

# MUSICDISCOVER

## Identification Automatique des Instruments de Musique

### INTRODUCTION

La Reconnaissance Automatique des Instruments de Musique (RAIM), d'une manière intuitive, signifie que, étant donné un enregistrement musical, l'ordinateur essaye d'identifier quelles parties de la musique sont exécutées par quels instruments de musique.

La recherche en RAIM s'est développée au cours des 10 dernières années particulièrement en raison de son utilisation comme composant à l'intérieur des moteurs de recherche intelligents de musique, qui permettent de rechercher de la musique sur Internet ou dans des mémoires de masse, dans les ensembles de fichiers MP3 personnels, en utilisant des critères intelligents tels que le style ou la composition musicale - par opposition aux recherches impliquant seulement l'information textuelle accompagnant les fichiers musicaux. D'autres utilisations en RAIM incluent l'intégration et l'amélioration d'autres tâches en Recherche d'Information Musicale comme la transcription automatique et l'alignement de partition, et comme outil dans les applications pour les compositeurs et les studios d'enregistrement.

La RAIM est un processus complexe impliquant beaucoup d'étapes difficiles. Les diverses étapes du processus en RAIM étudiées dans MusicDiscover incluent l'obtention et le formatage de bases de données de sons d'apprentissage et de test, le calcul des Descripteurs de caractéristiques des sons, la purge automatique des bases de données, la pondération des caractéristiques et la réduction de dimension de l'espace des Descripteurs de caractéristiques, et en conclusion, la classification des sons comme provenant de tel ou tel instrument. Enfin, l'évaluation du reconnaiseur est également importante et non triviale en RAIM.

Ce chapitre traite en détail les différentes étapes de la recherche en RAIM effectuée dans le projet MusicDiscover et, tout en “remplissant des trous” dans la théorie, présente de nouvelles techniques et méthodes pour beaucoup des tâches, la RAIM des sons isolés, des solos et de la musique polyphonique et multi-instrumentale.

## ARCHITECTURE DU SYSTÈME ET CONTRIBUTIONS

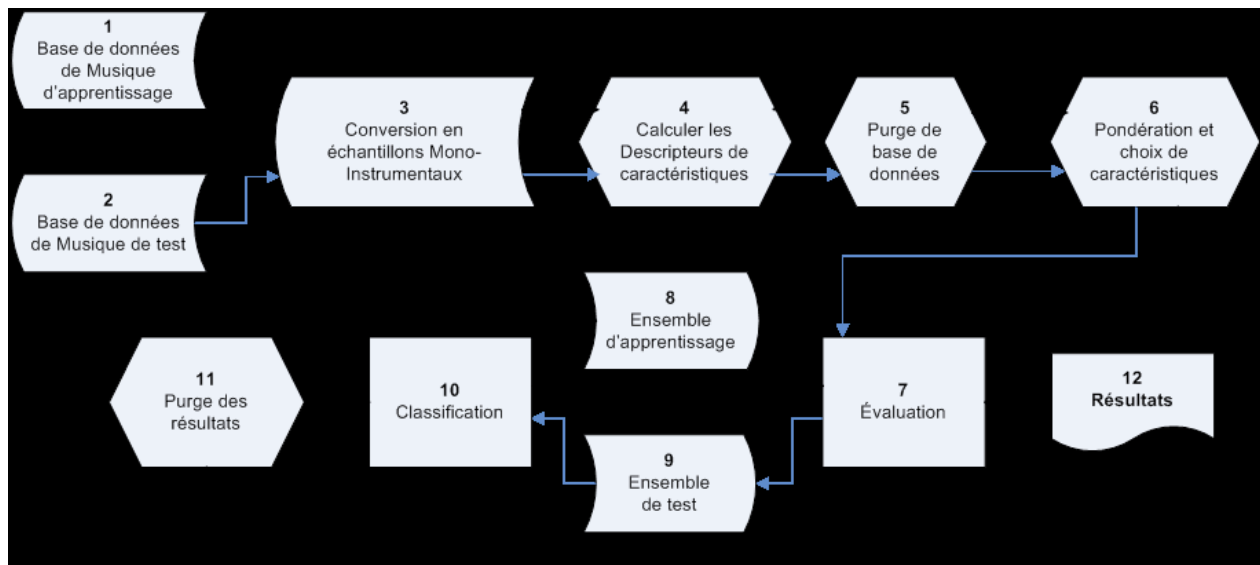


Figure 1. Un organigramme du processus en RAIM

La contribution de MusicDiscover au champ de recherches en RAIM est expliquée dans le contexte des différents composants de notre système de RAIM comme représenté sur la figure 1 :

