

RAPPORT d'activité
2006

Rapport de synthèse
Rapport détaillé

Responsable : Xavier Rodet

L'équipe Analyse-Synthèse étudie le signal audio-numérique, soit pour en extraire tous types d'informations, soit pour le synthétiser selon certaines propriétés sonores et musicales, soit pour le transformer en fonction de besoins compositionnels, sans oublier des applications dans d'autres domaines comme le cinéma, le multimédia, la téléphonie ou même des domaines industriels. L'un des aspects les plus nouveaux est l'intérêt marqué par les compositeurs et les partenaires de l'Ircam pour le traitement et la synthèse de la voix. Ainsi un effort important lancé en 2005 sur ce thème trouve des débouchés nombreux cette année (projet VIVOS) avec des travaux sur l'expressivité, la reconstitution d'une voix parlée, ou la transformation de nature ou d'identité d'une voix.

Un autre aspect qui connaît un grand développement est celui du traitement par le contenu qui cherche à trouver automatiquement des caractéristiques des œuvres audio, telles que la hauteur, les instruments, les styles, etc. (projet Ecoute). Cela signifie que la caractérisation de telle ou telle partie du signal musical, en temps et/ou en fréquence, non seulement fournit des informations musicales précieuses mais encore permet d'effectuer des traitements spécifiques, adaptés au problème et qui sans cette connaissance ne pourraient réussir. Ainsi, de nouveaux outils pour les compositeurs et musiciens pourront être conçus et construits. Il faut souligner que ces recherches et les développements qui en découlent, sont reliés naturellement aux Groupes de Travail avec les Compositeurs, « Voix » et « Orchestration » notamment.

Le domaine des modèles physiques se renforce dans l'équipe, notamment grâce à la composante UMR-CNRS, avec un chargé de recherche et les travaux de trois stagiaires qui poursuivent en doctorat, et le projet blanc ANR-CONSONNES. Les travaux concernent l'analyse-synthèse de phénomènes non linéaires, des débouchés vers des implémentations en temps réel, les méthodes d'inversion entrée/sortie de modèles d'instruments et le début d'un projet « mécatronique » associant mécanique des instruments et commande électronique.

Enfin les travaux de *développement* sont marqués par plusieurs contrats de l'extérieur notamment avec des sociétés américaines comme MakeMusic. De progrès importants ont été obtenus dans les méthodes d'analyse et de synthèse, par exemple la réduction du biais dans l'estimation des paramètres des sinusoïdes dans le logiciel SuperVP ou la synthèse et la pré-écoute des modèles sinusoïdaux en temps réel dans le logiciel

AudioSculpt. Pour ce dernier, les améliorations sont nombreuses, telles que le copier-coller et l'écoute de zones marquées sur le sonagramme, la sélection par niveau d'amplitude (baguette magique) ou la réduction de bruit en temps réel.

1.1 Modèles d'analyse et de synthèse du signal audio

Les progrès dans l'analyse du signal, c'est-à-dire dans la compréhension de sa structure et l'estimation de ses paramètres autorisant ainsi des traitements sonores de haute qualité, nécessitent des modèles, de structure, de production, ou de perception. De plus en plus, l'apprentissage automatique des modèles donne des résultats complémentaires voire meilleurs que ceux qui dépendent de l'expert. En conséquence, un point important est celui de la disponibilité de bases de données de sons et de musiques, permettant cet apprentissage et de bases de données de référence permettant l'évaluation. C'est une des raisons du projet MusicDiscover, avec bien sûr le développement de méthodes d'accès au contenu (Cf. ce paragraphe plus loin). Cependant, l'étude et le développement de nouvelles méthodes performantes d'analyse du signal reste tout à fait nécessaire. Nous y travaillerons notamment dans la recherche des fondamentaux (hauteurs) simples et multiples, essentiels pour le traitement des flux polyphoniques, le traitement du vibrato, l'adaptation automatique de la taille de la fenêtre d'analyse ou l'apprentissage automatique de modèles d'instruments.

1.1.1 Etude d'algorithmes pour l'adaptation automatique de la résolution temps fréquence

Pour améliorer la qualité des analyses et des signaux transformés, nous cherchons à adapter la résolution temporelle et fréquentielle au signal. Pendant l'année 2006 nous avons commencé à étudier des algorithmes basés sur des mesures statistiques de la transformée du signal qui ont été proposées dans la littérature [Hamnane06a]. L'étude menée a été purement théorique et nous allons compléter ces travaux avec une étude plus expérimentale en 2007.

Participant : S. Hamnane (stage ATIAM), A. Roebel.

1.1.2 Modèles d'enveloppes spectrales des instruments

La modélisation des enveloppes des instruments est une technique qui pourra servir pour plusieurs problèmes. Les exemples intéressants pour les travaux de l'équipe sont l'estimation des fréquences fondamentales multiples, la transposition avec préservation d'un timbre naturel, la reconnaissance des instruments et la séparation des sources sonores. Au cours de l'année 2006, nous avons étudié les propriétés de plusieurs modèles d'enveloppes. Les instruments ciblés sont harmoniques et le principe fondamental est donc la représentation des amplitudes des partiels des signaux par une superposition linéaire

d'enveloppes de base. Les enveloppes de base sont obtenues grâce à une PCA des enveloppes d'une collection de signaux tests qui ont été sélectionnés parmi les signaux de la base RWC. Pour l'entraînement des modèles, nous avons utilisé deux instruments différents appartenant à la même classe d'instruments et utilisant un domaine de fréquence fondamentale d'une octave. Un résultat intéressant de nos expériences est la comparaison entre différentes représentations fréquentielles [Burred06a]. Les deux versions représentent les fréquences des partiels soit en utilisant le numéro de l'harmonique soit en utilisant la fréquence du partiel. Les expériences nous ont montré que la représentation fondée sur les fréquences crée une meilleure compression des données.

Participants : A. Roebel, J. Burred (stage)

Collaboration extérieure: équipe systèmes de communications, Université Technique de Berlin, Allemagne

1.1.3 Détection des débuts de notes.

La segmentation des signaux audio en notes individuelles est une tâche souvent demandée. L'évaluation de notre algorithme de détection des transitoires dans le contexte de la tâche « Onset detection » de MIREX 2005 nous avait montré que l'adaptation de l'algorithme de détection des transitoires pour le problème de la détection des onsets nécessitait encore des améliorations. Deux améliorations ont été implémentées en 2006. Premièrement, le modèle statistique qui évalue la probabilité des transitoires dans des bandes fréquentielles a été modifié. Avec la nouvelle implémentation, la taille des bandes n'est plus constante et les statistiques sont prélevées en utilisant des bandes de largeurs variables tel que le paramètre du modèle, qui était difficile à régler avant, devient moins sensible. Deuxièmement, un nouveau paramètre de contrôle a été ajouté pour permettre le contrôle de la distance minimale entre les transitoires détectés. Les résultats de l'évaluation de l'algorithme lors de MIREX 2006 montrent que les changements de l'algorithme ont amélioré les résultats d'une façon significative [Roebel06c].

Participant : A. Roebel

1.1.4 Estimation de la fréquence fondamentale des instruments percussifs

Les algorithmes d'estimation de la fréquence fondamentale qui ont été développés pendant les dernières années ne fonctionnent que si les signaux sont suffisamment harmoniques. Pour les instruments dits « pitched percussions » cette condition n'est pas remplie. Un deuxième problème qui se pose pour l'analyse de ce type d'instruments est le fait que les notes résonnent très longuement. Pour améliorer les résultats des estimations de la fréquence fondamentale pour ces instruments, nous avons mis au point un algorithme qui nous permet la détection des onsets et des fréquences fondamentales pour les instruments percussifs. L'algorithme se base sur le module de détection des transitoires discuté plus haut qui nous permet de séparer les parties spectrales des notes consécutives et, ensuite, la détection de la fréquence fondamentale de la note actuelle.

Participants : A. Roebel

1.1.5 Estimation de fréquences fondamentales multiples (contexte polyphonique)

Dans les années précédentes, un algorithme d'estimation de fréquences fondamentales (F0s) multiples a été développé pour le cas où le nombre de F0s est connu. Pendant l'année 2006, le travail s'est focalisé sur l'estimation du nombre de sources sonores. Le premier essai était l'application de la fonction « score » proposée en 2005 pour estimer le nombre de F0s. Un nombre croissant de F0s est testé jusqu'à ce qu'il n'y ait plus d'amélioration du score. Cette méthode a été intégrée dans un système de suivi pour les enregistrements des instruments monodiques [Yeh06b].

Nous avons trouvé que la limite d'amélioration du score varie avec le nombre de sources et le rapport signal à bruit. En rapport avec ce problème nous avons considéré l'utilisation de modèles probabilistes du spectre résiduel pour mieux différencier les composantes sinusoïdales qui seraient expliquées par une combinaison de F0s. Un modèle du spectre résiduel pour une estimation du niveau bruit adaptée localement dans le plan temps-fréquence a été proposé [Yeh06c]. Cette estimation du niveau bruit est une étude complémentaire de la classification de pics spectraux dans le cas polyphonique. Le modèle du spectre résiduel nous permet de classifier des pics sinusoïdaux perturbés par les pics proches dont la sinusoïdalité n'est pas observée. Un spectre relatif au niveau du bruit pourrait également être défini, ce qui faciliterait l'utilisation de l'amplitude spectrale en échelle *log*.

Le modèle du spectre résiduel a été intégré dans l'estimation des F0s. Une combinaison de F0s doit expliquer la plupart des pics sinusoïdaux. La difficulté rencontrée dans l'estimation du nombre de F0s a été l'extraction de F0s ayant une relation harmonique entre eux parce qu'il existe alors plusieurs façons de décomposer une séquence de partiels.

