



**HAL**  
open science

# SYSTÈME DE CAPTATION OPTIQUE POUR LA TRANSCRIPTION AUTOMATIQUE DE LA MUSIQUE DE CITHARE MALGACHE MAROVANY

Dorian Cazau, Olivier Adam, Marc Chemillier

► **To cite this version:**

Dorian Cazau, Olivier Adam, Marc Chemillier. SYSTÈME DE CAPTATION OPTIQUE POUR LA TRANSCRIPTION AUTOMATIQUE DE LA MUSIQUE DE CITHARE MALGACHE MAROVANY. Journées d'Informatique Musicale, May 2013, Paris, France. hal-03112215

**HAL Id: hal-03112215**

**<https://hal.science/hal-03112215v1>**

Submitted on 16 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SYSTÈME DE CAPTATION OPTIQUE POUR LA TRANSCRIPTION AUTOMATIQUE DE LA MUSIQUE DE CITHARE MALGACHE *MAROVANY*

*Dorian Cazau, Olivier Adam*  
Laboratoire d'Acoustique Musicale (LAM)  
UPMC, 11 rue de Lourmel - 75015 PARIS  
cazau@lam.jussieu.fr

*Marc Chemillier*  
Centre d'analyse et de mathématiques sociales  
EHESS, 198 avenue de France - 75013 Paris  
chemilli@ehess.fr

## RÉSUMÉ

Dans cet article, nous introduisons un système de captation optique original dédié à l'analyse musicale de la cithare malgache. Ce système a été appliqué à l'extraction automatique d'information musicale, assistée par des méthodes de pré-traitement pour la détection et la caractérisation acoustique des notes de musique. L'objectif visé est de réaliser une transcription systématique du répertoire musical de la cithare *marovany* en contexte de transe *tromba*, et de classifier les airs musicaux récurrents en terme de devises identificatoires. A travers ce travail nous recommandons pour remplir cet objectif l'utilisation d'un enregistreur multicanal à capteurs optiques synchronisés, en avançant les qualités technologiques suivantes : captation individuelle des cordes, fort rapport signal sur bruit (forte sensibilité au déplacement des cordes / insensibilité aux sources acoustiques externes), démarcation systématique des notes par des silences issues du contact doigt-corde. De telles caractéristiques simplifient grandement la tâche délicate de transcription automatique de musique polyphonique en milieu bruité.

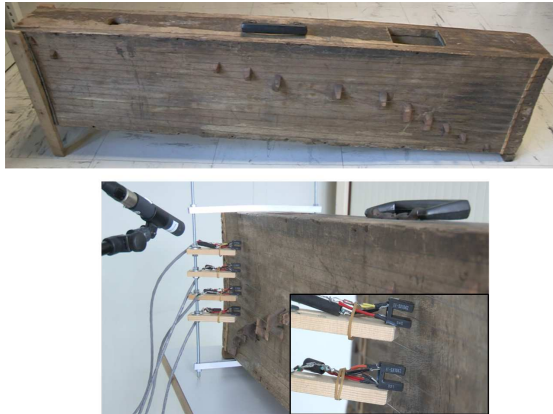
## 1. INTRODUCTION

La *marovany* est une grande cithare montée sur une caisse rectangulaire en bois de récupération. Ses cordes métalliques en câbles de freins de moto peuvent mesurer jusqu'à 1 m 20, et sont tendues des deux côtés de la caisse. Elles sont ensuite clouées sur celle-ci, reposant à chaque extrémité sur un chevalet (en bois ou en métal), et sont surélevées par des tasseaux de bois placés à différents endroits suivant la hauteur que l'on veut donner à la corde. Elles forment, comme pour la cithare tubulaire (*valiha*), une échelle diatonique alternée. D'un point de vue musical, le répertoire joué à la cithare *marovany* se veut comme un flot mélodique arpégé. Bien qu'il n'y ait pas de véritable écriture verticale dans son répertoire, à l'exception de quelques accords (c'est-à-dire au minimum deux notes simultanées) ponctuels, les notes sont jouées très rapidement, sont rarement étouffées à l'intérieur d'une phrase musicale et les cordes résonnent par sympathie les unes avec les autres. Ces caractéristiques font qu'un certain nombre de notes se retrouvent très sou-

vent en simultanéité et confèrent à cette musique une complexité d'analyse digne d'une musique polyphonique. Le type de bois, les dimensions et le nombre de cordes utilisés ne sont pas fixés. On peut en effet trouver des cithares en bois léger ou dur, de longueur pouvant varier de 1 à 2 mètres de long, possédant 8, 10, 12 cordes sur des tasseaux allant de 2 cm de haut à 0.5 cm (des tasseaux plus étroits sont connus pour donner plus de son car les cordes sont plus près de la caisse, sonnante *mafo be*, « plus fort »). La cithare étudiée ici (photo à la figure 1) compte 13 cordes sur chacune des deux faces, avec une tessiture couvrant plus de deux octaves.