- **Étapes 1 et 2** – la Base de données Musicale d'apprentissage est une collection d'enregistrements musicaux étiquetés en instruments, qui est employée pour faire la RAIM de la Base de données de test. Ces bases de données de musique peuvent contenir des sons isolés, des exécutions solos ou de la musique Multi-Instrumentale et Polyphonique (MIP).

- “*Les bases de données de sons isolés*” sont les bases de données de sons où chaque échantillon est un enregistrement d'une seule note de musique.

Les publications et les rapports de MusicDiscover traitant la RAIM dans les bases de données de sons isolés incluent : (Livshin et Rodet 2006 p1), (Livshin et Rodet 2006 p2) et (Livshin 2007 r19).

- “*Une exécution solo*” signifie un enregistrement monophonique ou polyphonique d'un morceau musical exécuté par un seul instrument de musique.

Tandis que les enregistrements solos de piano existent en abondance, les solos d'autres instruments, particulièrement monophoniques et semi-monophoniques sont plutôt rares et difficiles à trouver. Dans ce projet on a insisté sur le fait que chaque solo devait être un enregistrement authentique d'un morceau musical différent, enregistré dans un environnement différent et joué par un musicien différent afin de faire une évaluation croisée significative en RAIM (Livshin et Rodet 2003 p2).

Ce fait a rendu l'obtention de cette collection de solos extrêmement difficile et longue, et c'est une contribution notable en soi. Notre base de données solo a été créée en coopération avec S. Essid de l'ENST (autre partenaire dans MusicDiscover). Elle contient plus de 150 solos, y compris de musique classique, moderne et ethnique, recueillis à partir de CD commerciaux (contenant de nouveaux ou vieux enregistrements) et de fichiers MP3 trouvés sur l'Internet, joués et enregistrés par des professionnels et des amateurs.

Les publications et les rapports de MusicDiscover traitant la RAIM dans les bases de données solos incluent : (Livshin et Rodet 2004 p1) et (Livshin et Rodet 2004 p2).