Nous avons aligné une collection de morceaux polyphoniques de "RWC Classical" en utilisant le programme "Alignement de Partition" de l'équipe. Les références en notes MIDI doivent toutefois être corrigées à la main. Ce travail a été effectué en partie dans le cadre du projet MusicDiscover de l'ACI « Masse de Données » (voir ci-dessous) de l'équipe et servira dans d'autres études, comme la reconnaissance des instruments.

Participants : C. Yeh (thèse), A. Roebel

Collaborations internes : A. Livshin (thèse)

1.1.6 Classification des pics spectraux

L'algorithme de classification des pics spectraux (en pics sinusoïdaux et pics bruités) développé dans les années précédentes est une technique qui est utilisé dans plusieurs applications. Pendant l'année 2006 nous avons commencé à travailler pour améliorer la performance et l'interface de contrôle de l'algorithme. A cause du problème du contrôle de l'algorithme par l'utilisateur final, le descripteur qui représente la durée du signal n'est pas utilisé dans les applications. Le contrôle utilisé actuellement permet à l'utilisateur de choisir la

probabilité de mauvaise classification d'un pic bruité. Cette probabilité n'est directement liée avec le seuil de classification que si un seul descripteur est utilisé. Une alternative possible pourrait utiliser la relation minimale entre le niveau d'un pic sinusoïdal et le niveau du bruit de fond (SNR) pour lequel une classification est demandée. L'utilisation de cette valeur de contrôle nécessitera de savoir lier le SNR avec les seuils de classification. En 2006 nous avons commencé à développer un modèle simple qui nous permettra d'exprimer les relations entre les descripteurs du pic sinusoïdal modulé et le niveau de bruit de fond. Actuellement ce modèle n'a été évalué que pour la fenêtre de Hanning. Nous poursuivons ces études afin de pouvoir l'intégrer dans les applications existantes.

Participants : A. Roebel

Collaborations extérieures : M. Zivanovic (Université de Pampelune)

1.1.7 Modèles sinusoïdaux

Les modèles sinusoïdaux sont importants pour l'analyse et la représentation des signaux à un niveau d'abstraction moyen (au dessus du spectre à court terme). Les analyses sinusoïdale servent actuellement, par exemple, dans le contexte du projet orchestration, de la classification des instruments etc. Un problème bien connu qui se pose lors de l'analyse sinusoïdale est le biais de l'estimation standard lorsque les sinusoïdes ne sont pas stationnaires. Afin d'améliorer les résultats des analyse dans cette situation nous avons développé un nouvel algorithme.

L'estimateur basé sur la maximisation de la vraisemblance nécessite une démodulation fréquentielle du signal pour obtenir les paramètres standard pour les sinusoïdes non stationnaires. L'idée du nouvel algorithme est alors d'obtenir le signal démodulé à partir d'une estimation initiale de la pente fréquentielle. Avec trois valeurs voisines de la pente fréquentielle, un modèle polynomial d'ordre 2 de la relation entre l'amplitude du pic démodulé et la pente fréquentielle peut être établi. A partir de ce modèle polynomial, la pente optimale peut alors être estimée et, après démodulation finale, les paramètres de la sinusoïde peuvent être obtenus avec un algorithme standard. L'astuce principale de la méthode proposée est le fait que les démodulations sont calculées dans le domaine spectral en n'utilisant que le pic à analyser de sorte que l'algorithme peut être utilisé pour l'estimation des paramètres sinusoïdaux dans les signaux à plusieurs composantes. Une version initiale de l'algorithme a été publiée [Roebel06b].

Participants : A. Röbel

1.1.8 Emergence temporelle de bruits non voulus

Dans le cadre d'une collaboration industrielle avec Renault et Sncf, une étude a été menée sur l'émergence temporelle des bruits parasites à l'intérieur des habitacles de voitures (automobile/TGV). Côté SNCF, ces bruits proviennent par exemple des vibrations de tablette et de garnissage. Côté Renault, les émergences temporelles proviennent de différentes sources, comme les sièges, divers habillages ou la planche de bord qui constitue, dans le

cadre de cette étude, la seule source prise en compte mais possédant de multiples parasites (boite à fusibles, boîte à gants, habillage sous le volant et divers accessoires, ...). Pour les industriels, la problématique est la suivante : ces bruits n'ont pas d'intérêt pour le client et dégradent la qualité perçue du véhicule, il ne faut donc pas les entendre émerger du bruit de fond. D'un point de vue scientifique, l'objectif de cette étude est de définir dans quelles conditions ces émergences temporelles ne sont pas entendues dans le bruit de fond en proposant des critères objectifs capables de quantifier cette perception et de prédire le seuil de détection d'un bruit parasite mélangé à un bruit de fond.

Le travail s'est articulé en plusieurs étapes : a) étude bibliographique centrée sur les derniers développements des modèles de sonie partielle (notamment celui de Moore et Glasberg), , b) préparation du corpus de sons à partir des enregistrements réalisés par Renault et SnCF (avec notamment des problèmes d'extraction/segmentation des séquences), c) test de catégorisation visant à réduire le corpus de test en définissant des classes de sons et des prototypes associés, d) test de détection sur ce corpus réduit visant à définir le seuil d'audibilité d'un parasite dans un bruit de fond, e) implémentation d'un algorithme qui, à partir des caractéristiques morphologiques du parasite et du bruit de fond, donne une probabilité d'émergence [Minard06a].

Participants : A. Minard, N. Misdariis, P. Susini

Collaborations internes : G. Peeters, X. Rodet (équipe A/S)

Collaborations externes : S. Langlois (Renault), F. Poisson (SNCF)

1.1.9 Analyse des sons produits par des bébés

Ce projet étudie les précurseurs de la parole dans le signal sonore produit par des bébés. Nous cherchons une description du son s'associant à des informations sur les mouvements des mains ou du regard qui montre une évolution, ou non, vers le langage. Dans ce but, nous avons réalisé en 2006 un prototype pour la reconnaissance automatique du bébé et de sa mère dans des enregistrements bruités. Les différents points que nous avons abordés sont donc :

- 1) Analyse du son, extraction de descripteurs (MFCC, F0, apériodicité, ...)
- 2) Sélection de descripteurs
- 3) Classification avec GMM

Participant : M. Derio (thèse)

Collaborations internes : , G. Peeters, X. Rodet,

Collaborations extérieures: V. Desjardins, J. Kiss (projet PILE) et Fondation EADS

1.1.10 Modélisation et reconnaissance en temps réel d'événements musicaux

Ce sujet concerne la segmentation et la reconnaissance d'événements musicaux en temps-réel, dans un flux audio monophonique (un seul instrument, non polyphonique). Le travail de cette première année a permis de préciser les problématiques de la thèse, par la définition d'un scénario applicatif visé pour lequel un compositeur spécifie un vocabulaire musical (au

sens d'un ensemble d'unités acoustiques de granularité temporelle semblable à une note, mais dont les attributs distinctifs peuvent être sur d'autres plans que la hauteur fixe - articulation, mode de jeu, dynamique par exemple), un apprentissage détermine la structure des modèles, et une reconnaissance est effectuée au cours de l'interprétation. Le contexte applicatif visé est celui de l'interaction musicien-machine (un interprète et une partie électronique).

En nous inspirant de la littérature fournie qui existe à propos de la modélisation d'unités acoustiques par des modèles de Markov cachés (en reconnaissance de parole, en transcription musicale ou en suivi de partition), nous avons mené des études préliminaires afin de cerner la pertinence et les limites de tels modèles dans notre cas, au travers de 3 axes principaux : la modélisation sans expert (sélection de modèles à partir des données d'apprentissage), l'apprentissage sur peu de données (qui impose l'utilisation de modèles parcimonieux) et le décodage dans un flux.

Pour les deux premiers axes, nous avons mis en oeuvre une méthode de sélection de modèle par un critère BIC (Bayesian information criterion) sur des données synthétiques. Les résultats préliminaires semblent indiquer que pour une tâche de classification donnée, notre méthode permet de sélectionner des modèles moins complexes que les modèles génératifs responsables des données. Pour la reconnaissance dans un flux, nous avons mis en place une méthode de décodage Viterbi à court terme, en nous inspirant de l'algorithme Short-Time DTW proposé l'année précédente dans l'équipe Analyse/Synthèse par H.Kaprykowsky [Kaprykowsky06a].

Participants : J. Bloit (thèse)

Collaboration interne: X. Rodet, équipe Interactions Musicales Temps-Réel

1.2 Traitement et synthèse de la voix

Dans le cadre de productions internes ou extérieures à l'Ircam, de nombreuses demandes concernent le traitement et la synthèse de la voix. Nous avons donc monté et obtenu un important projet RIAM, nommé VIVOS, sur la transformation et synthèse de la voix en collaboration avec France-Télécom, l'IRISA, le Studio de doublage Chinkel et la société de jeux vidéos BeTomorrow. Ce projet a été tout particulièrement remarqué et apprécié par le jury d'experts du RIAM et a commencé en 2006. Plusieurs recherches sont donc menées dans ce domaine dans notre équipe : segmentation automatique de la parole, synthèse à partir du texte, travaux sur la transformation d'identité, travaux sur l'expressivité, travaux avec des compositeurs (J. Harvey, J. Fineberg), enfin Groupe de Travail avec des compositeurs sur la Voix.

