implementations of speech detection models, and speaker diarization methods, based on [3] have been integrated to TimeSide. While these models are well suited to radio-news recordings, the current developpement tasks consist to adapt them to the particular case of ethnographic archives.

In the context of this project, researchers from Ethnomusicological, Speech Processing and MIR communities are working together to specify the tasks to be addressed by automatic analysis tools.

5.1 The method of a new interdisciplinary research

In this research program, groups from different backgrounds are working together to specify the automatic analysis tools: IT developers, humanities researchers (anthropologists, ethnomusicologists, ethnolinguists) and specialists in speech processing and MIR. The first challenge was to initiate a common interest and a mutual understanding. In this process, DIADEMS gave us the opportunity to improve our understanding on the link between the semantics and acoustics of voice production. As a prelimirary work we attempted to first define vocal categories with a particular interest for liminal oral productions. At the border between speech and song, utterances such as psalmody or recitation are at the center of an old debate in ethnomusicology[15]. Gathering specialists from various fields, DIADEMS project goes well beyond the usual disciplinary boundaries. Our aim, through the study of a large range of audio components (pitch range, syllabic flow, metric, polyphonic and so on) is to define and characterize the variability of vocal productions, keeping in mind the semantic aspects. By doing so, we wish to reduce the traditional gap in academic studies between sounds and semantics and to propose combined analytical tools for the study of vocal production[16].

One of the goals of the DIADEMS project is also to provide useful tools for musical analysis such as detection of musical instrument families, analysis of musical content (tonal, metric and rythmic features), musical similarities and structure (chorus localisation, musical pattern replication).

The study follows three steps :

1. The development of tools and selection of a representative corpus for each tool
2. The evaluation of the proposed automatic analysis, in addition to the human-led evaluations carried on the corpus selected
3. The development of a visual interface with an ergonomic access and import of the results in the database

5.2 Automatic tools for assisting indexation and annotation of audio documents

A first concern was to develop an automated annotation component that could differentiate spoken from sung voice and from instrumental music. Regarding that task, the state-of-the-art methods have usually been specifically designed to fit the needs of radio broadcast data (i.e. clear recordings produced in studios) and are not adapted to face the sonic diversity of ethnomusicological field recordings. For these, more refined detection tools were needed to pick up sound events such as overlapping speeches, speech over music, as well as instrumental music mixed with singing and/or spoken interventions. Beyond the implementation of tools detecting the beginning and the end sound signatures of magnetic, mechanical and digital recorders as well as tape noises and silences, numerous algorithms allow for complex automated analysis for a wide range of combinations of vocal and instrumental sounds.

[12] `http://www.irit.fr/recherches/SAMOVA/DIADEMS/en/welcome/`

[13] IRIT (Institute of research in computing science of Toulouse), LABRI (Bordeaux Computer Science Research Laboratory), LIMSI (Laboratory of computing and mechanics for engineering sciences), LAM (String instruments - Acoustic - Music, Jean Le Rond d'Alembert Institute)

[14] LESC (Laboratory of Ethnology and Comparative Sociology), MNHN (National Museum of Natural History)

[15] A colloquium on liminal utterances between speech and song will be organised by the International Council for Traditional Music (ICTM) in May 2015 and hosted by the Centre of research in Ethnomusicology (CREM). A round table will be dedicated to the presentation of the main results and findings of the ANR project Diadems

[16] As an example, research will be conducted on the recognition of ”icons of crying” in lamented utterances. As defined by Urban in [21], ”icons of crying” include cry break, voice inhalation, creaky voice and falsetto vowels.

5.2.1 Analysis of recordings sessions

A primary concern of people dealing with audio materials coming from field recordings is to quickly and automatically identify the position of the start and the stop time of a recording session.

In the case of digital recorder, the proposed system for localizing the start of recording sessions is based on the observation that on such recorders, the powering of the recorder implies a characteristic and quite reproducible perturbation of the signal as illustrated in Figure 4. Therefore, this perturbation can be modeled by using some reference models of the shape of the temporal energy. On the analyzed recordings, the evolution of the temporal energy of low energy segments is then compared to the models by an euclidean distance using the best alignment of the energy shape.

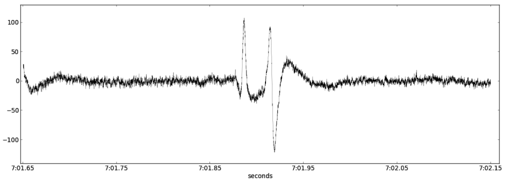


Figure 4: Typical perturbation produced by the recorder powering, used to recognize session start

5.2.2 Analysis of speech and singing voice segments

Quick identification and localization of spoken sections, particularly in rather long recordings, are relevant for all the disciplines involved in the project. The difficulties inherent in the sound materials led to the development of tools to automatically detect the occurrences of speech when performed simultaneously or alternatively with music; when numerous speakers interact and/or overlap with each others, with or without additional music or noises; and when voices modulate from speech to song, using a wide range of vocal techniques (recitation, narration, psalmody, back channel, and so on). The developed algorithms allow the analysis of the syllabic flow and the prosody of the speech. Figure 5 shows a visual example of how speech segmentation is rendered.