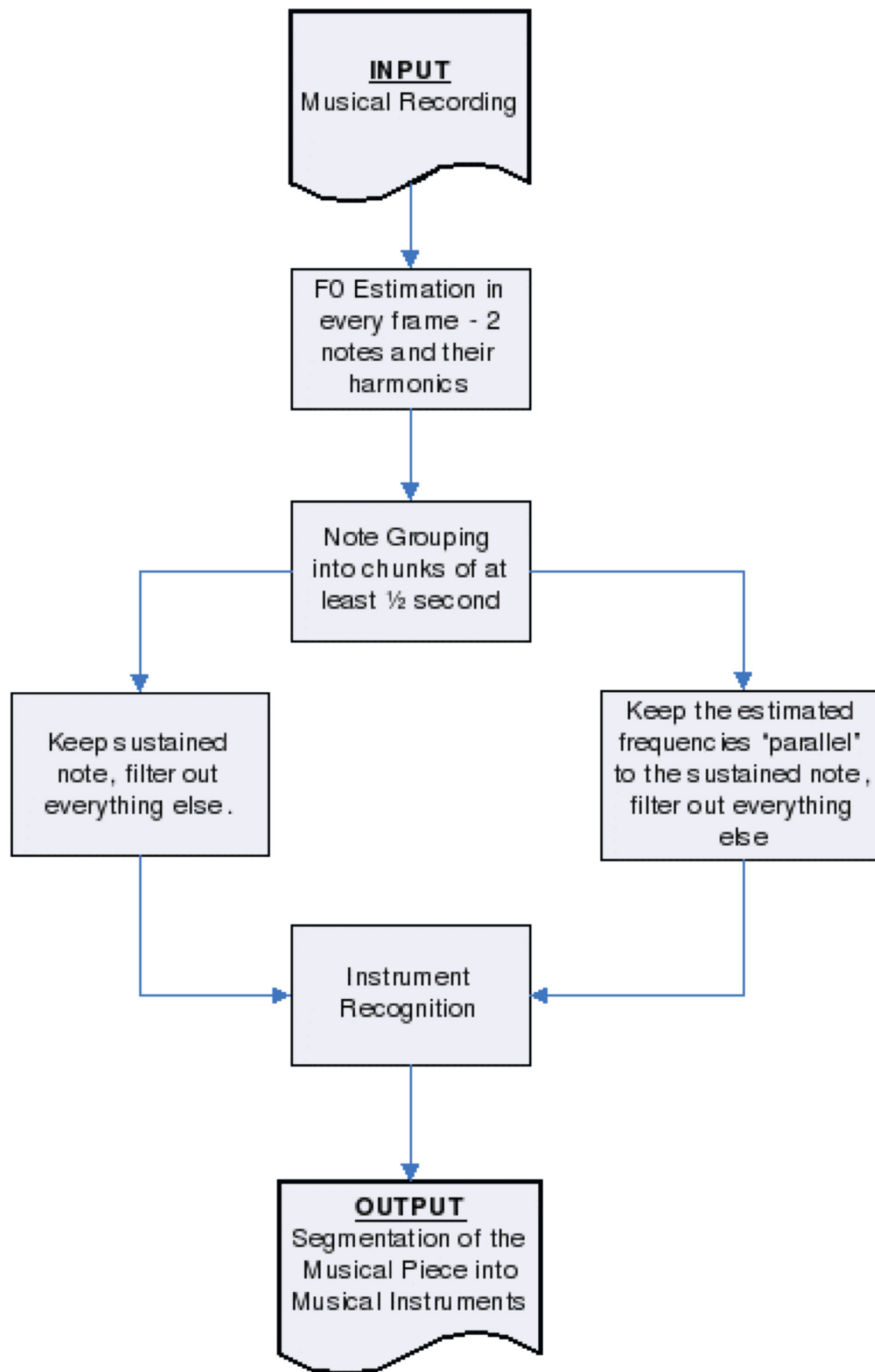
- *La musique Multi-Instrumentale et Polyphonique (MIP)* signifie la musique où plusieurs instruments musicaux jouent en même temps. La RAIM dans la musique MIP a été faite principalement sur de vrais enregistrements de duos et sur des mélanges artificiels de solos de 2 à 5 instruments différents mixés ensembles.

Les publications et les rapports de MusicDiscover traitant la RAIM dans la musique MIP incluent tous les rapports (Livshin 2007 r14) (Livshin 2007 r20) à l'exclusion de (Livshin 2007 r19).

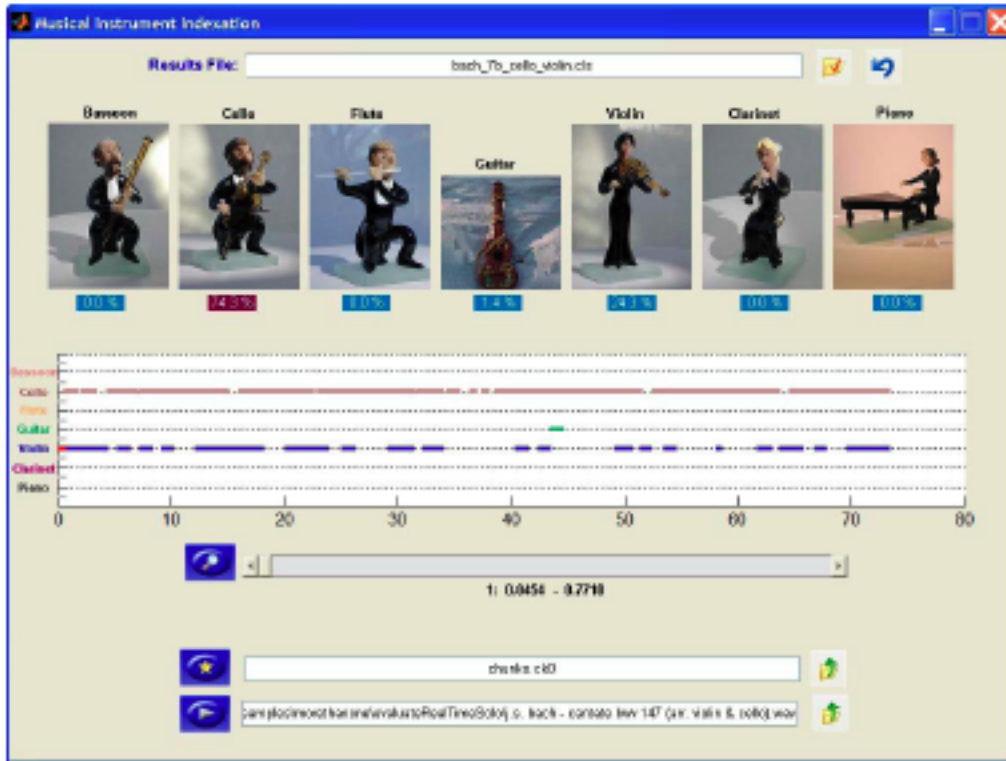
- **Etape 3** – quelles que soient les bases de données de musique originales, les méthodes en RAIM dans MusicDiscover exigent leur conversion en échantillons de sons courts, chacun contenant seulement des sons d'un seul instrument de musique. Dans le cas de RAIM de sons isolés, les bases de données viennent déjà sous la forme désirée. Pour faire la RAIM sur des solos, ceux-ci sont découpés en petit morceaux qui se recouvrent, monophoniques ou polyphoniques selon les instruments.