1.2.1 Estimation de l'enveloppe spectrale « True Envelope »

La technique d'estimation de l'enveloppe spectrale, dites « true envelope », qui a été développé à l'équipe en 2005 est utilisée avec un grand succès dans plusieurs applications,

et nous avons continué de la développer. Le premier point est l'estimation de la phase de la fonction de transfert de l'enveloppe. Fondé sur l'hypothèse que la fonction transfert est un filtre à phase minimale, la phase de l'enveloppe spectrale peut directement être calculée à partir des coefficients cepstraux. Une évaluation expérimentale de la qualité d'estimation de la méthode nous a montré un problème lié au fait que les signaux harmoniques n'ont dans la réalité pas de pic spectral à fréquence 0 qui reflète la valeur de la fonction de transfert de la source sonore. A cause de ce pic manquant, les résultats de la méthode sont très souvent perturbés autour des premiers partiels. En raison de cette perte d'information l'enveloppe n'est plus échantillonnée d'un façon régulière et l'ordre du modèle doit être réduit d'un facteur 2. Une amélioration a pu être obtenue par une approche en deux étapes. Premièrement une estimation avec ordre réduit est calculée pour obtenir une valeur à la fréquence 0. Fondé sur le spectre ainsi complété une deuxième estimation est effectuée en utilisant l'ordre théorique optimal pour la fréquence fondamentale. Les expériences [Roebel07a] montrent que, avec cette modification, la méthode « true envelope » avec sélection de l'ordre à partir de la fréquence fondamentale donne des résultats presque toujours aussi exacts que la méthode DAP avec ordre optimisé par minimisation d'erreur. L'avantage majeur de la méthode « true envelope » est que la sélection de l'ordre peut se faire à partir d'une caractéristique du signal (F0) observable ce qui n'est pas le cas pour les modèles auto régressifs (DAP).

Participants : A. Roebel

1.2.2 Estimation des parties voisées et non voisées

La fréquence de coupure entre les parties voisées et non voisées (VUF) d'un spectre est un descripteur souvent demandé. A sein de l'équipe il existait déjà un algorithme pour l'estimation du VUF qui était fondé sur le calcul de l'énergie du résiduel d'un modèle sinusoïdal. Le problème de cette méthode est que les résultats dépendent d'un façon significative de l'évolution de la fréquence fondamentale. Afin de réduire cette dépendance, nous avons développé un nouvel algorithme qui est basé sur une démodulation locale du signal utilisant une estimation de la fréquence fondamentale préalable. Afin de réduire le temps de calcul l'algorithme ne détermine plus un modèle sinusoïdal explicite mais il utilise la classification des pics spectraux en pics sinusoïdaux et pics bruités. La classification des pics est alors utilisée pour estimer l'énergie du résiduel par bandes fréquentielles et cette estimation est alors utilisée pour déterminer le VUF. La nouvelle approche donne des résultats convaincants avec un temps de calcul presque négligeable.

Participants : A. Roebel, S. Farnier, X. Rodet

1.2.3 Projet VIVOS : transformation et synthèse de voix expressives pour les applications multimédia

Les images de synthèse ont envahi de nombreux domaines multimédias, dessins animés, jeux vidéos et films notamment. Parallèlement à ce phénomène de fond, la voix reste aujourd'hui le parent pauvre en la matière : elle est la plupart du temps simplement

enregistrée par des acteurs, synchronisée souvent de façon « manuelle » avec le mouvement des personnages et n'utilise presque aucune technique de synthèse, sauf à de rares exceptions.

Le but du projet est donc de permettre l'utilisation de voix de synthèse dans le multimédia en général et peut-être dans d'autres applications artistiques comme le théâtre. Ce projet est nommé « transformation et synthèse de voix expressives » car il apparaît que la question de l'expressivité est au centre du projet et conditionne très fortement l'utilisation de voix de synthèse en multimédia. Parmi les problèmes principaux posés à la recherche, citons :

Ce sont des voix spécifiques qui doivent être entendues,

La synthèse doit être de très haute qualité

La destination créative et artistique impose de pouvoir modifier les caractéristiques des voix à volonté en fonction des effets particuliers ou artistiques désirés.

L'objectif que se fixe ce projet nécessite un certain nombre de moyens conceptuels et logiciels sans lesquels la chaîne de production de voix de synthèse ne pourrait fonctionner, et qui sont complémentaires :

Les méthodes:

- 1) Transformation d'identité de voix
- 2) Transformation de type et de nature de voix
- 3) Synthèse de voix expressive
- 4) Synthèse à partir du texte
- 5) Synthèse à partir de corpus d'acteurs et personnages
- 6) Interfaces utilisateur graphiques interactives

Les utilisations :

- 1) Doublage simple, ou avec effets, expressivité
- 2) Synthèse à partir du texte simple, ou avec effets et expressivité
- 3) Post-processing

Ce projet a été élaboré en collaboration avec France-Télécom, l'IRISA, le Studio Chinkel et la société BeTomorrow. Il a été tout particulièrement remarqué et apprécié par le jury d'experts du RIAM et a commencé en 2006.

Une des tâches du projet est de mettre au point un logiciel interactif de traitement et de synthèse à partir du texte. Le traitement est donc fait en synergie avec le développement de SuperVP (pour l'amélioration et l'extension des traitements) et de AudioSculpt (pour les question d'interface utilisateur), la synthèse part du système Talkapillar développé précédemment dans l'équipe.

1.2.3.1 Transformation de type et de nature de voix

Afin d'établir une bibliothèque de transformations de type et de nature de la voix parlée, nous avons effectué les travaux suivants :

- 1) mise en place des méthodes PSOLA et HNM (et SuperVP),
- 2) test perceptif préliminaire de comparaison de ces méthodes en transposition constante,
- 3) programmation d'une première version d'une interface d'utilisateur en Matlab pour transformer le type et la nature de la voix à partir de paramètres intuitifs,
- 4) optimisation des paramètres et étude des combinaisons optimales de SuperVP et PSOLA

pour transformer une voix d'homme en une voix de femme et l'inverse sans artéfact,
5) production des transformations de voix d'homme en voix de femme pour le film "Les amours d'Astrée et Céladon" (Rohmer, 2007).

Participants : X. Rodet, S. Farnier

Collaborations internes : A. Röbel

Collaborations extérieures : France-Télécom, IRISA, Studio Chinkel et société BeTomorrow

1.2.3.2 Transformation de l'identité de la voix

Cette thèse a pour finalité de trouver une fonction de transformation qui soit capable de donner l'identité de la voix d'un locuteur « source » à un locuteur « cible » à l'aide de l'extraction, de l'apprentissage et de la modification des paramètres du signal (fréquence fondamentale, enveloppe spectrale, excitation) liés au locuteur.

Le travail pendant l'année 2006 a compris la mise au point d'un système modulaire d'analyse, alignement, apprentissage et synthèse de voix entre corpus parallèles de parole basé sur la méthode PSOLA. Une étude a été menée sur les techniques d'estimation de l'enveloppe spectrale et l'évaluation d'une méthode pour la sélection du type et de l'ordre optimal du modèle lors de l'estimation pour les signaux harmoniques [Villavicencio07a],[Roebel07a]. L'étude de la problématique pour l'apprentissage des paramètres du locuteur nous a permis d'établir plusieurs voies de recherche qui sont actuellement en cours d'évaluation finale dans le but d'améliorer la transformation de l'identité du locuteur tout en gardant une haute qualité du signal de synthèse.

Participants : F. Villavicencio (thèse), Xavier Rodet.

Collaboration interne : A. Roebel

1.2.3.3 Transformation de l'expressivité d'une voix: Analyse, modélisation, synthèse

Cette thèse, débutée au mois de Septembre 2005, a pour objectif de conférer de l'expressivité à une voix, par transformation. Il s'agit donc de modéliser chaque expressivité par un ensemble de fonctions de transformations des paramètres du signal de la voix (fréquence fondamentale, enveloppe spectrale, source d'excitation).

Les travaux pendant l'année 2006 ont permis de mettre en évidence l'importance du contexte phonétique et pragmatique [Beller06c] :

- mise en place d'un nouveau protocole pour l'enregistrement d'un corpus de parole expressive multi-locuteur (dans le projet VIVOS);
- mise en place d'une nouvelle interface de base de données [Beller06b].
- algorithmes d'estimation et modification de trajectoire de formant.
- étude de signaux électro-glottographiques : estimation du quotient ouvert et des instants d'ouverture/fermeture, concluant à une dépendance du quotient ouvert en fonction du contexte phonétique.
- Etude de l'influence de l'expressivité sur le débit [Beller06a] et de la relation débit/articulation selon les expressivités concluant à la dépendance des variations des paramètres prosodiques en fonction du contexte pragmatique (accentuation).

Les variations de paramètre acoustique de la prosodie (f_0 , énergie, débit, qualité vocale,

degré d'articulation), induites par les expressivités nécessitent une transformation dépendante du contexte pragmatique et phonétique. C'est pourquoi nous commençons à utiliser des réseaux bayesiens afin créer des modèles par apprentissage.

Participants : X. Rodet, G. Beller (thèse).

1.2.3.4 Segmentation automatique de la parole

La synthèse à partir du texte que nous développons est fondée sur la technique de sélection d'unités dans un large corpus. Il faut donc être capable de segmenter automatiquement de la parole enregistrée en phonèmes. Ceci est aussi utile dans tous les traitements de parole. Nous développons donc un logiciel de segmentation automatique en phones par Modèles de Markov Cachés. Le premier travail a porté sur la segmentation monolocuteur et avec la « tool box » HTK. L'apprentissage des modèles a été effectué sur une partie de notre base de données pour la synthèse, un enregistrement d'environ 4 heures de parole. Les premiers tests de segmentation ont été faits par Décodage Acoustico-Phonétique en phonèmes, sans utiliser le texte. Sur un ensemble de phrases de test segmentées à la main, ce logiciel a permis d'obtenir environ 94% de bonnes reconnaissance de phonèmes [Morris06a]. C'est déjà un excellent résultat puisque l'utilisation du texte permettra d'augmenter le taux et de détecter certaines des erreurs pour correction manuelle [Morris06b].