En plus de ses qualités purement musicales, l'étude de la cithare sur caisse *marovany* présente un intérêt majeur du point de vue anthropologique par le fait que sa musique est utilisée dans le culte de possession *tromba*. La transe est induite musicalement, en ce sens qu'elle est provoquée et entretenue par des stimuli musicaux. En contexte de *tromba*, le mode de jeu de la *marovany* est constitué essentiellement de motifs mélodiques, enchaînés et transformés progressivement à travers de multiples répétitions obsédantes. L'instrument est souvent accompagné d'un hochet dit *kantsa*, formé d'une boîte de conserve remplie de graines et clouée sur un manche en bois, fournissant le support rythmique de cette musique. Cette fonction sociale de la *marovany* détermine probablement son répertoire musical. En effet, les musiques de possession prennent souvent la forme de devises identificatoires [1], associant une formule musicale à une entité symbolique extramusicale. Une partie au moins du répertoire de la *marovany* pourrait ainsi se présenter comme une table d'associations et de correspondances entre formules musicales et certaines divinités [2]. Un autre aspect de sa fonction dans le culte de possession *tromba* est qu'elle participe à l'effervescence collective propice au déclenchement de la transe, en contribuant à faire « chauffer » l'atmosphère. Le moyen musical de cet échauffement est peut-être l'utilisation de formules musicales jouées selon une progression judicieusement contrôlée par le musicien. Afin de mieux comprendre ces deux aspects du lien de la cithare avec la transe, une analyse de son répertoire doit être réalisée, avec un inventaire systématique des formules musicales, si possible en relation avec des indices comportementaux, et des associations air/divinité. L'étude de ce répertoire spé-

cifique, fondée sur des critères musicaux, permettra d'apporter des éléments inédits à la question fascinante des relations entre la musique et la transe. Dans une perspective plus large, les études de la transe qui abordent la question de fond des mécanismes neurophysiologiques [3, 4], bénéficieraient des données musicales extrêmement précises et complètes fournies par nos capteurs.



**Figure 1.** Photo de la cithare malgache (en haut) et du dispositif de captation optique (en bas)

Les cithares malgaches, en particulier la *valiha*, considérée comme l'instrument national et de laquelle dérive la *marovany*, ont fait l'objet de nombreuses recherches [5, 6] d'ordre ethnomusicologique. Cependant, il n'existe pas à notre connaissance de projet d'analyses et de classifications systématiques du répertoire des airs musicaux de la cithare *marovany*, en fonction de leurs propriétés musicales (rythmiques, structurelles, modales et acoustiques) et extra-musicales (contenu sémantique et symbolique). Dans le but de fournir un aperçu significatif de la relation transe-musique à Madagascar, des investigations de terrain doivent être entreprises, avec l'aide de systèmes automatisés pour l'extraction et la caractérisation de patterns musicaux. Ces données traitées par l'informatique pourront ensuite servir à évaluer statistiquement certaines occurrences musicales en fonction de différentes séances de transe, et de dresser des corrélations entre des patterns musicaux et comportementaux au cours d'une séance de transe (par exemple, la manière dont un musicien injecte des indices de renouvellement musicaux pour relancer l'intérêt ou intensifier la danse). Une étude systématique de ces concordances permettrait d'établir le catalogue raisonné du répertoire commun aux joueurs de *marovany* en contexte de transe. L'automatisation de l'analyse de cette musique est rendue impérieuse par l'aspect laborieux que la tâche manuelle de transcription représente, sachant que des séances de transe peuvent durer plusieurs heures et qu'il n'existe aucun support écrit de musique. Aussi, la complexité de sa transcription (vitesse de jeu, caractère polyphonique, milieu bruité par des sources acoustiques) peut entraîner une grande variabilité dans les résultats, sans estimation possible de la qualité de la transcription. Durant les trances étudiées, ces transcriptions optiques sont accompagnées d'enregistrements audiovisuels

(des exemples de vidéo et d'extraits audio peuvent être trouvés sur la page web [7]), qui permettent l'analyse parallèle des indices comportementaux rapportés ci-dessus, mais ne constituent pas un support optimal pour la transcription musicale car comportant plusieurs sources musicales compétitives dans un environnement bruyant. Pour répondre à cette problématique, ce papier présente en section 2 un système de captation optique dédié à l'enregistrement in situ d'airs musicaux de la *marovany*. Ce système de captation a ensuite été intégré à une chaîne d'acquisition et de traitements visant à l'extraction automatique d'informations musicales, présentée en section 3.