Quand la musique classifiée est MIP, nous proposons deux méthodes afin d'extraire les notes musicales séparées :

**1) Réduction de source** - un programme de détection de  $f0s$  multiples (Yeh, Röbel et Rodet 2005) est employé afin d'estimer les frontières des notes (où chaque note commence et finit) et les partiels harmoniques. La détection de  $f0s$  multiples, qui est très importante pour la Recherche d'Information Musicale, a été l'un des domaines de recherche dans MusicDiscover, voir ce chapitre dans ce rapport. Ensuite, chaque note de musique, à son tour, est différenciée du reste de la musique jouant en même temps, en réduisant le niveau de volume de toutes les autres notes jouant en même temps qu'elle. De cette façon, métaphoriquement, le projecteur "RAIM" est tournée vers chaque note individuellement.



**Figure 2. Le procédé de Réduction de Source, une version pour des duos, y compris une stratégie d'Anti-Note**



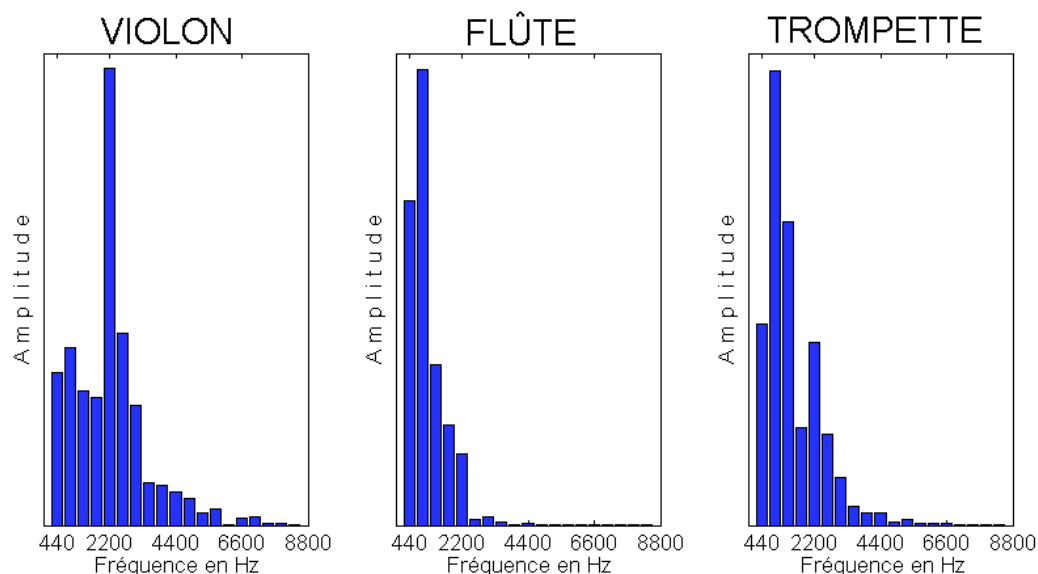
**Figure 3. Notre système de Réduction de Source**

Dans (Livshin et autres 2005), nous avons intégré un module RAIM de Réduction de Source avec un module d'évaluation de  $f0s$  multiples pour créer un système qui produit automatiquement des partitions à partir de la musique MIP enregistrée, en trouvant d'abord les différentes notes dans le morceau musical et en les arrangeant ensuite dans des portées suivant leurs instruments de musiques reconnus.