Participants : X. Rodet, A. Morris, C. Veaux, G. Beller (thèse), A. Marty (stage), P. Cyrta (stage)

1.2.3.5 Synthèse à partir du texte

Pour la synthèse à partir du texte, les travaux effectués pendant l'année 2006 ont pris comme point de départ le système de synthèse Talkapillar développé ces dernières années à l'Ircam en cherchant à améliorer principalement la phonétisation et la sélection d'unités.

La phonétisation utilisée jusque là par Talkapillar a été remplacée par le logiciel de phonétisation Liaphon (Laboratoire d'Informatique d'Avignon). Un des intérêts de ce logiciel est d'être basé sur un système de règles, ce qui offre la possibilité de générer différents styles d'élocution en modifiant les règles utilisées pour les liaisons et élisions.

Après un état de l'art des dernières techniques utilisées en synthèse, la sélection d'unité mise en oeuvre dans Talkapillar a été repensée. Nous développons actuellement une procédure de pré-sélection des unités basée sur un apprentissage par arbre de décision. Cette approche permet la modélisation de relations plus complexes entre les espaces symbolique et acoustique, ainsi qu'une validation plus aisée puisqu'il s'agit d'un apprentissage hors-ligne.

Participants : X. Rodet, Christophe Veaux, G. Beller (thèse), A. Marty (stage), P. Cyrta (stage)

1.2.3.6 Apprentissage de la corrélation de l'enveloppe spectrale et de la F0 : application à la transposition de la voix parlée

Notre travail a consisté en l'amélioration des modèles actuels de transposition de la voix parlée – modèles reposant sur la conservation de l'enveloppe spectrale. Nous avons mis en évidence l'importance de l'enveloppe spectrale lors de la transposition sur des exemples,

puis modélisé la corrélation entre la F0 et l'enveloppe spectrale par un apprentissage non-supervisé par classes phonétiques, en montrant de meilleurs résultats que le modèle de conservation [Obin06a]. Les points à améliorer sont les suivants : utilisation d'un analyseur phonétique performant pour une modélisation probabiliste de la classe phonétique en cours, détermination automatique du nombre optimal de gaussiennes, et conservation de la dynamique par la modélisation de la différence à apporter à l'enveloppe source pour atteindre l'enveloppe cible.

Participants : N. Obin (M2 ATIAM), X. Rodet.

Collaboration interne : F. Villavicencio.

1.2.4 Synthèse de parole en temps réel

Le système Talkapillar [Beller06b] permet de créer une voix par une sélection intelligente de fragments sonores dans une base de données, en fonction du texte que l'on souhaite faire prononcer. Le système de synthèse par concaténation d'unités en temps réel « cataRT » [Schwarz06a] a été complété afin de permettre la synthèse vocale en temps réel :

- mise en place d'un séquenceur de descripteurs ;
- interface textuelle permettant le Text-to-Speech en temps réel ;
- Procédure d'échange de données entre Talkapillar et cataRT afin de profiter de segmentations et analyses off-line;
- possibilité d'utiliser la prosodie pour la synthèse de phrases musicales en temps réel ;
- modification graphique de la prosodie en utilisant la méthode PSOLA.

La synthèse de parole en temps réel permet l'exploration immédiate de base de données, l'affichage et la sélection de descripteurs pour des applications artistiques. CataRT-speech est utilisé lors d'un séminaire littérature-musique.

Participants : X. Rodet, G. Beller (thèse).

Collaboration interne : D. Schwarz (équipe Système Temps-Réel).

1.2.5 Analyse et synthèse de chants polyphoniques corses

Ce projet a pour but l'analyse de chants polyphoniques corses et la synthèse de voix chantées, associées aux gestes de la mimophonie. Ces gestes sont captés par deux gants (un pour chaque main) que portent le chanteur.

Les travaux réalisés en 2006 sont :

- 6) grâce à l'analyse de chants, nous avons recréé des voix de chanteurs de polyphonies corses,
- 7) nous avons également créé des gants de données permettant de capter les gestes de la mimophonie,
- 8) puis nous avons établi un langage informatique de la mimophonie afin de faire correspondre à chaque mot du langage, un geste de la mimophonie.

Il nous reste (avec l'équipe Applications Temps Réel) à associer à chaque geste, une voix chantée synthétisée qui soit la plus juste possible.

Participants : X. Rodet, J.S. Gualtieri (Thèse)

Collaborations internes : équipe Applications Temps Réel, E. Flety

Collaborations externes : Université de Corse, Association E Voce di U Cumune

1.3 Traitement par le contenu et nouveaux outils pour la musique

Pendant l'année 2006, nos travaux en traitement par le contenu du signal musical se sont amplifiés, en particulier dans le cadre du groupe de travail avec les compositeurs Orchestration et des projets Semantic HIFI, MusicDiscover et Ecoute. Ils auront également, entre autres buts, celui de fournir de nouveaux outils pour la création, notamment pour permettre de spécifier des traitements dans Super-VP et AudioSculpt ou dans Max/MSP, ou de fournir des annotations automatiques (AudioSculpt).

1.3.1 Extraction d'informations rythmiques pour le projet Semantic HIFI

L'outil d'extraction d'informations relatives au rythme d'un morceau de musique, développé dans le cadre du projet Semantic HIFI, permet la détection du tempo (variable au cours du temps), l'estimation de la position des « battues » et l'estimation de la métrique. L'algorithme repose sur la définition de « templates » mètre-battue-subdivision. Une phase de test intensif sur une très large base de données (les trois bases de données de test utilisées pour ISMIR2004) conduit à apporter des améliorations à l'algorithme initial. Les templates sont revus et permettent maintenant une meilleure discrimination entre les caractéristiques propres à chaque métrique. Ce nouvel algorithme ainsi que son évaluation a fait l'objet d'une publication de Journal [Peeters06b]. Le portage de l'algorithme d'estimation d'informations rythmiques (ircambeat) en C/C++ a été effectué et intégré dans le système final SemanticHIFI.

Participants : G. Peeters

Collaborations internes : P. Tisserand

Collaborations extérieures : Partenaires du projet SHF

1.3.2 Extraction de tonalité pour le projet Semantic HIFI

L'outil d'extraction d'informations relatives à la tonalité et au mode, développé dans le cadre du projet Semantic HIFI, permet une estimation de la tonalité et du mode à partir d'une analyse brute du signal audio. Le système développé en 2005 reposait sur la comparaison de la suite de vecteurs « Profil de Classe de Hauteur » à un ensemble de profils de tonalité dérivés d'expériences cognitives (Krumhansl, Temperley, ...) et étendus aux cas des signaux polyphoniques selon Gomez. Les limitations de cette approche sont une forte

hypothèse sur le profil des amplitudes des harmoniques et une non prise en compte des modulations de tonalité dans un morceau. Comme solution, deux nouvelles approches ont été étudiées en 2006. La première repose sur l'entraînement d'un ensemble de 24 HMMs sur une base de données annotée en tonalité et mode. Une technique de lissage par permutation circulaire des profils est proposée afin d'éviter un sur-apprentissage du système et d'augmenter sa robustesse. La deuxième approche repose sur une extension de la méthode DFT/FM-ACF [Peeters06a] (combinaison d'une représentation de périodicité fréquentielle et d'une représentation temporelle portée dans le domaine fréquentiel) au cas des signaux polyphoniques. Cette méthode appelée Harmonic Peak Subtraction, permet de réduire l'influence des harmoniques supérieures de chaque note sans nécessiter l'estimation explicite des hauteurs et garde donc un coût de calcul réduit. Les deux algorithmes ont été testés positivement sur une large (300 morceaux) base de données de musique baroque, classique et romantique incluant de la musique pour clavier, orchestre symphonique et orchestre de chambre. Ces approches ont fait l'objet de deux publications de congrès [Peeters06c] et [Peeters06d]. Dans les deux systèmes, un pré-processeur permet d'estimer le diapason global du morceau (écart par rapport à un diapason de 440Hz). Ce pré-processeur a fait l'objet d'un contrat avec la société Make Music en temps que logiciel d'estimation de diapason (ircamtuning). Le portage de l'algorithme d'estimation de tonalité et mode (ircamkeymode) en C/C++ a été effectué et intégré dans le système final SemanticHIFI.

Participants: G. Peeters

Collaborations internes : P. Tisserand

Collaborations externes : participants du projet SHF

1.3.3 Projet Ecoute

Le projet Ecoute démarré courant 2006 propose une chaîne éditoriale hypermédia permettant de constituer des collections sonores musicales, de les indexer finement au niveau inter-document et aussi intra-document. Dans ce projet, l'équipe Analyse-Synthèse est en charge du développement des technologies d'indexation. En 2006, les recherches se sont essentiellement concentrées sur deux points : segmentation automatique d'un flux audio de type radiophonique, catégorisation d'un morceau de musique en genre musicale.

