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 *CNRS-Musée de l'Homme*'s audio archives are among the most important collections of ethnomusicological recordings in Europe. Yet, as it continuously expands and as new audio technologies develop, questions linked to the preservation, the archiving and the availability of these audio materials have arisen. Since 2007, ethnomusicologists and engineers have thus joined their efforts to develop a scalable and collaborative web platform for managing and increasing access to digitalised 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 produces various automatic annotation, segmentation and musicological analysis as it includes a set of audio analysis plug-ins and wraps several audio feature extractions libraries that have been developed in the interdisciplinary research project called DIADEMS. In this paper, we will introduce 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.

*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*

1. INTRODUCTION

In social sciences, as very large scientific databases become available and rapidly increase by both numbers 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 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 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* [19, 9]. 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 [16, 10, 17]. 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 annotations 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 1. Its flexible and streaming safe architecture is represented in Figure 2. The main features of *Telemeta* are:

- Pure HTML5 web user interface including dynamical 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 (now english and french)

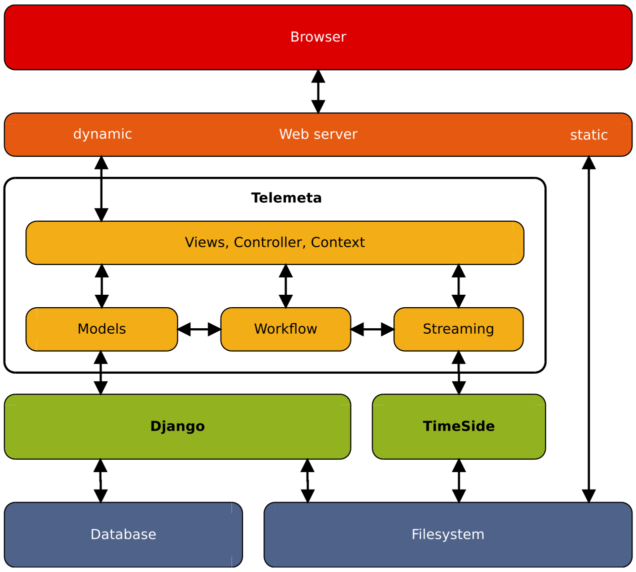

Figure 2: Telemeta architecture

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 enrichment of the materials in the database. One of the major challenges is thus 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.

[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

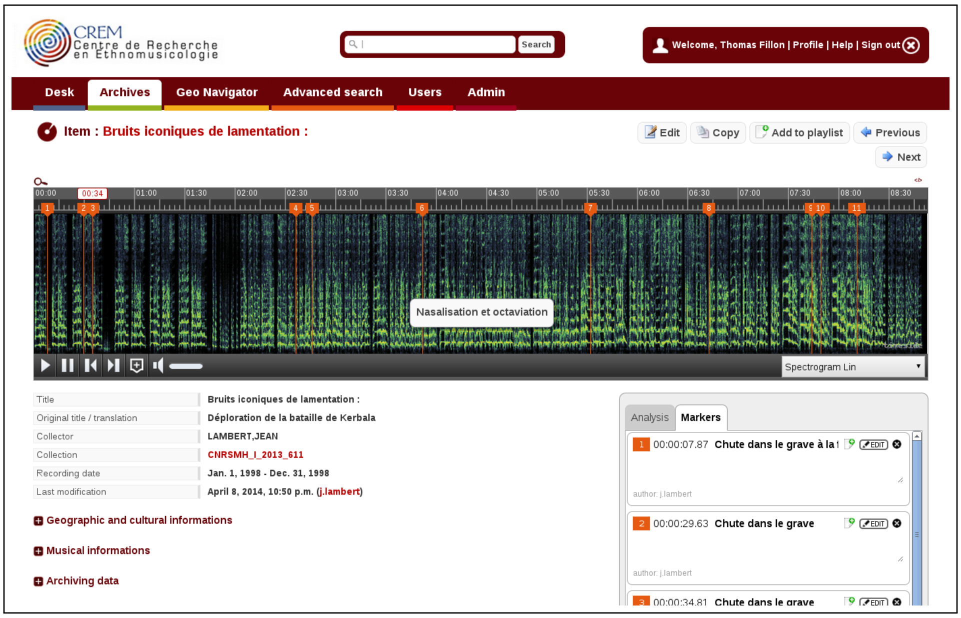

Figure 1: Screenshot excerpt of the Telemeta web interface

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. 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.