Publications et rapports de MusicDiscover traitant la Réduction de Source : (Livshin et Rodet 2004 p2), tous les rapports de (Livshin 2007 r14) à (Livshin 2007 r18), et (Livshin et autres 2005).

**2) Resynthèse harmonique** – on a montré dans (Livshin et Rodet 2006 p1) et (Livshin et Rodet 2006 p2) que pour la RAIM de sons isolés, employer seulement l'information présente dans la série harmonique suffit pour obtenir des taux élevés d'identification d'instrument. Par conséquent, les partiels estimés pour une note, tels que fournis par un programme de détection de  $f0s$  multiples (Yeh, Röbel et Rodet 2005) peuvent être employés directement

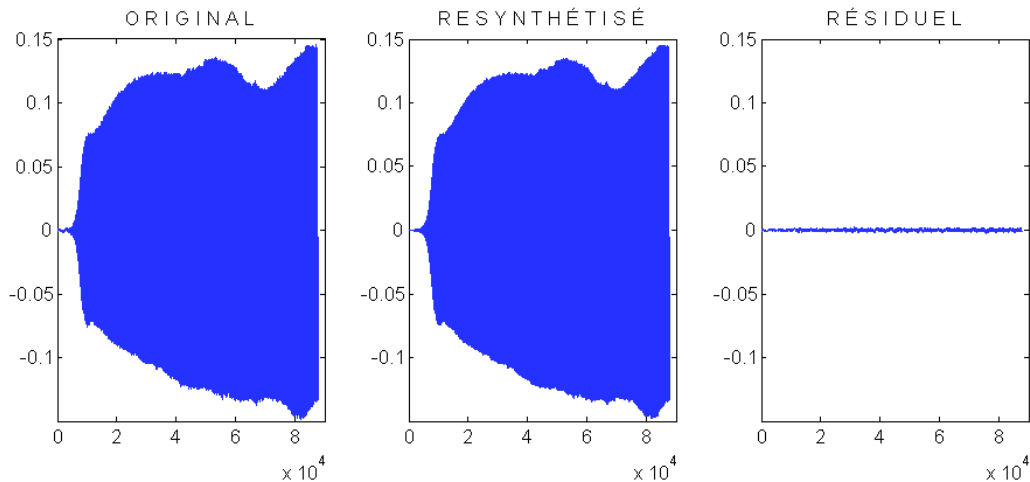
pour synthétiser chaque note d'un morceau MIP et donc être employés pour la RAIM.



**Figure 4.** Les premiers 20 partiels harmoniques d'une note A4 jouée par différents instruments

Publications et rapports de MusicDiscover traitant de Resynthèse Harmonique : (Livshin 2007 r17), (Livshin 2007 r18) et (Livshin 2007 r20).

- **Etape 4** – quand les courts échantillons de sons sont prêts, diverses statistiques sont calculées sur chaque échantillon afin d'obtenir les différentes caractéristiques du son. Ces statistiques s'appellent les descripteurs de caractéristiques. Voir (Peeters 2004) pour une liste des descripteurs de caractéristiques utilisés dans ce travail.



**Figure 5.** De gauche à droite, Echantillon original de Clarinette, l'échantillon synthétisé à partir de sa série harmonique, et le résidu par soustraction.