1.3.3.1 Développement d'un système générique d'indexation

Les deux applications visées reposent sur des techniques d'indexation (l'une temporelle dans le flux, l'autre globale au morceau) il a été décidé de développer un système unique permettant de gérer ces deux cas de figures. Le système se veut générique et modulaire et est paramétré spécifiquement pour chaque application. Le système développé est basé sur le celui qui avait été développé dans le cadre du projet CUIDADO pour l'indexation des échantillons en catégories de type « instrument de musique ». Ce système comporte une étape d'apprentissage et une étape de classification. L'apprentissage lui même comporte trois étapes : extraction des descripteurs, sélection automatique des descripteurs

(permettant l'adaptation à tout nouveau problème) modélisation des classes. Le système est réécrit de manière à 1) faciliter la définition de nouveaux problèmes (communication avec l'extérieur au travers de simples fichiers textuels, possibilité de re-diriger des classes donc possibilité de hiérarchisation des classes, vérification des fichiers pour contrôle de cohérence) 2) permettre l'indexation temporelle (lecture de définition et écriture de classes temporelles au format wavesurfer) 3) permettre l'indexation de fichiers longs (jusqu'à 24heures pour Dalet) 4) modulariser la partie extracteur (possibilité de lier de nouveaux extracteurs). Le système inclut les procédures de tests habituelles (« ten-fold », « cross-database validation »). Les modélisations statistiques ont été étendues de manière à inclure les type GMM et HMM. Enfin les algorithmes de sélection de caractéristiques (IRMFSP) et d'analyse linéaire discriminante (LDA) ont été réécrits de manière à réduire l'influence d'un éventuel déséquilibre de représentation des classes lors de l'apprentissage.

Participants : G. Peeters

Collaboration interne : partenaires du projet Ecoute.

1.3.3.2 Segmentation automatique d'un flux radiophonique en catégorie Speech/Music pour le projet Ecoute

La segmentation de flux radio s'effectue avec le partenaires industriel Dalet. Celui-ci spécifie les besoins et fournit des bases audio représentatives des radios de ses clients. Il s'agit donc d'un travail sur des données applicatives réelles et sur des définitions de classes industrielles. Pour le développement des technologies, les corpus suivants sont utilisés : RadioFrance, Key103 (radio UK de Manchester) et Traffic FM. Les catégories proposées par Dalet sont subdivisées en catégories du métier et en catégories audio. L'ensemble forme 7 classes différentes. Le système présenté ci-dessus est appliqué sur un ensemble générique de caractéristiques. Le « rappel » moyen obtenu est de 58% pour les 7 classes et monte à 95.4% en considérant les seules classes musique et parole. Un ensemble important de confusions est observé entre certaines catégories. Des études préliminaires ont été effectuées afin de remplacer l'indexation de certaines catégories par un système d'identification audio. La généralisation du système est testée en entraînant le système sur RadioFrance et en l'appliquant pour l'indexation de Key103 (deux radios très différentes). Le « rappel » moyen obtenu est de 93%.

Participants : G. Peeters

Collaboration externe : partenaires du projet Ecoute.

1.3.3.3 Indexation en genre musicale pour le projet Ecoute

L'indexation automatique de morceau de musique en genre et humeur musical (music genre, music mood) s'effectue avec le partenaire industriel WMI. Celui-ci spécifie les catégories considérées par ses clients et fournit une base d'exemples de ces catégories. En raison d'un retard de livraison de cette base, seule une étude préliminaire a été effectuée sur des bases de données publiques issues de la communauté ISMIR. Six catégories sont considérées, 'classical', 'electronic', 'jazz_blues', 'metal_punk', 'rock_pop', 'world'. Le système présenté ci-dessus est appliqué à ce problème et permet d'atteindre un « rappel » moyen de 78% pour la base ISMIR2004. Un nouveau type de classifieur à deux étages est proposé. Le premier

étage effectue une classification préliminaire de chaque trame du signal de manière indépendante. Le deuxième étage modélise l'histogramme d'appartenance de toutes les trames d'un morceau à toutes les classes. Ceci permet de tenir compte de la non-homogénéité acoustique d'un morceau.

Participants : G. Peeters

Collaboration interne : partenaires du projet Ecoute.

1.3.3.4 Extraction automatique de suites d'accords à partir de l'analyse d'un signal audio musical pour le projet Ecoute

Ce travail sur l'estimation de suites d'accords d'un morceau de musique continue celui sur l'estimation de tonalité et de mode. Il a commencé dans le cadre d'un stage de master recherche [PAPADOPOULOS06a] et se poursuit au travers de la thèse de H. Papadopoulos. Cette thèse vise à estimer l'ensemble des paramètres de contenu musical (tempo, métrique, tonalité, accords, mélodie, ...) de manière conjointe. Cette thèse s'inscrit dans le cadre du projet Ecoute. En 2006, la recherche s'est concentrée sur la détection automatique de suites d'accords d'un signal audio musical polyphonique complexe. Le système développé comprend une première partie qui effectue l'extraction de vecteurs représentant l'importance des différentes hauteurs à un instant donné (vecteur de « chroma »). Ces informations sont de nature probabiliste ; l'objectif est d'estimer la suite d'accords au cours du temps la plus probable. Plusieurs méthodes ont été proposées et comparées. Ces méthodes sont basées sur des modèles de Markov cachés et se distinguent les unes des autres selon la manière d'aborder différents points (probabilités d'observation, matrice de transition, apprentissage). L'évaluation du système a été faite sur une base de données relativement large (110 morceaux de musique populaire).

Participants: H. Papadopoulos (stage), G. Peeters

Collaborations internes : C. Yeh

Collaborations externes : participants du projet Ecoute

1.3.4 Montage du projet Quaero de l'All

L'équipe Analyse/Synthèse de l'Ircam est fortement impliquée dans le montage du projet Quaero de l'All. Dans ce montage, l'Ircam coordonne les activités d'indexation musicale du projet.

Participants : G. Peeters, X. Rodet

Collaborations internes : H. Vinet

Collaborations extérieures : ENST, IRISA, LIMSI, Thomson, FT, Exalead

1.3.5 Reconstruction de timbre pour l'orchestration

Le projet, commencé en 2005, a pour but de construire des outils informatiques pour aider les compositeurs à orchestrer leurs oeuvres (Groupe de Travail Orchestration). La première approche est la suivante : l'utilisateur fournit un son cible et un ensemble d'instruments

disponibles, le système doit alors trouver la combinaison d'instruments qui se rapproche le plus du son cible. Une première interface graphique a été réalisée sur la base de l'algorithme de 2005 pour permettre à des compositeurs de tester le système et ainsi confirmer l'approche choisie. Ensuite, un nouveau système a été développé. Ce système est basé sur des méthodes d'apprentissage automatique. Cela permet d'une part de généraliser les connaissances extraites des bases d'échantillons, c'est-à-dire d'apprendre les possibilités d'un instrument en général et non d'une instance particulière de cet instrument jouée par un instrumentiste donné. D'autre part, cela permet d'extrapoler des connaissances absentes des bases d'échantillons, comme des hauteurs manquantes ou des combinaisons de modes de jeu inédites. Un ensemble de nouveaux descripteurs du timbre a été introduit, permettant par exemple de prendre en compte les modulations de fréquence et d'amplitude, ou encore la rugosité des sons. La collaboration avec G. Carpentier a permis de développer un algorithme génétique capable de trouver des combinaisons contenant un grand nombre d'instruments en un temps très réduit et de mettre en place les bases nécessaires à l'interaction entre le compositeur et le système.

Participant : D. Tardieu (thèse), X. Rodet, G. Peeters

Collaborations internes : Equipe Représentation musicale (Grégoire Carpentier), Equipe Service en Ligne, Groupe Orchestration (Y. Maresz, G. Drouin), projet MusicDiscover

1.3.6 Projet MusicDiscover : Bases de données audio pour l'apprentissage et la référence

Dans les années passées, nos prévisions concernant le développement et l'utilisation de bases de données se sont confirmées. C'est en particulier le sujet du projet MusicDiscover de l'ACI « Masse de Données » obtenu en 2004 par l'équipe pour 3 ans (jusqu'à 2007) et en collaboration avec d'autres centres de recherche. Les principaux points traités en 2006 sont :

- constitution de bases de données polyphoniques de référence ;
- mise en formes des fichiers, analyses et ajouts de méta-données ;
- utilisation, pour le projet de reconnaissance des instruments, d'estimation de fondamentaux multiples.

Trois travaux de recherche doctorale participent à ce projet. La question de la description structurée des méta-données extraites du signal audio sera abordée en 2007.

Participants : C. Yeh (Thèse), A. Livshin (Thèse), D. Tardieu (thèse)

Collaborations internes : équipes Systèmes Informatiques et Services en Ligne

Collaborations extérieures : G. Richard (ENST), L. Chen (LIRIS).

1.3.6.1 Reconnaissance des instruments

Ce travail de thèse est mené dans le cadre du projet MusicDiscover de l'ACI Masse de Données. Le son des instruments de musique entretenus (tels que les vents, les bois, etc.) est composé d'une série de partiels harmoniques et d'une partie non-harmonique nommée parfois *résiduel*. Nous avons étudié la possibilité de reconnaître automatiquement les instruments musicaux en utilisant seulement la série harmonique, car cela permettrait de simplifier considérablement la reconnaissance automatique des instruments dans la musique

polyphonique. Pour des sons isolés de 10 instruments, [Livshin06a] a montré que des niveaux de reconnaissance très proches pouvaient être obtenus en utilisant seulement la série harmonique (91%) ou le signal entier (95%). Dans [Livshin06c], en employant seulement le résiduel non-harmonique nous avons atteint un taux de reconnaissance de 78%. Aussi, sur les partiels harmoniques seulement, en employant le programme d'estimation automatique de F0 de C. Yeh (voir ci-dessus le paragraphe Estimation de fréquences fondamentales multiples), nous avons atteint, sur des mixages polyphoniques et multi-instrumentaux de 5 instruments différents, des taux de reconnaissance d'instruments de 69%, 60%, 58%, 40% pour 2 à 5 instruments respectivement.

Participants : G. Peeters, A. Livshin (thèse), J. J. Burred (stage)

Collaborations internes: C. Yeh (thèse).

Collaborations extérieures : G. Richard (ENST), L. Chen (LIRIS).

1.3.7 Collaboration multimodale pour l'indexation

Cette thèse poursuivie également au CEA pendant les années précédentes a été soutenue en Avril 2006 [Delezoide06]. Elle a permis d'obtenir notamment un système intégré d'indexation incluant la parole, le son et la vidéo et prenant en compte la multimodalité des documents [Delezoide06a], [Delezoide06b]. En particulier, les travaux sur la séparation musique/parole pourront être utilisés dans d'autres projets (Ecoute, MusicDiscover).

Participants : B. Delezoide (thèse)

Collaboration extérieure : C. Fluhr (CEA)

1.4 Modélisation physique : simulation en temps-réel et problèmes inverses

La modélisation physique a connu un développement important dans l'équipe, notamment grâce à la composante UMR-CNRS, avec un chargé de recherche et les travaux de trois stagiaires dont deux (Mastère ATIAM) poursuivent en doctorat (une bourse EDITE et une bourse conjointe CNRS-CEA), et le projet blanc ANR-CONSONNES. Les travaux concernent l'analyse-synthèse de phénomènes non linéaires, des débouchés vers des implémentations en temps réel, les méthodes d'inversion entrée/sortie de modèles d'instruments et le début d'un projet « mécatronique » associant mécanique des instruments et commande électronique.

1.4.1 Ondes progressives dans les tubes à section variable

Un travail a été présenté au Congrès Français d'Acoustique sur les « Ondes découplées et ondes progressives pour les problèmes mono-dimensionnels d'acoustique linéaire » [Helie06c]. Il a permis d'obtenir des décompositions systématiques d'ondes acoustiques (notamment dans les tubes à sections variables) en la somme de deux ondes progressives

(aller/retour) même pour des géométries non triviales. L'intérêt sera d'aboutir dans un second temps à des systèmes modulaires et faible coût pour simuler des résonateurs d'instruments à vent.

Le stage de Rémi Mignot [Mignot06a] a permis d'avancer significativement dans l'obtention de structures simulables en temps réel pour les résonateurs d'instruments à vent: des structures dites de « Kelly-Lochbaum » très économiques pour la synthèse en temps réel ont été obtenues pour le cas de tubes droits, coniques ou évasés de courbure constante (type pavillon). Ce stage se poursuit depuis octobre 2006 par une thèse (dirigée par D. Matignon, ENST et encadrée par Th. Hélié-CNRS-Ircam).

Participant : Th. Hélié, Rémi Mignot (stage)

Collaboration extérieure : D. Matignon (ENST), Projet blanc ANR-CONSONNES

1.4.2 Représentations intégrales pour phénomènes à mémoire longue et pertes visco-thermiques

Les pertes visco-thermiques et la variation de section des tubes acoustiques ont un effet important, y compris du point de vue perceptif. Mais ils introduisent des opérateurs non-triviaux et difficiles à simuler. Depuis quelques années, des résultats théoriques et méthodes d'approximation ont permis d'aboutir à des simulations efficaces permettant le temps réel pour la synthèse sonore. Deux articles ont été publiés en revue [Helie06d, Helie06e] et un troisième en congrès [Helie06f]. Des recherches visant à encore améliorer les approximations et, si possible, garantir l'optimalité sont toujours en cours.

Participant : Th. Hélié, R. Mignot (stage)

Collaboration extérieure : D. Matignon (ENST), Projet blanc ANR-CONSONNES

1.4.3 Convergence et erreur garantie pour systèmes faiblement non linéaires

Les séries de Volterra sont un outil que nous utilisons beaucoup en synthèse depuis quelques années pour simuler des phénomènes faiblement non linéaires (cf. ci-dessous). Leur efficacité est reconnue dans la communauté scientifique mais aucun résultat garantissant la convergence ou la bornitude de l'erreur due à une troncature (ce qui est toujours le cas en pratique) n'était encore prouvé. Un travail initié en 2005 vient de permettre l'obtention de tels résultats, d'abord sur les systèmes différentiels avec une non-linéarité d'ordre 2. L'intérêt est aujourd'hui de pouvoir contrôler quantitativement l'erreur commise et ainsi d'affiner en connaissance de cause la simulation temporelle de ces systèmes. Un article de revue a été soumis en 2006 (parution probable en 2007 dans International Journal of Control).

Participant : T. Hélié

Collaboration extérieure : B. Laroche (Laboratoire des Signaux et Systèmes-Université d'Orsay)

1.4.4 Propagation non linéaire

Ce phénomène apparaît dans de nombreux systèmes acoustiques ou vibratoires dès que les niveaux sont suffisamment forts. C'est un thème de recherche très actif: (1) les résultats obtenus par les séries de Volterra dans le cas du tube droit (cuivrage des sons) sont l'objet d'une demande de dépôt de brevet au CNRS; (2) des recherches préliminaires ont été poursuivies pour étendre ces résultats au cas de tubes à section variable dans le cadre d'un stage (M. Rébillat, ENS Cachan); (3) des résultats sur le cas de l'équation de Westervelt (propagation acoustique non linéaire dans un milieu amorti) ont été présentés en congrès [Helie06b]; (4) le cas d'une corde en forte déformation a été simulé par les séries de Volterra lors du stage ATIAM de D. Roze [Roze06a] qui poursuit en thèse (bourse CNRS-CEA); (5) un algorithme de résolution rapide et exacte d'une classe d'ondes progressives amorties avec propagation non linéaire est à l'étude en collaboration (Th. Hélie, C. Vergez).

Participant : T. Hélie, M. Rébillat (stage), D. Roze (stage), J. Bensoam

Collaboration extérieure : C. Vergez (LMA-CNRS, Marseille)

1.4.5 Simulation de circuits faiblement non linéaires

Une simulation faible coût d'un circuit analogique faiblement non linéaire a été proposée et présentée en congrès [Helie06a] sur le cas d'un filtre VCF Moog. Des versions temps-réel sous forme de PLUG IN VST ont depuis été implantées. Les résultats encourageants sont à l'origine d'un sujet de stage pour 2007 sur les (pré-) amplificateurs de guitares et autres modèles physiques de circuits audio.

Participant : Th. Hélie

Collaboration interne : R. Muller (stage)

1.4.6 Problèmes inverses entrée/sortie

Piloter un instrument à vent virtuel pour qu'il reproduise un son donné, autrement dit, inverser le système reste encore aujourd'hui un enjeu. Plusieurs techniques ont déjà été étudiées à l'IRCAM (approche mathématique et par apprentissage). Ce problème mal posé et difficile est toujours à l'étude. L'étude d'approches utilisées en automatique et robotique (fonction de Lyapunov, platitude, etc) ont débuté sur un modèle « instrument type » simplifié pour mettre au point une technique robuste.

Participant : T. Hélie

Collaborations extérieures : B. d'Andréa-Novet (ENSMP), Projet blanc ANR-CONSONNES

1.4.7 Robotisation de bouche artificielle pour cuivres

Depuis octobre 2006, un projet mécatronique a été mis en place en partenariat entre l'Ecole des Mines de Paris et l'équipe Acoustique-Instrumentale. Le projet consiste à piloter les éléments asservissant les paramètres (masse, raideur, amortissement) des lèvres

synthétiques (chambres en silicone) et du jeu (pression de bouche, appui de l'instrument sur les lèvres, enfoncement des pistons d'une trompette). Le but est de disposer d'une maquette permettant de faire des expériences automatisées, contrôlées et reproductibles sur le jeu des cuivres. La première version prototype devrait être terminée pour mai 2007.

Participant : T. Hélié

Collaborations internes : équipe Acoustique-Instrumentale, Atelier de mécanique

Collaborations extérieures : J. Senpauroca (ENSMP), Projet blanc ANR-CONSONNES

1.5 Développements

Un des changements majeurs qui ont marqué les développements de cette année est l'arrivée des processeurs Intel dans les ordinateurs Apple. Les modifications des logiciels qui ont été nécessaires afin d'assurer un fonctionnement correct sur les deux architectures sont considérables, surtout parce que le compilateur CodeWarrior n'est plus maintenu, ce qui nous a obligé à changer de compilateur pour le logiciel AudioSculpt. De plus, afin de supporter aussi les architectures Powerpc, il était nécessaire de mettre au point un mécanisme de création de binaires universels pour toutes les bibliothèques et exécutables.

De nouveau, comme en 2005, nous avons eu plusieurs contrats de l'extérieur notamment avec la société américaine MakeMusic. Les contrats concernent notamment les méthodes d'estimation de la fréquence fondamentale et de détection des frappes de batterie. Ces demandes nécessitent des systèmes de compilation autres que le système GNU, et nous avons donc été amenés à changer l'organisation de nos sources afin de pouvoir plus facilement adapter les configurations à d'autres compilateurs et systèmes d'exploitation.

1.5.1 Configuration des projets avec cmake

Comme décrit ci-dessus, des demandes fréquentes de l'extérieur nous obligent à supporter tous les compilateurs et systèmes d'exploitation utilisés aujourd'hui. Nous nous sommes donc intéressés aux méthodes de configuration du code source et de génération des configurations de compilateur plus portables que les fichiers make et le système autoconf. Après quelques tests initiaux le choix s'est porté sur le configurateur de système de compilation cmake (<http://www.cmake.org>). Ce générateur de configuration peut générer des configurations pour un nombre impressionnant de compilateurs actuels. Au cours de l'année 2006, la plupart des bibliothèques et logiciels ont été mis sous la gestion de cmake.

Participant : A. Roebel

1.5.2 Logiciel SuperVP

Plusieurs modules dans le logiciel SuperVP ont été améliorés et un certain nombre de fonctionnalités ajoutées. Pour l'organisation du code, nous avons introduit un niveau hiérarchique de modules basé sur la dérivation des objets. La nouvelle organisation nous a permis une simplification significative du code.

Concernant les fonctionnalités, les changements les plus importants sont:

1. le support des calculs en plusieurs « threads » d'exécution (traitements parallèles),
2. l'utilisation de la classification sinusoïde/bruit pour le re-mixage des composants,
3. un nouveau mode de traitement des sons stéréos qui permet la préservation de l'image stéréo, une amélioration de la détection des transitoires (onsets) qui a été évaluée, avec des résultats très satisfaisants, lors de MIREX 2006,
4. l'intégration d'un nouvel algorithme pour la détection de la fréquence fondamentale pour les instruments percussifs,
5. la mise en place d'un mécanisme pour la gestion de paramètres en temps réel,
6. une réorganisation du module de ré-échantillonnage et de la gestion des paramètres des transpositions dynamiques qui améliore la qualité de cette transformation d'une façon significative.

En collaboration avec l'équipe IMTR, nous avons créé plusieurs modules Max/MSP qui, basés sur la bibliothèque dynamique de SuperVP, offrent aux utilisateurs de Max/MSP le bénéfice de fonctionnalités de SuperVP.

Participant : A. Roebel

Collaborations intérieures : N. Schnell (équipe IMTR)

1.5.3 Bibliothèque PM2 et modèle sinusoïdal

En plus les travaux liés au problème des binaires universels et de l'utilisation du gestionnaire de projet cmake, l'implémentation de l'estimateur des paramètres des sinusoïdes a avancé en 2006 avec une version initiale de la réduction du biais de l'estimateur pour le cas d'une fréquence fondamentale variable. Les résultats obtenus avec la nouvelle version de l'estimateur concordent avec les résultats obtenus lors de l'évaluation expérimentale (voir partie recherche ci-dessus). Les améliorations des estimations de paramètres de partiels se traduisent dans une réduction de l'énergie du résiduel du modèle qui peut aller jusqu'à -3dB pour les sons avec vibrato. L'estimateur implémenté actuellement correspond à une version initiale de la recherche et il sera encore amélioré dans l'avenir proche.

Une deuxième amélioration que nous avons implémenté est la synthèse et la pré-écoute des modèles sinusoïdaux en temps réel qui sera bientôt accessible dans l'application AudioSculpt.

Participant : A. Roebel

1.5.4 Logiciel AudioSculpt

AudioSculpt est un logiciel d'analyse, transformation et synthèse sonore, utilisant les noyaux de calcul SuperVP et Pm2. En 2006 les versions 2.6 et 2.7 ont été publiées. Ils introduisent des nouveautés, ainsi qu'une amélioration de la stabilité et l'ergonomie. Le code source a été considérablement modernisé, et la transition vers le système de développement XCode et les processeurs Intel est conclue.

Les nouveautés de 2006 comprennent :

- 1) Copier-coller de régions définies en temps et fréquence sur le sonagramme

- 2) Ecoute de zones marquées sur le sonagramme
- 3) Annotation des sons par note MIDI et texte
- 4) Crayon harmonique
- 5) Sélection par niveau d'amplitude (baguette magique)
- 6) Préservation partielle de l'enveloppe spectrale
- 7) Nouvelles analyses (VUV, « pitched percussion »)
- 8) Réduction de bruit en temps réel

Participant : N. Bogaards

Collaborations internes : A. Roebel

Collaborations extérieures: A. Lithaud (compositeur)

1.5.5 Bibliothèques SDIF et Easdif

Afin de faciliter l'accès des utilisateurs extérieurs, les bibliothèques SDIF et Easdif ont été transférées au serveur de projets de logiciels libres Sourceforge. La préparation de ce transfert a été assurée par D. Schwarz (IMTR). Les travaux de développement ont été concentrés autour du support SWIG pour SDIF et Easdif. Grâce aux avancées récentes du projet SWIG, l'interface Easdif<->SWIG a pu être amélioré.

Pour faciliter l'importation des données stockées en SDIF dans des logiciels ne supportant pas SDIF, les logiciels SDIFConverter et Pm2Convert ont été développés, qui sont capables de convertir le plupart des analyses provenant de AudioSculpt, SuperVP et Pm2 en texte.

Participant : N. Bogaards, A. Roebel

Collaborations internes : D. Schwarz (IMTR) , J. Bresson (RepMus)

1.5.6 Xspect

Le logiciel Xspect est une application qui est indispensable pour les travaux de l'équipe. Pour répondre aux demandes, en relation notamment avec le projet VIVOS, plusieurs nouvelles analyses ont été ajoutées à Xspect. En particulier:

1. lectures des fichiers SDIF,
2. écoute des régions temps-fréquence,
3. analyse F0 sur un segment temporel,
4. interface donnant accès à un nouveau estimateur de F0.

En 2006 nous avons porté Xspect sur MacOSX et Xspect a été intégré dans le catalogue des logiciels de l'équipe Analyse-Synthèse du Forum IRCAM.

Participant : A. Roebel, N. Bogaards, X. Rodet

Publications :

Articles parus dans une revue à comité de lecture

[Helie06d] Hélie, T., Matignon, D., « Diffusive Representations for Analyzing and Simulating Flared Acoustic Pipes with Visco-thermal Losses », *Mathematical Models and Methods in Applied Sciences*, Avril 2006, vol. 16, n° 4, pp. 503-536

- [Helie06e] Hélie, T., Matignon, D., « Representation with poles and cuts for the time-domain simulation of fractional systems and irrational transfer functions », *Journal of Signal Processing, Special Issue on Fractional Calculus Applications in Signals and Systems*, 2006, vol. 86, pp. 2516-2528
- [Peeters06b] Peeters, G., « Template-based estimation of time-varying tempo », .» *EURASIP Journal on Advances in Signal Processing, Special Issue on Music Information Retrieval Based on Signal Processing*, Décembre 2006
- [Roebel06a] Roebel, A., « Adaptive Additive Modeling with Continuous Parameter Trajectories », *IEEE Transactions on Speech and Audio Processing*, Juillet 2006, vol. 14, n° 4, pp. 1440-1453
- [Vincent06a] Vincent, E., « Musical source separation using time-frequency source priors », *IEEE Trans. on Audio, Speech and Language Processing*, 2006
- [Vergez06a] Vergez, C., Tisserand, P., « The BRASS Project, from Physical Models to Virtual Musical Instruments: Playability Issues », *Lecture Notes in Computer Science*, Mai 2006, vol. 2006, n° 3902, pp. 24-33

Actes de congrès ou de colloque avec comité de lecture

- [Beller06a] Beller, G., Hueber, T., Schwarz, D., Rodet, X., « Speech Rates in French Expressive Speech », *Speech Prosody 2006*, Dresden, 2006
- [Burred06a] Burred, J., Roebel, A., Rodet, X., « An Accurate Timbre Model for Musical Instruments and its Application to Classification », *Workshop on Learning the Semantics of Audio Signals*, vol. 1, Athens, 2006
- [Boutard06a] Boutard, G., S. Goldszmidt, et al. (2006). Browsing inside a Music Track, the Experimentation Case Study. Workshop on Learning the Semantics of Audio Signals (LSAS), Athens, Greece
- [Delezoide06a] Delezoide, B., « Multimedia movie segmentation using low-level and semantic features », *ACM Multimédia*, 2006
- [Delezoide06b] Delezoide, B., « Multimedia movie segmentation using low-level and semantic features », *Axmedia*, 2006
- [Helie06a] Hélie, T., « On the use of Volterra series for real-time simulations of weakly nonlinear analog audio devices: application to the Moog ladder filter », *DAFx*, Montréal, 2006
- [Helie06b] Hélie, T., « Résolution d'une équation des ondes faiblement non linéaire par les séries de Volterra et décomposition modale », *CFA*, Tours, 2006
- [Helie06c] Hélie, T., « Ondes découplées et ondes progressives pour les problèmes mono-dimensionnels d'acoustique linéaire », *CFA*, Tours, 2006
- [Helie06f] Hélie, T., Matignon, D., Mignot, R., « Criterion design for optimizing low-cost approximations of infinite-dimensional systems: towards efficient real-time simulation », *13th IFAC workshop on Control Applications of Optimisation*, Cachan, 2006
- [Kaprykowsky06a] Kaprykowsky, H., Rodet, X., « Globally Optimal Short-Time Dynamic Time Warping Application to Score to Audio Alignment », *ICASSP*, Toulouse, 2006
- [Livshin06a] Livshin, A., Rodet, X., « The importance of the non-harmonic residual for automatic musical instrument recognition of pitched instruments », *AES Convention 120*, Paris, 2006
- [Livshin06c] Livshin, A., Rodet, X., 2006. "The Significance of the Non-Harmonic "Noise" Versus the Harmonic Series for Musical Instrument Recognition," In Proceedings of the International Symposium on Music Information Retrieval (ISMIR'06).
- [Peeters06a] Peeters, G., « Music Pitch Representation by Periodicity measures based on Combined Temporal and Spectral Representations », *ICASSP*, Toulouse, 2006
- [Peeters06c] Peeters, G., « Musical key estimation of audio signal based on HMM modeling of chroma

vectors », *DAFX*, Montreal, 2006

[Peeters06d] Peeters, G., « Chroma-based estimation of musical key from audio-signal analysis », *ISMIR*, Victoria, 2006

[Roebel06b] Roebel, A., « Estimation of partial parameters for non stationary sinusoids », *International Computer Music Conference (ICMC)*, New Orleans, 2006, pp. 167-170

[Roebel06c] Roebel, A., « Onset Detection in Polyphonic Signals by means of Transient Peak Classification : mIREX 2006 -- Onset Detection », *International Symposium on Music Information Retrieval/MIREX*, Victoria, 2006

[Tardieu06a] Tardieu, D., Carpentier, G., Assayag, G., Rodet, X., Saint-James, E., « IMITATIVE AND GENERATIVE ORCHESTRATIONS USING PRE-ANALYSED SOUNDS DATABASES », *SMC*, Marseille, 2006

[Villavicencio06b] Villavicencio, F., Röbel, A., Rodet, X., « Improving LPC spectral envelope extraction of voiced speech by True-Envelope estimation », *ICASSP*, Toulouse, 2006

[Yeh06a] Yeh, C., Roebel, A., « Adaptive noise level estimation », *9th Int. Conf. on Digital Audio Effects (DAFx'06)*, 2006

[Yeh06b] Yeh, C., Roebel, A., Rodet, X., « Multiple F0 tracking in solo recordings of monodic instruments », *120th AES Convention*, 2006

Actes de congrès ou de colloque sans comité de lecture

[Beller06b] Beller, G., marty, a., « Talkapillar: outil d'analyse de corpus oraux », *Rencontres Jeunes Chercheurs de l'Ecole Doctorale 268*, paris, 2006

[Beller06c] Beller, G., « Analyse et Synthèse de l'Expressivité dans la Voix Parlée », *JJCAAS*, Lyon, 2006

[Yeh06c] Yeh, C., Roebel, A., « Adaptive noise level estimation », *Workshop on Computer Music and Audio Technology (WOCMAT'06)*, Taipei, 2006

Travaux universitaires (thèses, mémoires) et rapports de stage

[Delezoide06c] Delezoide, B., « Modèles d'indexation multimedia pour la description automatique de films de cinéma », UNIVERSITÉ PARIS VI – PIERRE ET MARIE CURIE, 2006

[Hamnane06a] Hamnane, S., « Traitement et représentation temps-fréquence des sons avec résolution adaptative », Paris 6, 2006. [Master M2 ATIAM]

[Mignot06a] Mignot, R., « Modélisation de résonateurs d'instruments à vent », Paris 6, 2006. [Master M2 ATIAM]

[Obin06a] Obin, N., « Apprentissage de la corrélation de la F0 et de l'enveloppe spectrale : application à la transposition de la voix parlée », Paris 6, 2006. [Master M2 ATIAM]

[PAPADOPOULOS06a] PAPADOPOULOS, H., « Extraction automatique d'une suite d'accords à partir de l'analyse d'un signal audio musical », Université de Cergy-Pontoise / ENSEA, 2006. [Master SIC / diplôme d'ingénieur]

[Roze06a] Roze, D., « Simulation d'une corde avec fortes déformations par les séries de Volterra », Paris 6, 2006. [Master M2 ATIAM]

Rapports de recherche ou de fin de contrat

[Morris06a] je vais le rajouter ...

[Morris06b] je vais le rajouter ...

Conférences invitées

Roebel, A., « Spectral envelope estimation and its applications for pitch shifting and sound transformation », Inventionen 2006, Berlin, 2006.

Roebel, A., « Spectral envelope estimation and its applications for pitch shifting and sound transformation », Lange Nacht der Wissenschaften de l'Université Technique de Berlin, 2006.

Barroso, E., Roebel, A., « Spectral Modelling Synthesis », Inventionen 2006, Berlin, 2006.

Colloques et séminaires

Séminaire Ircam (21/03/2006) Peeters.Barthélémy, Orchestration

Conférence European (24/04/2006) G. Peeters "Caractérisation du timbre des sons musicaux"

Conférence Ircam Portes Ouvertes (11/06/2006), H. Vinet, G. Peeters "Projet Européen Semantic HIFI"

Conférence Journée Valorisation du patrimoine de la création musicale (15/06/2006), G. Peeters "Systèmes d'identification et de navigation intra-documentaires"

Séminaire Ircam (13/09/2006), Peeters/Papadopoulos, « Détection tonalité, accords »

Roebel, A., « Transiente, Sinusoide und Rauschen : Detektion von Signalkomponenten und signalangepasste Transformation » (transitoires, sinusoides, bruits : détection des composantes et transformation adapté au contenu du signal), séminaire de recherche a l'institut de la science de la communication, 2006.

Bloit, J., "Modélisation et reconnaissance en temps réel d'événements musicaux", séminaire Ircam, décembre 2006.

Bloit, J., "Modélisation et reconnaissance d'événements musicaux", journées des jeunes chercheurs en audition, acoustique musicale et signal audio, INSA, Lyon, octobre 2006.

X. Rodet, R. Cahen : UTILISATION DE L'HAPTIQUE DANS L'EXPLORATION MUSICALE - PROJET PHASE: PLATEFORME HAPTIQUE D'APPLICATION SONORE POUR L'ÉVEIL MUSICAL,, Séminaire : MUSIQUE ET TECHNOLOGIES / RECHERCHE ET PÉDAGOGIE, Ircam, 16 Juin 2006

Jurys de thèse

X. Rodet, rapporteur de la thèse de J.-J. Aucouturier, Juin 2006, ENST.

X. Rodet, membre du jury de la thèse de G. Picard, Novembre 2006, ENST.

X. Rodet, membre du jury de la thèse de C. Dubois, Novembre 2006, Ecole Centrale de Nantes.

Review d'articles et propositions

Emissions radiophoniques et télévisées, entretiens journalistiques, animations

G. Peeters : Télévision (13/01/2006) France 3, 19/20, édition Ile-de-France

Autres activités de formation

Thèses soutenues

[Delezoide06c] Delezoide, B., « Modèles d'indexation multimedia pour la description automatique de films de cinéma », UNIVERSITÉ PARIS VI – PIERRE ET MARIE CURIE, Avril 2006

Activités d'enseignement

Roebel, A., « Analyse, Modellierung und Transformation von Audiosignalen » (Analyse, Modélisation

et Transformation des signaux audios), séminaire et travaux dirigés donnés lors du séjour comme professeur invité (Edgar-Varèse professor for computer music) à l'Université Technique de Berlin en Allemagne, 48h au semestre d'été, 2006.

J. Bloit, vacances Paris 6 : Initiation à l'informatique pour scientifiques (TPs, 25h), Programmation récursive (TPs, 12h)

G.Beller, moniteur Paris 8: Initiation à l'acoustique musicale (cours, 32H), Initiation à l'audio numérique (32H).

Projets et Contrats

Projet RIAM VIVOS «Transformation et synthèse de voix expressives pour les applications multimédia »

Ce projet (2006-2008) dont l'Ircam est leader, est mené avec France-Télécom, l'IRISA, le Studio de doublage Chinkel et la société de jeux vidéos BeTomorrow. Le but du projet est de permettre l'utilisation de voix de synthèse dans le multimédia en général et peut-être dans d'autres applications artistiques comme le théâtre. Ce projet est nommé « transformation et synthèse de voix expressives » car l'expressivité est au centre du projet et conditionne très fortement l'utilisation de voix de synthèse en multimédia.

Projet RIAM Ecoute

Le projet Ecoute (2006-2008) propose une chaîne éditoriale hypermédia permettant de constituer des collections sonores musicales, de les indexer finement au niveau inter-document et aussi intra-document.

Projet ACI Masse de Données MusicDiscover « Bases de données audio pour l'apprentissage et la référence »

Ce projet (2005-2007) est mené en collaboration avec l'ENST et le LIRIS. Le projet étudie des méthodes d'analyse, d'apprentissage et de reconnaissance dans les bases de données audio.

Projet blanc ANR-CONSONNES

Ce projet blanc GIP-ANR (2006-2009) est sur le contrôle des sons et l'inversion des modèles physiques d'instruments de musique. Il est mené en collaboration avec le LMA CNRS et l'ENST- UMR5141.

Projet mécatronique

Ce projet débuté en 2006 est en partenariat avec l'Ecole des Mines de Paris. Le projet consiste à piloter les éléments asservissant les paramètres de lèvres synthétiques et du jeu d'une trompette.

Projet Européen Semantic HIFI

Ce projet terminé en 2006 concernait la navigation, le partage, l'interaction, l'écoute et l'exécution sur les futurs systèmes de haute fidélité.

Montage du projet Quaero de l'All

L'équipe Analyse/Synthèse de l'Ircam est fortement impliquée dans le montage du projet Quaero de l'All. Dans ce projet d'explorateur multimédia du Web, l'Ircam coordonne les activités d'indexation

musicale du projet.

Contrat avec la société de production REZO FILMS

Ce contrat concerne la production des transformations de voix d'homme en voix de femme pour le film "Les amours d'Astrée et Céladon" de E. Rohmer, qui doit sortir dans les salles en 2007.

Contrat avec l'association E Voce di U Cumune et l'Université de Corse

Ce projet a pour but l'analyse de chants polyphoniques corses et la synthèse de voix chantées, associées aux gestes de la mimophonie.

MakeMusic

Plusieurs contrats ont été réalisés ou sont en cours avec la société MakeMusic et concernent l'aide à l'apprentissage des instruments de musique pour l'élève avec leur logiciel SmartMusic.

Contrat « Emergence » avec Renault et Sncf

une collaboration industrielle avec Renault et Sncf, une étude a été menée sur l'émergence temporelle des bruits parasites à l'intérieur des habitacles de voitures (automobile/TGV).

Contrat « Etude des sons des bébés » avec l'association PILE

Ce projet étudie les précurseurs de la parole dans le signal sonore produit par des bébés, avec l'association PILE et l'hôpital Necker