As illustrated in Figure 3, 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 that can be combined with visual representations. 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 on Figure 2, 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.

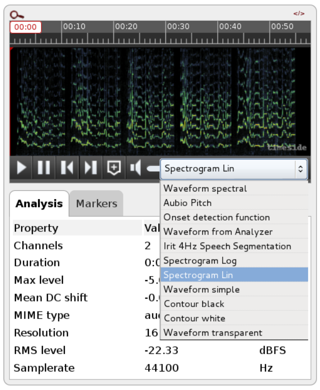

Figure 3: Selection of various sound representation.

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

One specificity of the Telemeta architecture is to rely on an external component, TimeSide[5], that offers audio player

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

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 4 illustrates the overall architecture of TimeSide together with the data flow between TimeSide and the Telemeta web-server.

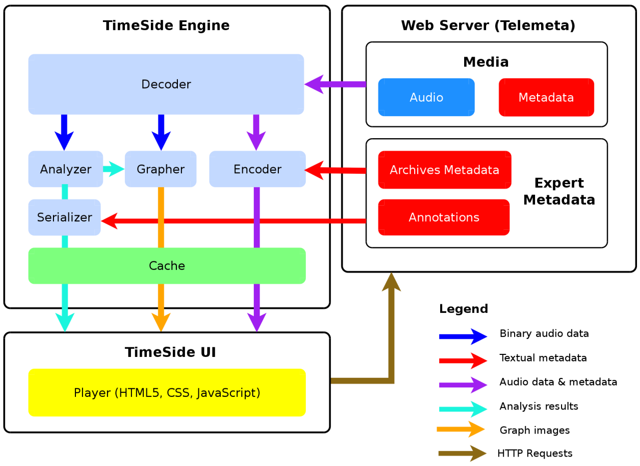

Figure 4: TimeSide engine architecture and data flow with Telemeta web-server

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 visualisation (waveform, spectrogram)
- Multi-format support: reads all available audio and video formats through Gstreamer, transcoding 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 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] [12] and Vamp plugins[8]. As a open-source framework and given its architecture and the flexibility provided by Python, the implementation of any audio and music analysis algorithm can be considered. Thus, it makes it a very convenient framework for researchers in computational ethnomusicology to develop and evaluate their algorithms. 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 like Numpy 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.

[6]http://aubio.org/
[7]https://github.com/Yaafe/Yaafe
[8]http://www.vamp-plugins.org

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 (Research Center for Ethnomusicology). 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 have 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 comes 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 many 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 in Mai 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 had 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

[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/

some use of the platform to foster and/or deepen individual research. The unexpectedly broad uses of the sound archives once digitalised and accessible emphasise the necessity and the benefits of such database. From an archive stand, 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 to 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, others' material or both. When use for education, the platform provides a wide array of teaching material to illustrate students' work as well as support teaching curricula.

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 program whose consortium includes research laboratories from both *Science and Technology of Information and Communication*[13] domain and *Musicology and Ethnomusicology*[14] domain and Parisson, a company involved in the development of Telemeta.

The goal of Diadems project is to develop computer tools to automatically index the recording content directly from the audio signal to improve the access and indexation of this vast ethnomusicological archive. Numerous ethnomusicological recordings contain speech and other types of sounds that we categorized as sounds from the environment (such as rain, insect or animal sounds, engine noise and so on) and sounds generated by the recording (such as sound produced by the wind in 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 their content, 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 most exhaustive annotation of the audio materials.

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

[13] IRIT (Institute of research in computing science of Toulouse), LABRI (Bordeaux laboratory of research in computer science), 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)

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. Automatic analysis of these recordings requires methods having a stronger robustness. Preliminary Implementations of speech detection models, and speaker diarisation 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 these methods to the particular case of ethnographic archives.