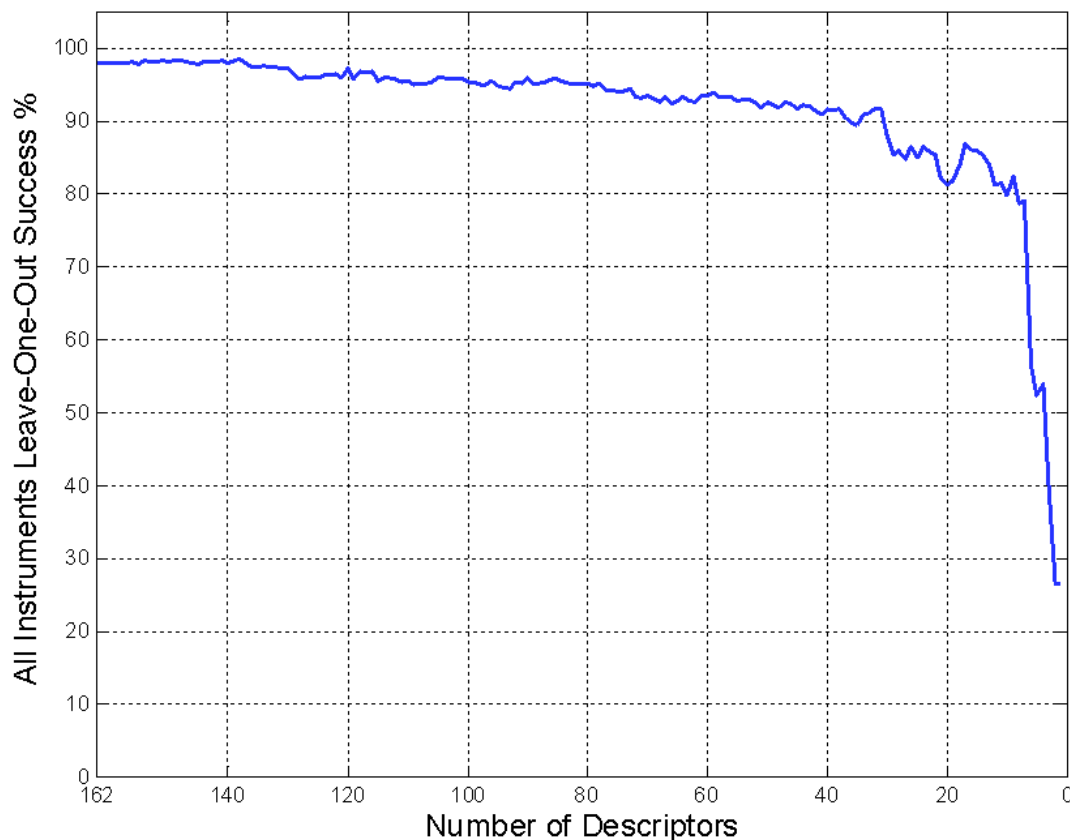
- **Etape 5** – dans le cas où la base de données de sons d'apprentissage peut contenir des échantillons qui sont mal enregistrés ou mal étiquetés, il est salutaire de la nettoyer afin d'empêcher des erreurs de classification. Il est également possible d'enlever ou marquer les échantillons anormaux de l'ensemble de test qui sont susceptibles d'être classifiés inexactement, dépendant des conditions de l'application. De tels processus de nettoyage s'appellent 'purger la base de données' et sont présentés et discutés dans (Livshin 2006 r19).

Nous avons obtenu le meilleur taux en utilisant base de données en purgée suivant la méthode MIQR que nous avons proposée et développée. Cette méthode MIQR a réussi dans 70.1% des cas l'enlèvement "des mauvais" échantillons dans une base de données de sons volontairement souillée de mauvais échantillons tout en enlevant seulement 0.9% des échantillons originaux et bons. Si l'on permet un taux de fausse alarme plus élevé, MIQR peut enlever jusqu'à 90.4% des "mauvais" échantillons contre 8.8% de "bons".

- **Etape 6** – afin d'améliorer la séparation des classes et réduire la dimensionnalité des données, les descripteurs de caractéristiques sont pondérés et les matrices de caractéristiques sont transformées pour obtenir un espace de caractéristiques de faible dimension. La « sélection de caractéristiques » est utilisée et signifie que les meilleurs descripteurs pour un problème spécifique sont choisis.