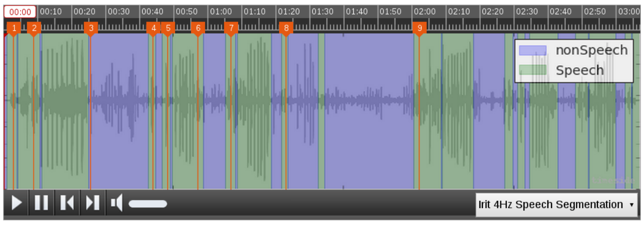


Figure 5: Detection of spoken voices in a song

Speech segmentation, with 2 features: 4 Hz modulation energy and entropy modulation.

Speech signal has a characteristic energy modulation peak around the 4 Hertz syllabic rate [19]. In order to model this property, the signal is filtered with a FIR band pass filter, centered on 4 Hertz. Entropy modulation is dedicated to discriminate between speech and music [16]. We first evaluate the signal entropy ($H = -\sum_{i=1}^{k} p_i \cdot log_2(p_i)$, where p_i denotes the probability of event i). Entropy modulation values are usually larger for speech than for music. This measure is used to compute the entropy modulation on each segment.

Speech activity detection based on GMM models.

Speech activity detection is a prerequisite for several speech-related tasks to be integrated in the platform such as speech segmentation, speaker diarization and so on. For this approach, a Gaussian Mixture Model system operating at the frame level, is used to learn speech spectral properties from annotated data. MFCC, together with their first and second time derivative, log energy and zero crossing rate audio features are used to train the GMM models to discriminate speech frames from non speech frames. ETAPE corpus [9] (speech-based TV archives in French), and data obtained from the French Center of Research and Teaching on Amerindian Ethnology (carnival rituals in Mayan language) were used to train two distinct sets of models. Both models have been serialized and integrated to TimeSide together with the GMM-based speech activity detection system, allowing the final user to choose a given model according to the properties of the media to analyze.

5.2.3 Analysis of music segments

The DIADEMS project aims to provide useful tools for musical analysis in both research and teaching frameworks. So, the detection of segments of instrumental music along with the recognition of the different musical instrument categories is needed. Pushing the detection further into details, the implemented tools provide musicological information to support sound analysis (such as tonal, metric and rhythmic features) and allow detecting of similarities in melody, harmony and rhythm as well as musical pattern replications.

Music segmentation, with 2 features based on a segmentation algorithm.

This segmentation is provided by the Forward-Backward Divergence algorithm, which is based on a statistical study of the acoustic signal [2]. The speech signal is assumed to be composed of a sequence of quasi-stationary units that can be seen as alternate periods of transient and steady parts (steady parts are mainly vowels). We characterize each of these units by an Auto Regressive (AR) Gaussian model. The method consists in performing a detection of changes in AR models. Indeed, music is usually much more constant than speech, that is to say the number of changes (segments) will be smaller for music than for speech. To estimate this, we count the number of segments per second of signal. The number of segments is the first discriminative feature for music segmentation. The segmentation algorithm generally produces longer segments for music than for speech. We chose to model the segment duration by an inverse Gaussian distribution (or Wald distribution) which is indeed the second feature providing music segmentation.

Monophony / Polyphony segmentation.

A ”monophonic” sound is defined as one note played at a time (either played by an instrument or sung by a singer), while a ”polyphonic” sound is defined as several notes played simultaneously. The parameters extracted from the signal come from the YIN algorithm, a well known pitch estimator [6]. Besides F0, this estimator provides an additional numerical value that can be interpreted as the inverse of a confidence indicator (the lower the value, the more reliable the estimated pitch). Considering that when there is a single note the estimated pitch is fairly reliable, and that when there are several simultaneous notes, the estimated pitch is not reliable, we take the short term mean and variance of this ”confidence indicator” as parameters for the monophony / polyphony segmentation. The bivariate distribution of these two parameters is modelled using Weibull bivariate distributions [12]. An example of the segmentation produced by this method is illustrated in Figure 6

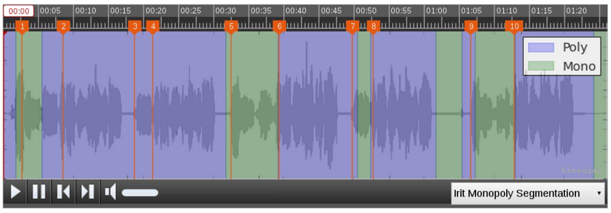


Figure 6: Detection solo and duo parts

Automatic instrument classification.