In the context of this project, both researchers from Ethnomusicological, Speech and Music Information Retrieval 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 on speech and Music Information Retrieval (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 also useful tools for musical analysis such as tools for 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 follow 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 man-led (human) evaluations carried on the corpus selected,

[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 [20], "icons of crying" include cry break, voice inhalation, creaky voice and falsetto vowels.

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

–> TF, section à relire.

Il faut également changer la taille et le contenu des figures après mise à jour diadems.telemta.org ou directement depuis TimeSide

At first, a primary component annotation in Speech, Music and Singing Voice is consolidated. Speech and music detection is generally studied in rather different contexts – usually broadcast data. The diversity of available recordings implies the introduction of new acoustic features and a priori knowledge coming from content descriptions. Singing voice detection is emerging and a genuine research effort is needed. Exploiting the complementarity of the three approaches – speech, music, singing voice – will provide robust means to detect other components, such as speech, speech over music, overlap speech, instrumental music, a cappella voice, singing voice with music, etc.

5.2.1 *Analysis of recordings sessions*

- Detection of start and stop of the magnetic recorder
- Detection of start and stop of the mecanic recorder
- Detection of start and stop of the digital recorder

–> IRIT : Insérer Description de "irit-noise-startSilences" ?

5.2.2 *Analysis of speech and singing voice segments*

- Speech detection in a recording with music and speech alternation
- Speech detection in a recording with one speaker and group of speaker
- Speech detection in a recording with speaker and music mixed
- Segmentation in speaker turns for speech or singing voice
- Analysis of the syllabic flow and the prosody of the speech in a ritual context
- Detection of speakers overlap
- Detection of other categories of speaking voice: recitation, told, psalmody, back channel. . .

See figure 5

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 [18]. 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 [15]. We first evaluate the signal entropy ($H = \sum_{i=1}^{k} -p_i log_2 p_i$, with p_i =proba. 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.

–> LIMSI : Insérer Description de "LIMSI SAD" ?

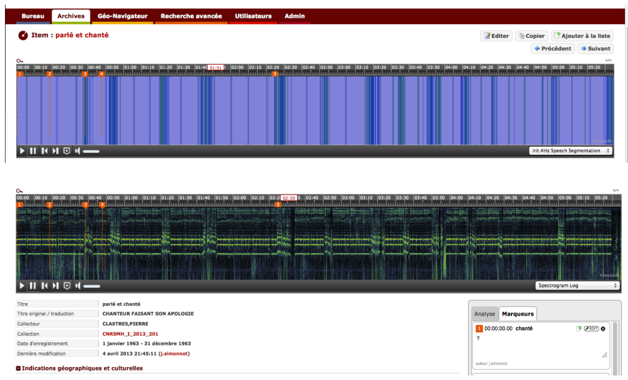

Figure 5: Detection of spoken voices in a song

5.2.3 *Analysis of music segments*

DIADEMS wishes to provide useful tools for musical analysis musicologists and music teachers:

- Detection of instrumental music
- Musical instrument family recognition
- Analysis of musical information (tonal, metric and rhythmic features)
- Musical excerpts characterisation using similarity measures (melody, harmony, rhythm,)
- Navigation inside a document using a structural analysis (chorus localisation, 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 by a string 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 unites 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 segments obtained with our segmentation algorithm are generally longer for music than for speech. We chose to model the segment duration by a Gaussian Inverse law (Wald law) which is indeed our second feature providing a 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 is, 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 [11].

See figure 6

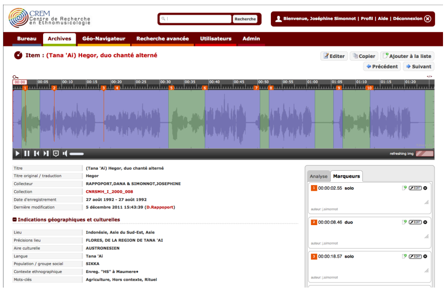

Figure 6: Detection solo and duo parts

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 [21] and later translated in [22]. This system is the most widely used system for classifying musical instruments by ethnomusicologists and organologists. together with 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.

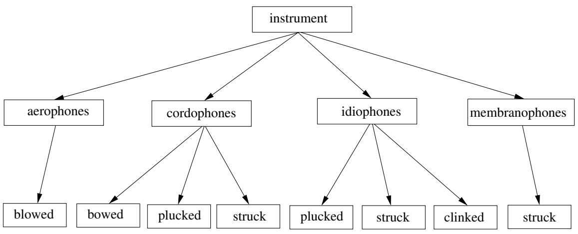

Figure 7: Musical instrument families

The proposed method is based on the supervised learning approach and uses a set of 164 acoustic descriptors proposed by Peeters *et al.* in [14]. This system (see Figure 8) applies the Inertia Ratio Maximization Features Space (IRMFSP) algorihtm [13] 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.

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

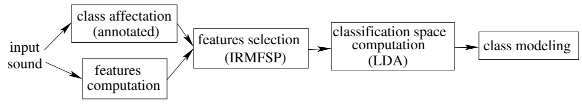

Figure 8: Training step of the proposed classification method.

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).

5.3 Preliminary results and evaluation protocol

–> Thomas : Comme on a pas de résultats quantitatif à montrer il faut peut-être revoir l'intitulé de cette sous-section

At the end of the first step of the project, interesting preliminary results have been obtained regarding sessions recordings, speech recognition, singing voice recognition and also musical instrument family classification.

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. Results visualised thanks to the Telemeta platform will be evaluated by the humanities community involved, through a collaborative work on line. One of the issues is to develop tools to generate results on line with the server processor, according to the capacities of the internet navigators. An other challenge is to manage the workflow, according to the users' expectations about their annotations. The validated tools will be integrated with a new design in the platform Telemeta.

Automatic instrument classification.

As shown in Figure 9, 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.

6. CONCLUSION

The Telemeta open-source framework provides the researchers in musicology with a new platform to efficiently distribute, share and work on their research materials. The platform has been deployed since 2011 to manage the *Sound archives of the CNRS - Musée de l'Homme* which is the most important european collection of ethnomusicocological resources. Furthermore, this platform is offered automatic music analysis capabilities through an external component, TimeSide that provides a flexible computational analysis engine together with web serialization and visualization capabilities. As an open-source framework TimeSide could be an appropriate platform for researchers in computational ethnomusi-

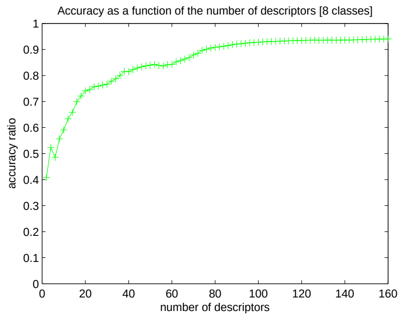

Figure 9: 3-fold cross-validation classification accuracy as a function of the number of optimally selected descriptors.

cology to develop and evaluate their algorithms. The benefits of this collaborative platform for the field of ethnomusicology apply to numerous aspects, ranging from musical analysis in a diachronic and synchronic comparative perspective, as well as long-term preservation of sound archives and support teaching material for education. Thanks to the collaborative nature of the platform, users can cooperate to continuously enrich metadata asso-ciated to sound archives.

7. ACKNOWLEDGMENTS

The authors would like to thank all the people that have been involved in Telemeta specification and development or have provide useful input and feedback. The project has been partially funded by the French National Centre for Scientific Research (CNRS), the French Ministry of Culture and Communication, the TGE Adonis Consortium, and the Centre of Research in Ethnomusicology (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 Transactions 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. *Audio, Speech, and Language Processing, IEEE Transactions on*, 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 *ISMIR*, 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. Submitted for publication.

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

[10] 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.

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

[12] 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 ISMIR 2010, Utrecht, Netherlands*, pages 441–446. International Society for Music Information Retrieval, 2010.

[13] 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.

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

[15] 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.

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

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

[18] 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.

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

[20] G. Urban. Ritual wailing in amerindian brazil. *American Anthropologist*, 90(2):385–400, 1988.

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

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