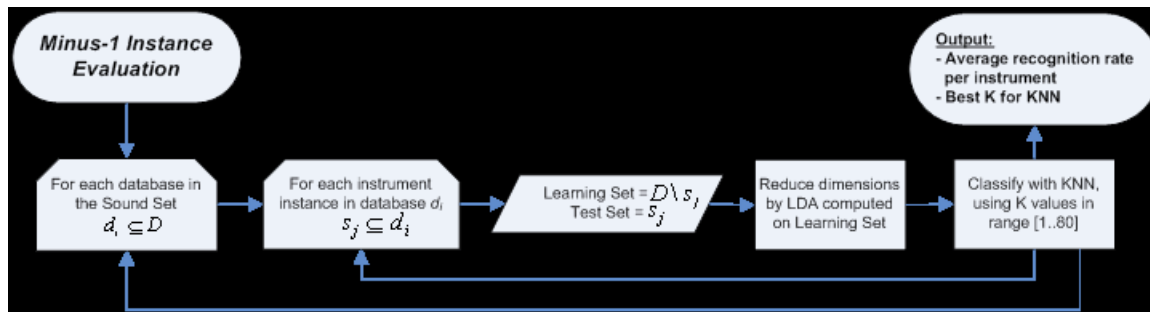
Voir (Livshin 2006 r16) pour choisir les meilleurs descripteurs pour effectuer la Réduction de Source de la musique MIP. Dans ce rapport, le Taux Mutuel moyen pour notre base de données de duos est augmenté de 83.2% à 86.2% en limitant l'ensemble de descripteur de caractéristiques aux descripteurs choisis automatiquement.



**Figure 6.** Elimination des descripteurs avec la méthode GDE. Taux de reconnaissance des instruments sur une base d'échantillons différentes de 16 instruments de musique.

- **Etape 7** –l'évaluation d'un processus en RAIM exige beaucoup de soin. Tandis que la méthode idéale d'évaluation devrait vérifier si un processus de RAIM identifie un instrument indépendamment des conditions d'enregistrement (ce qu'on nomme la « généralisation »), certaines méthodes d'évaluation, mal choisies, pourraient par erreur seulement vérifier s'il peut apprendre à identifier des types très limités d'échantillons (Livshin et Rodet 2003 r2).

Deux méthodes précises et « généralisantes » d'évaluation sont présentées et discutées : L' "Exemple Moins-1" dans (Livshin et Rodet 2006 p1) est employé pour l'évaluation en RAIM sur des sons isolés et "Moins-1 en solo" dans (Livshin et de Rodet 2004 p1) est employé pour les solos. La RAIM sur la musique MIP, comme fait dans ce projet, n'exige aucune technique spéciale d'évaluation.



**Figure 7.** Evaluation « Exemple Moins-1 » avec LDA et KNN

- **Étapes 8 et 9** – une méthode d'évaluation peut faire plusieurs tours de classification, en employant chaque fois différentes collections de données issues des ensembles d'apprentissage et de test pour la classification. Ces ensembles dans les étapes 8 et 9 sont des sous-ensembles des bases de données initiales de musique présentés par étapes 1 et 2.
- **Étape 10** – les vecteurs de caractéristiques sont classifiés par l'algorithme de classification dans des classes selon la taxonomie musicale désirée d'instrument.

Dans MusicDiscover, nous avons la plupart du temps employé l'analyse discriminante linéaire couplée à l'algorithme des K-Plus-Proche-Voisins pour la RAIM. Cette combinaison d'algorithme apparaît dans toutes les références de Livshin à partir de 2004.

- **Étape 11** – la purge semi finale des résultats de classification peut être faite après que les sons sont classifiés, en calculant des niveaux de confiance, en notant certaines des classifications en tant que probablement incorrectes et en les supprimant ou en les marquant. Une des méthodes possibles pour détecter et enlever des classifications douteuses est d'utiliser les mêmes algorithmes de purge que dans l'étape 5, comme fait dans (Livshin 2007 r20) ; les taux moyens d'identification avec la Réduction de Source sont augmentés de 54.1% à 55.1% et les

taux avec la Resynthèse Harmonique augmentent en moyenne de 56.9% à 58.6% pour la RAIM dans les -mélanges solo de 2 à 5 voix, en employant MIQR pour enlever des classifications douteuses.

- **Etape 12** – quand le processus d'évaluation est fini, les résultats d'évaluation en RAIM sont enregistrés.

Nous avons atteint les taux suivants en RAIM :

Pour la RAIM dans des sons isolés, nous avons réalisé un taux moyen d'identification d'instrument de 95.58% pour 10 instruments musicaux (Livshin et Rodet 2006 p2).

Pour des exécutions solos, un taux moyen d'identification de Moins-1-Solo de 88.13% et un taux d'identification en temps réel de 85.54% ont été réalisés pour 7 instruments musicaux dans (Livshin et Rodet 2004 p2).