For the detection of musical instrument, we choose to follow the Hornbostel–Sachs system of musical instrument classification as first published in [22] and later translated in [23]. This system is the most widely used system for classifying musical instruments by ethnomusicologists and organologists. It was extended by in more recent systems like the one proposed by Geneviève Dournon in [7] and the *RAMEAU* reference from the French national library (BnF)[17].

We choose to develop tools to detect the four musical instrument families (cordophones, aerophones, idiophones, membranophones), but also refine subdivisions related to the playing techniques, identifying whether each instrument is blowed, bowed, plucked, struck or clincked by focusing on the 8 classes of instrument shown in Figure 7.

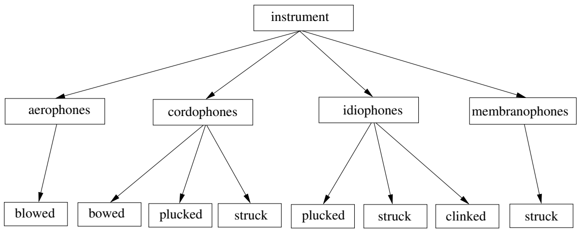


Figure 7: Musical instrument families

The proposed method is based on the supervised learning approach and uses a set of 164 acoustic descriptors proposed by Peeters *et al.* in [15]. This system (see Figure 8) applies the Inertia Ratio Maximization Features Space (IRMFSP) algorihtm [14] on annotated samples at the training step to reduce the number of features (to avoid overfitting) while selecting the most discriminative ones.

The Linear Discriminant Analysis (LDA) method [1] is used to compute the best projection space or linear combination of all descriptors which maximizes the average distance between classes while minimizing the distance between individuals of the same class.

Each class is simply modeled into the classification space by its centroid (mean vector computed over the selected descriptors of the individuals which compose the class). Thus, the classification task consists in projecting each tested sound described by a vector of descriptors to the classification space and to select the closer class (in term of Euclidean distance to the class centroid).

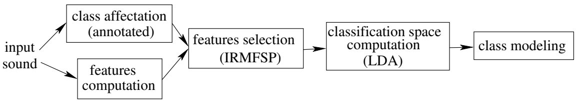

input sound
class affectation (annotated)
features computation
features selection (IRMFSP)
classification space computation (LDA)
class modeling

Figure 8: Training step of the proposed classification method

This simple and computationally efficient method obtains about 75% of accuracy using the 20 most relevant descriptors projected on a 8-dimensions discriminative space. In a more detailed study [8], this promising result was shown comparable with state-of-the art methods applied for the classification of western instruments recorded in studio condition.

5.3 Evaluation and sought improvements

At the end of the first step of the project, interesting preliminary results have been obtained regarding start of recording session detection, speech recognition, singing voice recognition and musical instrument family classification.

Through a collaborative work, ethnomusicologists, ethnolinguists and engineers are currently evaluating, correcting and refining the implemented tools, with the expectation that this work will lead to positive results, so these new tools can be integrated into the Telemeta platform.

The robustness of all these processing are assessed using criteria defined by the final users: teachers, students, researchers or musicians. Annotation tools, as well as the provided annotations, will be integrated in the digitalized database.

Further work on the user interface aims to enhance the visualization experience with time and frequency zooming capabilities, in the hope that it will improve the accuracy and the quality of time-segment based annotations. One of the remaining issues is to develop tools to generate results in line with the server processor and according to the capabilities of Internet browsers while managing the workflow.

6. CONCLUSION

The Telemeta open-source framework brings researchers in humanities and social sciences a new platform to efficiently distribute, share and work on their research on musical and sound materials. This platform offers automatic music analysis capabilities through the external component, TimeSide that provides a flexible computational analysis engine together with web serialization and visualization options. It brings an appropriate processing framework for researchers in computational ethnomusicology to develop and evaluate their algorithms. Deployed to manage the CNRS - Musée de l’Homme sound archives, the Telemeta platform has been conceived and adapted to generate tools in line with the needs of users.

Thanks to the collaborative nature of the platform, users can continuously enrich metadata associated with sound archives.

The benefits of this collaborative platform for the field of ethnomusicology apply to numerous aspects of research, ranging from musical analysis in a diachronic and synchronic

[17] http://catalogue.bnf.fr/ark:/12148/cb119367821/PUBLIC

comparative perspective, as well as the long-term preservation of sound archives and the support of teaching materials for education.

7. ACKNOWLEDGMENTS

The authors would like to thank all the people who have been involved in Telemeta specification and development or who have provided useful input and feedback. The project has been partly funded by the French National Centre for Scientific Research (CNRS), the French Ministry of Culture and Communication, the TGE Adonis Consortium, and the CREM.

8. ADDITIONAL AUTHORS

Additional authors:
Jean-Luc Rouas (LABRI, Bordeaux, France - email: jean-luc.rouas@labri.fr),
Julien Pinquier and Julie Mauclair (IRIT-UPS - Toulouse, France - email: {pinquier,mauclair@irit.fr}) and
Claude Barras (Université Paris-Sud / CNRS-LIMSI - Orsay, France - email: claude.barras@limsi.fr)

9. REFERENCES

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Blackwell, New York, USA, 1958.

[2] R. André-Obrecht. A new statistical approach for automatic speech segmentation. *IEEE Trans. on Audio, Speech, and Signal Processing*, 36(1):29–40, january 1988.

[3] C. Barras, X. Zhu, S. Meignier, and J. Gauvain. Multistage speaker diarization of broadcast news. *IEEE Trans. on Audio, Speech, and Language Processing*, 14(5):1505–1512, 2006.

[4] P. Brossier. *Automatic annotation of musical audio for interactive systems*. PhD thesis, Centre for Digital music, Queen Mary University of London, UK, 2006.

[5] C. Cannam, C. Landone, M. B. Sandler, and J. P. Bello. The sonic visualiser: A visualisation platform for semantic descriptors from musical signals. In *Proc. of the International Society for Music Information Retrieval Conference*, pages 324–327, 2006.

[6] A. De Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[7] G. Dournon. *The New Grove Handbook in Music*, chapter "Organology", Ethnomusicology, an Introduction, pages 245–300. Macmillan Press, dournon, geneviève edition, 1992.

[8] D. Fourer, J.-L. Rouas, P. Hanna, and M. Robine. Automatic timbre classification of ethnomusicological audio recordings. In *Proc. International Society for Music Information Retrieval Conference (ISMIR'2014)*, Taipei, Taiwan, Oct. 2014. Accepted for publication.

[9] G. Gravier, G. Adda, N. Paulson, M. Carré, A. Giraudel, O. Galibert, et al. The etape corpus for the evaluation of speech-based tv content processing in the french language. In *International Conference on Language Resources, Evaluation and Corpora*, 2012.

[10] E. Gómez, P. Herrera, and F. Gómez-Martin. Computational ethnomusicology: perspectives and challenges. *Journal of New Music Research*, 42(2):111–112, 2013.

[11] A. Julien Da Cruz Lima. The CNRS — Musée de l'Homme audio archives: a short introduction. *International Association of Sound and Audiovisual Archives journal*, 36, jan 2011.

[12] H. Lachambre, J. Pinquier, and R. André-Obrecht. Distinguishing monophonies from polyphonies using Weibull Bivariate distributions. *IEEE Trans. on Audio, Speech, and Language Processing*, 19(6):1837–1842, august 2011.

[13] B. Mathieu, S. Essid, T. Fillon, J. Prado, and G. Richard. Yaafe, an easy to use and efficient audio feature extraction software. In *Proc. of the International Society for Music Information Retrieval Conference, Utrecht, Netherlands*, pages 441–446, 2010.

[14] G. Peeters. Automatic classification of large musical instrument databases using hierarchical classifiers with intertia ratio maximization. In *115th convention of AES*, New York, USA, Oct. 2003.

[15] G. Peeters, B. Giordano, P. Susini, N. Misdariis, and S. McAdams. The timbre toolbox: Audio descriptors of musical signals. *Journal of the Acoustical Society of America*, 5(130):2902–2916, Nov. 2011.

[16] J. Pinquier, J.-L. Rouas, and R. André-Obrecht. A fusion study in speech / music classification. In *IEEE International Conference on Audio, Speech and Signal Processing, Hong-Kong, China*, april 2003.

[17] J. Simonnot. TELEMETA: an audio content management system for the web. *International Association of Sound and Audiovisual Archives journal*, 36, jan 2011.

[18] J. Simonnot, M.-F. Mifune, and J. Lambert. Telemeta: Resources of an online archive of ethnomusicological recordings. Panel presented at ICTM Study Group on Historical Sources of Traditional Music, Aveiro, Portugal, May 12-17 2014, 2014.

[19] H. T. and J. M. Steeneken. A Review of the MTF Concept in Room Acoustics and its Use for Estimating Speech Intelligibility in Auditoria. *Journal of the Acoustical Society of America*, 77(3):1069–1077, 1985.

[20] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, 2007.

[21] G. Urban. Ritual Wailing in Amerindian Brazil. *American Anthropologist*, 90(2):385–400, 1988.

[22] E. v. Hornbostel and C. Sachs. Systematik der musikinstrumente. *Zeitschrift für Ethnologie*, 46:553–590, 1914.

[23] E. v. Hornbostel and C. Sachs. The classification of musical instruments. *Galpin Society Journal*, 3(25):3–29, 1961.