## 2. SYSTÈME DE CAPTATION OPTIQUE

Plusieurs contraintes ont du être appréhendées dans le choix du système d'enregistrement. Des contraintes intrinsèques à la cithare *marovany* d'abord, incluant la rapidité de jeu, les différents modes d'attaque et d'étouffement des cordes, les segments occasionnels de polyphonie (accord) et les résonances par sympathie des cordes. L'accumulation de ces contraintes génèrent des signaux audio complexes à analyser. S'ajoutent à cela des contraintes externes, comme les autres sources acoustiques (principalement le hochet, les battements de l'assistance, les interjections vocales du possédé), les conditions environnementales (fortes humidité et chaleur) et techniques (batteries électriques rares et fragiles). Il est préférable d'éliminer les systèmes trop sensibles et pré-amplifiés gourmands en énergie (comme l'usage d'une alimentation fantôme 48V) qui pourraient se dégrader plus vite à cause de l'humidité et de la chaleur. L'ensemble de ces contraintes a été pris en compte dans le développement du système par captation optique présenté ci-dessous.

Les systèmes de captation optique ont déjà trouvé des applications dans le domaine de l'audio, comme la réalisation de mesures métrologiques de déplacements de cordes [8, 9] ou la MIDIfication<sup>1</sup> d'un piano à travers la technologie piano-bar développée par Moog [10, 11]. Notre système de captation, illustré à la figure 1, se rapproche de cette seconde application, bien que s'en démarquant aussi par la volonté d'intégrer aussi fidèlement que possible un grand nombre de paramètres physiques caractérisant le rendu acoustique de l'instrument. Le capteur optique choisi se présente sous la forme d'une fourche, illustré dans la zone agrandie de la figure 1. D'un côté, une DEL (Diode ElectroLuminescente) émet un faisceau lumineux de 0.5 mm de diamètre dans l'infrarouge. Son pic d'émission se situe autour de la longueur d'onde 940 nm. De l'autre, un phototransistor adapte son courant de sortie en fonction du flux lumineux qu'il reçoit. Son pic de sensibilité est à la longueur d'onde 850 nm. Ainsi, lorsque la corde, placée dans la fourche optique, passe au travers du rayon lumineux, elle bloque une partie du flux lumineux et module donc le courant de sortie du capteur. Pour obtenir une surface sensible supérieure au seul diamètre du

<sup>1</sup> . Néologisme signifiant la conversion in situ d'un instrument acoustique en son homologue MIDI.

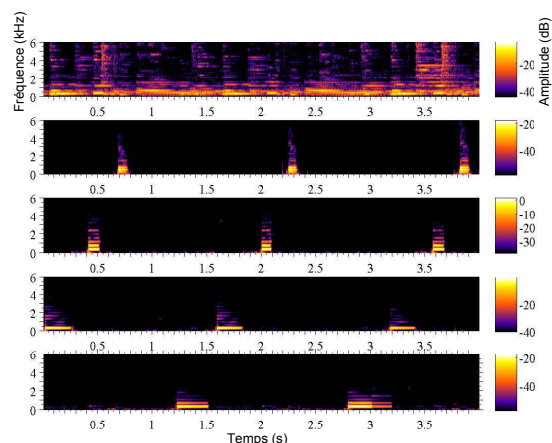
faisceau lumineux (0.5 mm), deux fourches optiques sont collées l'une à l'autre avec un léger décalage pour donner une surface couverte de 3.5 mm, puis mises en phases par des circuits de conditionnement électroniques. Le boîtier d'alimentation des fourches ne nécessite qu'une tension continue de 5V, et est isolé thermiquement, ce qui le rend bien adapté pour les conditions de terrain. Deux enregistreurs numériques portables ZOOM R16, permettant l'enregistrement de 2 x 8 pistes simultanément, servent enfin à l'acquisition des signaux optiques. Ceux-ci sont acquis par le logiciel Cubase (asservi par un ordinateur portable emmené sur le terrain). Un signal audio de référence est aussi enregistré simultanément avec un microphone Neumann KM 184 mt. La fréquence d'échantillonnage de tous les enregistrements est de 44.1 kHz, avec 16 bits pour la numérisation.

Chaque corde est ainsi équipée d'un tel capteur optique, qui sont tous fixés sur une barre commune, verticale et extérieure à l'instrument (voir photo 1). Deux contraintes ont été prises en compte pour le placement de ces capteurs. Il a fallu d'une part éviter que le système soit trop intrusif et altère la jouabilité de l'instrument<sup>2</sup>. Et d'autre part, le point de mesure de déplacement de la corde peut biaiser la mesure de son amplitude, puisque ce déplacement se compose d'une superposition de modes vibratoires, faits de ventres et de noeuds, le fait de placer un capteur en un noeud modal revient à supprimer la contribution énergétique de ce mode au déplacement global. Pour répondre à ces deux contraintes, la barre est installée au plus près du chevalet, pour qu'ainsi la zone de jeu soit moins dérangée, et que les capteurs soient placés approximativement sur la pente ascendante du ventre suivant directement le noeud du chevalet, qui est commun à tous les modes.

La figure 2 représente les spectrogrammes du signal audio et des signaux optiques (après pré-traitement, voir section 3) enregistrés sur un air *Sojerina*. Les signaux optiques offrent un fort rapport signal sur bruit, ainsi que des transitoires d'attaque raides. L'indépendance de chaque corde est bien respectée, chaque capteur détectant uniquement la vibration de la corde sur lequel il est placé. On peut aussi remarquer la bonne séparabilité des notes successives sur une corde, les blancs inter-notes correspondant aux instants de contact entre le doigt et la corde. En définitive un tel système de captation revient à décomposer un signal audio multi-sources complexe en ses composantes simples, ramenant plus particulièrement le traitement d'une séquence polyphonique en celui de plusieurs séquences monophoniques conjointes.

Nous présentons maintenant une courte étude comparative des enveloppes temporelles entre les signaux optiques et audio. Après avoir pré-traité les signaux (voir section 3), cinq descripteurs acoustiques classiques ont été calculés sur un ensemble de 30 paires de notes {audio;optique}, jouées isolément et laissées en résonance

2. Une première appréciation positive a déjà été donnée par un musicien malgache sur ce sujet [12], mais une prochaine version du système visera notamment une meilleure compacité.



**Figure 2.** Exemple de spectrogrammes du signal audio (graphe du haut) et de sa décomposition en quatre signaux optiques (graphes dessous), extrait d'un air *Midégana*

libre. Les paires ont été sélectionnées aléatoirement pour leur diversité. Les descripteurs traités sont définis comme suit :

**AT** , le temps d'attaque (en s) est défini par le temps nécessaire pour le signal d'atteindre 95 % de son énergie maximale  $E_{max}$

$$s(n = AT) = 0.95E_{max} \quad (1)$$

**D** , la durée physique du signal (en s) sera définie par le temps durant lequel l'énergie du signal est compris entre 5 % et 95 % de son énergie maximale  $E_{max}$

$$D = \{n/s(n) > 0.05E_{max} \ \& \ s(n) < 0.95E_{max}\} \quad (2)$$

**E** , le niveau énergétique rms (en Pa)d'un signal défini par

$$E(k) = \sqrt{\frac{1}{K} \sum_{k=1}^K |(x + kN)|^2} \quad (3)$$

calculé pour K fenêtres de N échantillons ;

**DH** , l'indicateur d'harmonicit  (adimensionn ) renseigne sur le contenu harmonique d'un signal. Il consiste   balayer la densit  spectrale X d'un signal avec un filtre en peigne Filt, dont la fr quence fondamentale varie dans une plage donn e (ici [140 ; 400] Hz , ce qui est grossi rement la tessiture de la cithare) [13]. Lorsque les vall es de ce filtre coïncident avec les pics du filtre, leur produit est minimal. Math matiquement, on a alors

$$DH = \min\left(\frac{E_{pond}}{E_{init}}\right) \quad (4)$$

avec  $E_{init} = \sum |X(k)|^2$  and  $E_{pond} = Filt(k, k_0)$   
 $E_{init}$ , avec Filt d fini par  $Filt = 2(1 - |\cos(\frac{\pi F}{F_0})|)$ .

Descripteurs	At	AD	dB	DH
$E(\Delta_D)$	0.0129	0.4366	0.23	0.26
$\sigma(\Delta_D)$	0.0028	0.11	0.15	0.09

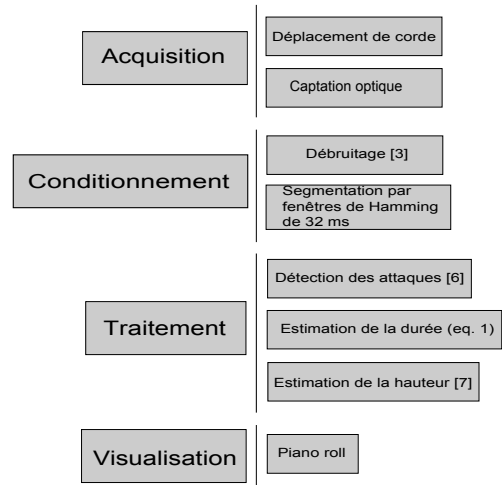
**Table 1.** Moyenne E et écart-type  $\sigma$  des différences acoustiques absolues entre des signaux de types optiques et audio pour les descripteurs acoustiques At, AD, dB, DH

Les descripteurs E et DH sont évalués relativement à une référence arbitraire respective au type optique ou audio. Les différences acoustiques absolues entre ces deux types de signaux sont simplement quantifiées avec l'opérateur  $\Delta_D = |D_{Audio} - D_{Optique}|$ , où D représente un descripteur acoustique donné. Le tableau 1 présente les résultats de cet opérateur, montrant que la distorsion d'information est minimale si on se restreint aux informations sur l'enveloppe temporelle. Par contre, le contenu spectral montre des différences plus significatives. Une tendance intéressante étant notamment que le contenu harmonique des signaux optiques est plus fort, ce qui peut s'expliquer par le fait qu'une mesure directe de déplacement de corde privilégie sa fréquence fondamentale et ses harmoniques dans le comportement vibratoire observé, en minimisant l'effet de couplage aux modes plus complexes de la table d'harmonie. La différence sur l'amplitude peut être significative si la vibration de la corde excite fortement certains modes de la table, générant des sons plus entretenus à travers un transfert d'énergie optimal des cordes vers la table.

### 3. APPLICATION À L'EXTRACTION AUTOMATIQUE D'INFORMATIONS MUSICALES

Généralement, la recherche d'information pour la transcription musicale peut se décomposer en plusieurs niveaux : les informations de bas niveau (la hauteur, l'attaque et la durée des notes), et les informations de haut niveau qui font appel à des notions plus globales de la musique (tempo, chiffrage de mesure, tonalité, reconnaissance d'instrument). Notre système de transcription comprend un dispositif d'acquisition de données (décrit ci-dessus) pour enregistrer la cithare et une partie logicielle intégrant un algorithme d'analyse qui déterminera la durée et la hauteur des notes jouées, qui sont les deux informations de bas niveau essentielles permettant d'écrire une partition, auxquelles on ajoutera leurs amplitudes. Ces informations seront ensuite regroupées pour ne faire qu'un fichier au format MIDI. Ce fichier pourra ainsi être lu et édité sur n'importe quel séquenceur audio ou logiciel de partition.

Une fois les signaux optiques correctement acquis, leur transcription ne présente pas de difficulté particulière. La figure 3 représente un diagramme-bloc des différentes fonctions constituant notre chaîne de traitement, de l'acquisition des signaux optiques à l'algorithme de détection et de caractérisation des notes de la *marovany*. Le pré-traitement des signaux consiste en les étapes suivantes. Pour des soucis de mémoire informatique, des séquences de 5s sont d'abord importées dans le logiciel Matlab. Un



**Figure 3.** Diagramme-bloc des fonctions de l'algorithme de détection et de caractérisation acoustique des notes

filtrage adaptatif [14] est appliqué pour optimiser le rapport signal sur bruit, celui-ci étant notamment détérioré par des bruits parasites d'origine électronique ou de résonance par sympathie des cordes<sup>3</sup>. En rentrant des segments de signaux dont les propriétés spectrales sont représentatives de ces bruits, l'algorithme [14] permet de les soustraire au signal utile en réduisant le résidu par moindres carrés. Une fenêtre de Hamming glissante de 512 échantillons (soit 11.6 ms, donnant une résolution temporelle de l'ordre de grandeur des temps d'attaque) avec une superposition de 256 échantillons balaye ensuite l'ensemble de la séquence. Les fronts d'attaque de chaque note sont détectés en utilisant une différence spectrale, avec prise en compte de l'incrément de phase, telle qu'introduit par Bello et al. [15] :

$$\hat{X}_{k,n} = |X_{k,n-1}| e^{j(2\phi_{k,n-1} - \phi_{k,n-2})} \quad (5)$$

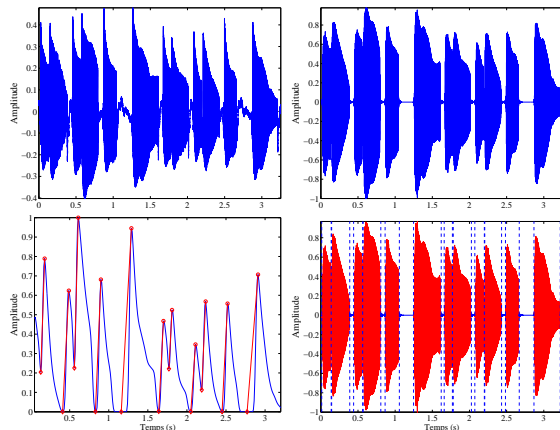
avec n l'indice de chaque fenêtre. Notre signal utile étant composé de sinusoides stationnaires, l'apparition d'un front d'attaque génère un pic dans l'erreur de prédiction définie par

$$r(n) = \sum_{n=1}^N |\hat{X}_{k,n} - X_{k,n}| \quad (6)$$

Un simple seuil est ensuite associé à ce résidu pour la validation d'une détection de notes ou non. A partir de cet onset détecté, le descripteur E (eq. 3) est calculé pour les fenêtres avoisinantes de façon à chercher le maximum local  $E_{max}(i)$  associé à la note i, en posant l'hypothèse que ce maximum est localisé près du front d'attaque, ce qui est vraisemblable pour les notes des instruments à cordes pincées. Ensuite, E est calculé sur l'ensemble des fenêtres qui suivent l'onset jusqu'à ce que sa valeur décroisse en-dessous de 5 % de  $E_{max}(i)$ , utilisée alors comme un seuil

<sup>3</sup>. La résonance par sympathie des cordes est en effet considérée dans notre situation de transcription bas niveau comme du bruit, bien que celui-ci prenne une place importante dans la constitution du timbre de l'instrument.

adaptatif d'énergie physique spécifique à la note  $i$ . Cette estimation nous permet de déduire la durée du signal (éq. 2) et l'amplitude par moyennage de l'énergie sur l'ensemble des fenêtres du signal. Une fois les notes détectées, la hauteur du signal à l'intérieur de cette fenêtre est enfin estimée par un algorithme robuste dérivé de la fonction d'auto-corrélation [16]. La figure 4 présente l'évolution de formes temporelles du signal à travers les différentes étapes de traitement.

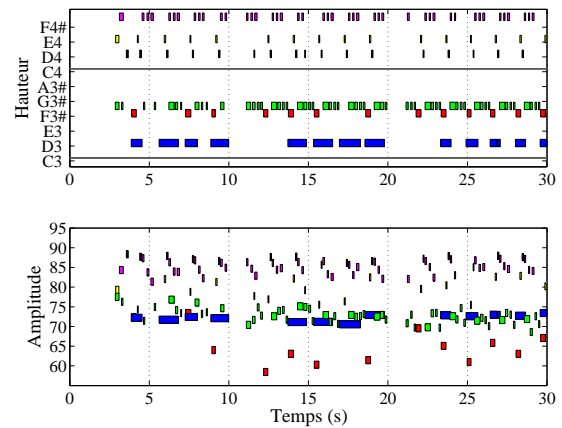


**Figure 4.** Evolution de formes temporelles du signal à travers les différentes étapes de traitement, avec de haut en bas et de gauche à droite : signal optique original, signal débruité, résidu  $r$  avec localisation des attaques, signal segmenté

Cet algorithme fut évalué sur des séquences préalablement étiquetées à la main, en reportant en plus de leur emplacement temporel absolu, leur durée, leur amplitude moyenne et leur hauteur. Une application brute de l'algorithme précédent aux séquences de type audio aboutit à des performances systématiquement inférieures à 55 % de détection de notes, alors que les séquences optiques donnent des résultats plus que satisfaisants (la tolérance fixée dans la détection des fronts d'attaque et de la durée étant respectivement de 32 ms et 0.5 s).

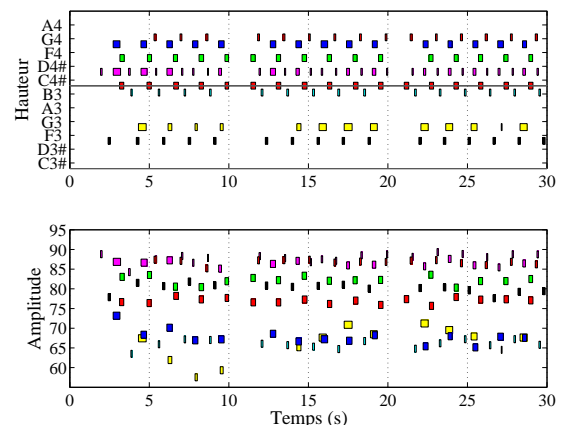
Nous proposons maintenant deux illustrations de cette méthode à travers les transcriptions automatiques de deux airs traditionnels joués à la cithare *marovany*, *Midegana* et *Sojerina* (les enregistrements sonores sont disponibles sur la page web [7]). Ces airs ont été enregistrés à Madagascar et joués par le musicien Velonjoro. Des problèmes techniques sur place ont rendu inutilisables plusieurs capteurs<sup>4</sup>, et chaque variante des airs a été partiellement reconstituée, avec plusieurs prises successives permettant d'enregistrer plusieurs séries de cordes. Malgré un recalage manuel et la superposition de prises distinctes, les signaux ont été jugés satisfaisants à l'oreille, la vitalité rythmique du jeu de Velonjoro étant bien préservée. Les figures 5 et 6 montrent les piano rolls de variantes de ces

4. Cette mission durant laquelle ces enregistrements ont été réalisés avait pour but de tester une version prototypique de notre système de captation [12]. La prochaine mission prévue pour Juin 2013 bénéficiera d'une version finalisée et entièrement opérationnelle de ce dispositif.



**Figure 5.** Piano roll résultant d'une transcription automatique d'un air *Midegana*

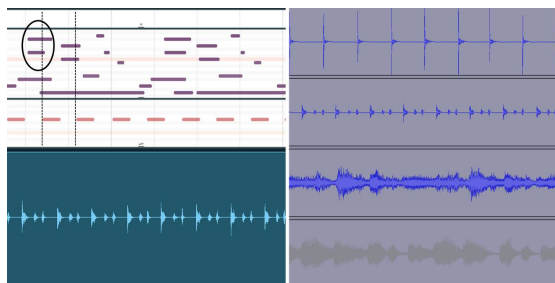
airs. Le piano roll est un moyen de représenter graphiquement un fichier MIDI. En ordonnée, il y a les différentes notes, représentées par des rectangles soit à travers leurs hauteurs (graphe du haut) ou leurs vélocités (graphe du bas) respectives, et en abscisse l'axe du temps. Chaque rectangle tracé sur le piano roll correspond à une note. Il est alors facile de visualiser les notes jouées au fil du temps.



**Figure 6.** Piano roll résultant d'une transcription automatique de *Sojerina*

La précision du système de transcription proposé est bien adaptée à la rapidité de jeu du musicien et rend bien compte de certaines subtilités des structures rythmiques caractéristiques du jeu de la *marovany*, comme on va maintenant l'expliquer brièvement. Le graphe gauche de la figure 7 superpose en fonction du temps la détection MIDI effectuée sur l'air *Sojerina* avec deux pistes correspondant à l'assise rythmique de cet air, les battements de main (ici représentés en clap MIDI) et la piste audio du hochet *kantsa*. Le graphe droit de la figure 7 montre la superposition des signaux audio de battements de main, de hochet, de l'original et de la transcription automatique MIDI (en grisé). En comparant l'audio de la cithare et celui de la dé-

tection MIDI, on remarque que les deux pistes sont bien synchronisées, bien qu'il manque des notes dans la partie MIDI, correspondant à des cordes manquantes (problème des capteurs défaillants présenté plus haut). On retrouve bien le caractère contramétrique de cette musique [17], la pulsation étant donnée par les battements de mains, on voit que l'accent du hochet est décalé par rapport à ce battement, en levée donc, c'est-à-dire à contre-temps. Plus précisément, il tombe sur la 2ème croche d'une subdivision ternaire de la pulsation (chaque pulsation se divise en trois croches). Dans la détection MIDI, il y a deux tierces qu'on remarque bien : RE-FA# et DO-MI, puis un arpège descendant SOL-RE-SI. Or ces deux tierces et le SOL sont plutôt placés sur l'accent du hochet, et non sur les frappements de mains, donc la cithare joue à contre-temps. Cette caractéristique est confirmée par la piste audio où l'on voit aussi bien dans l'original que dans la détection une intensité plus forte au moment des deux tierces, qui correspond aux accents du hochet.



**Figure 7.** A gauche, de haut en bas, superposition en fonction du temps la détection MIDI effectuée sur l'air *Sojerina* avec deux pistes correspondant à l'assise rythmique de cet air, les battements de main (ici représentés en clap MIDI) et la piste audio du hochet *kantsa*. Les deux barres verticales indiquent deux temps donnés par le clap MIDI, et la partie encerclée illustre la contramétrie du jeu, avec un placement entre à contre-temps. A droite, de haut en bas, la superposition des signaux audio de battements de main, de hochet, de l'original et de la transcription automatique MIDI (en grisé).

#### 4. CONCLUSION

Un système d'enregistrement par captation optique a été introduit dans ce travail et appliqué à la transcription automatique de la cithare *marovany*, instrument utilisé dans les séances de transe *tromba*. Parmi ces avantages technologiques, nous retiendrons l'acquisition d'un signal indépendant par corde, fort rapport signal sur bruit, forte sensibilité sur l'amplitude du mouvement. Il a été conçu à la fois pour répondre à des exigences immédiates d'études de terrain et pour assurer une méthode de transcription automatique robuste sur des projets de long terme tel qu'une classification systématique des airs de *marovany* en relation avec la transe.

La nécessité de retravailler sur des données audio se fera peut-être sentir pour récupérer des propriétés acous-

tiques plus complexes, incluant les propriétés vibratoires et acoustiques de l'instrument entier, comme le timbre de l'instrument, les modes d'ébranlement des cordes, l'intensité acoustique. Cependant, le système optique fournit un support fiable complémentaire pour les développements de la transcription à suivre, permettant la connaissance d'informations a priori susceptibles de simplifier des problèmes difficiles liés au travail direct sur l'audio, tels que le nombre de cordes vibrantes pour étudier les segments de polyphonie dans les morceaux.

Dans une perspective plus large, les applications de ce dispositif sont de deux ordres : anthropologique et artistique. Sur le plan anthropologique, les paramètres musicaux extraits de la classification du répertoire de la *marovany* pourraient conduire *in fine* à dégager une « théorie musicale » en la confrontant aux modèles heuristiques proposés par les musiciens autochtones, nous renseignant ainsi sur le rôle de la musique dans les mécanismes de déclenchement de la transe et sur le savoir-faire développé par les musiciens à cet effet.

A côté de cette application anthropologique, une autre perspective d'utilisation d'un tel système apparaît dans l'interaction musicale homme-machine. En effet, l'environnement informatique d'improvisation *ImproteK* [18] (développé à partir de l'environnement *OMax* [19] en partenariat avec l'IRCAM) permet de capter le jeu d'un musicien et d'improviser à partir des phrases enregistrées en ré-orchestrant les données MIDI avec les timbres générés par un échantillonneur. Des projets pourraient être mis en place pour que des musiciens traditionnels malgaches utilisent un tel système avec les données MIDI extraites par nos capteurs optiques. Des questions d'ordre plus esthétique (acceptabilité de formules musicales « dérivées » d'un catalogue connu, témoignage oral de cette capacité, intérêt musical dans l'amplification, la modification temps réel de paramètres musicaux, la ré-orchestration virtuelle d'un environnement musical) seront à considérer dans de futures recherches suivant cet axe.

#### 5. REMERCIEMENTS

Ce travail a été réalisé avec le soutien de l'ANR, au sein du Projet IMPROTECH ANR-09-SSOC-068. Il est issu en grande partie de deux projets de Master réalisés au LAM : *Transcription automatique de la cithare malgache* par Joachim Flocon-Cholet et *Etude de la cithare malgache en vue de la transcription automatique* par Mathilde Paul. Nous souhaitons aussi remercier Laurent Quartier pour la qualité de son apport technique à ce travail, ainsi que Velonjoro pour avoir participé à l'enregistrement de plusieurs variantes de *Midegana* et de *Sojerina*. La toolbox Matlab MIDItoolbox a été utilisée pour réaliser les pianos rolls présentés à la section 3.

#### 6. REFERENCES

- [1] Rouget, G. *La Musique et la transe. Esquisse d'une théorie générale des relations de la musique et de la*

- possession*, Paris, Gallimard, 1980.
- [2] Chemillier M. « Esthétique et rationalité dans les musiques de tradition orale », *Rapport de recherche au Ministère de la Culture*, 1998-2000.
- [3] Dianteill, E. et Hell, B. « Le possédé spectaculaire. Possession, théâtre et globalisation », *Gradhiva*, n°7, p 4-5, 2008.
- [4] Hell, B. « Négocier avec les esprits *tromba* à Mayotte. Retour sur le « théâtre vécu » de la possession », *Gradhiva*, n°7, p. 6-23, 2008.
- [5] Razafindrakoto, J. « Le timbre dans le répertoire de la valiha, cithare tubulaire de Madagascar », *Cahiers d'ethnomusicologie*, Vol. 12, p. 123-142, 1999
- [6] M. Domenichini, « Valiha », in « Stanley Sadie », ed. : *New Grove Dictionary of Musical Instruments*. London :Macmillan, Vol. 3, p. 705-706, 1984.
- [7] Chemillier, M. « Page web consacrée à la cithare *marovany* », <http://ehess.modelisationsavoirs.fr/marovany/index.html>, 2012-2013.
- [8] Seydoux, L. « Mesure de déplacement de cordes avec des fourches optiques », Mémoire de fin d'études, LAM (Laboratoire Acoustique Musicale), 2012.
- [9] Chabassier, J. « Modélisation et simulation numérique d'un piano par modèles physiques », Thèse de l'Ecole Polytechnique, 2012.
- [10] Mowat, W. « Bob Moog Piano Bar : MIDI Output device for acoustic pianos », *Sound On Sound*, mis en ligne sur [www.soundonsound.com/sos/mar05/...../articles/moogpianobar.html](http://www.soundonsound.com/sos/mar05/...../articles/moogpianobar.html), 2005.
- [11] Assayag G. and Bloch, G. *OMax. The Software Improviser*, Documentation version 2, IRCAM, May 2008.
- [12] Chemillier, M. « Compte-rendu de mission sur les capteurs optiques - Madagascar 29 juillet au 13 août 2012, Projet ANR IMPROTECH », mis en ligne sur <http://ehess.modelisationsavoirs.fr/marovany/capteurs/>, 2012.
- [13] Youngmoo, E. K. and Whitman, B. « Singer identification in popular music recordings using voice coding features », *ISMIR*, 2002.
- [14] Ephraim, Y. and Malah, D. « Speech enhancement using a minimum-mean square error log-spectral amplitude estimator », *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 33, p. 443-445, 1985.
- [15] Bello, J. P., Duxbury, C., Davies, M. and Sanders, M. B. « On the use of phase and energy for musical onset detection in the complex domain », *IEEE Signal Processing Letters*, Vol. 11, p. 553-556, 2004.
- [16] de Cheveigné, A. and Kawahara, H. « YIN, a fundamental frequency estimator for speech and music », *J. Acoust. Soc. Am.*, Vol. 111, p. 1917-1930, 2002.
- [17] Chemillier, M., Pouchelon, J., André, J. and Nika, J. « Le jazz, l'Afrique et la contramétricité », à paraître, *Anthropologie et société*, 2013.
- [18] Nika, J. and Chemillier, M. « ImproteK : intégrer des contrôles harmoniques pour l'improvisation musicale dans la filiation d'OMax », *Actes des Journées d'Informatique Musicale*, p. 147-155, 2012.
- [19] Assayag, G., Bloch, G., Chemillier, M., Cont, A. and Dubnov, S. « The OMax Project Page », available at <http://omax.ircam.fr>.