Pour la RAIM dans la musique MIP, les vrais enregistrements de duos ont produit un taux mutuel de 83.2% avec 100% de détection d'instrument, c'est-à-dire, en déterminant correctement quels instruments jouent dans les duos examinés. Un taux moyen d'identification de 56.9% a été atteint pour les mélanges solos de 2 à 5 instruments jouant simultanément en utilisant la technique Resynthèse Harmonique.

### ***Conclusion :***

La reconnaissance automatique d'instruments de musique est un champ multidisciplinaire large et complexe. Elle implique la connaissance de divers secteurs de recherches comprenant la théorie musicale, la physique audio, la théorie des instruments de musique, le traitement du signal et l'identification de modèle.

Tandis que MusicDiscover a notablement amélioré la situation courante dans les différents composants et étapes impliqués dans la RAIM, la recherche sur l'identification musicale automatique d'instruments est toujours loin d'être terminée. Créer un système qui pourrait identifier chaque instrument

participant à un concert classique, par exemple, au moins au niveau d'un auditeur humain, est encore loin du possible.

La recherche en RAIM récapitulée dans ce rapport apparaît en détail dans la thèse de PhD de Livshin, 2007.

## RÉFÉRENCES

- Livshin, A., 2004 r14. "Rapport 14 – résultats préliminaires d'identification d'instrument dans les duos", *rapport interne d'IRCAM*, juin 2004.
- Livshin, A., 2005 r15. "Rapport 15 – issues de réduction de source et expériences", *rapport interne d'IRCAM*, janvier 2005.
- Livshin, A., 2005 r16. "Rapport 16 – choisissant les meilleures caractéristiques pour l'identification d'instrument dans la musique Multi-Instrumentale", *rapport interne d'IRCAM*, mai 2005.
- Livshin, A., 2006 r17. "Harmonique Resynthesis – du rapport 17 contre la réduction de source (partie 1 de 2) - identification d'instrument dans solos et duos", *rapport interne d'IRCAM*, novembre 2006.
- Livshin, A., 2007 r18. "Harmonique Resynthesis – du rapport 18 contre la réduction de source (partie 2 de 2) - identification d'instrument dans mélanges solos Multi-instrumentaux et Polyphoniques", *rapport interne d'IRCAM*, janvier 2007.
- Livshin, A., 2007 r19. "Rapport 19 – améliorant l'uniformité d'une base de données de sons", *rapport interne d'IRCAM*, février 2007.
- Livshin, A., 2007 r20. "Rapport 20 – légères améliorations des taux multi-instrumentaux polyphoniques d'identification en utilisant MIQR", *rapport interne d'IRCAM*, mai 2007.
- Livshin, A., 2007. "IDENTIFICATION AUTOMATIQUE DES INSTRUMENTS DE MUSIQUE", *Thèse De PhD*.
- Livshin, A., Rodet, X. 2003. "l'importance de l'évaluation en travers de base de données dans le son musical d'instrument  
Classification : une approche critique ", dans *les démarches du colloque international sur la recherche documentaire de musique (ISMIR'03)*.
- Livshin, A., Rodet, X., 2004 p1. "Identification d'instrument au delà des notes séparées – classant les enregistrements continus," dans *les démarches de la conférence internationale de musique d'ordinateur (ICMC'04)*.
- Livshin, A., Rodet, X., 2004 p2. "L'identification musicale d'instrument dans les enregistrements continus," dans *la conférence internationale du Th'7 de DAFx 2004 sur l'acoustique de Digital effectue Proc.*, pp 222-227.
- Livshin, A., Rodet, X., 2006 p1. "L'importance du résiduel non-harmonique pour l'identification musicale automatique d'instruments des instruments lancés," dans *les démarches de la convention audio de société de technologie (AES 2006)*.
- Livshin, A., Rodet, X., 2006 p2. "La signification du son "non-harmonique" contre la série harmonique pour l'identification musicale d'instrument," dans *les démarches du colloque international sur la recherche documentaire de musique (ISMIR'06)*.

Livshin, A., Yeh, C., Röbel, A., Rodet, X. 2005. "Une forme d'onde au système de conversion du MIDI," *rapport non publié d'IRCAM*.

Peeters, G. 2004. "un grand ensemble de caractéristiques audio pour la description de sons (similitude et classification) dans le projet de CUIDADO". *CUIDADO I.S.T. Rapport 2004 De Projet*. URL : [http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters\\_2003\\_cuidadoaudiofeatures.pdf](http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf)