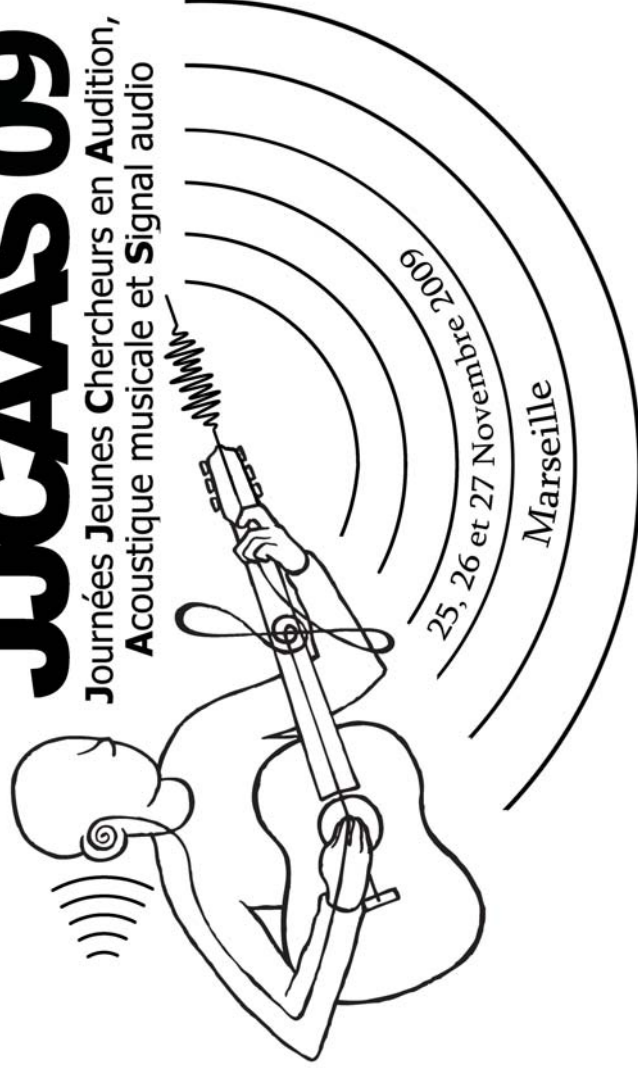


JJCAAS 09

Journées Jeunes Chercheurs en **Audition,**
Acoustique musicale et **Signal audio**



Journées des Jeunes Chercheurs en
Audition, Acoustique,
traitement du Signal audio

JJCAAS 2009

25, 26 et 27 novembre

Marseille

Laboratoire de Mécanique et d'Acoustique
UPR-7051

31 Chemin Joseph Aiguier
13402 Marseille, cedex 20
France

Partenaires des JJCAAS 2009 :



Programme des journées :

Mercredi 25 novembre	10h30	Accueil des participants
	10h45	Discours d'ouverture
	11h00	Visite des installations du LMA
	13h00	Buffet
	15h00	Discours de Jean Kergomard
	16h00	Conférence de Murray Campbell
	17h15	Pause et collation
Jeudi 26 novembre	17h30	Présentation des partenaires industriels
		Session 1
	8h45	Accueil des participants
	9h00	Session 2
	10h45	Pause et collation
	11h00	Session 3
	13h00	Buffet
Vendredi 27 novembre	15h00	Conférence de Jean-Claude Risset
	16h15	Pause et collation
	17h00	Visite du GMEM
	20h30	Repas de gala au Marseillois
	8h45	Accueil des participants
	9h00	Session 4
	10h45	Pause et collation
	11h00	Session 5
	13h00	Buffet
	14h30	Session 6
	16h30	Cocktail de clôture des JJCAAS 2009

Présentation des conférenciers invités :

Jean-Claude Risset (biographie issue de <http://brahms.ircam.fr>)

Jean-Claude Risset, né au Puy le 13 mars 1938, est à la fois musicien et chercheur en physique acoustique. Après une solide formation de pianiste auprès de Robert Trimaille (élève d'Alfred Cortot) qui lui donne l'envie d'entamer une carrière de pianiste, il découvre la composition entre 1961 et 1964 : André Jolivet l'engage à étudier l'écriture avec Suzanne Denarquez. Parallèlement, étudiant à l'École Normale Supérieure à Paris, il devient agrégé de physique en 1961 et Docteur d'État en Sciences Physiques en 1967 : il commence alors une carrière de scientifique, dans le domaine de l'électronique. Pionnier en informatique musicale, comme l'attestent ses travaux sur la synthèse sonore et en psychoacoustique, notamment lors de ses séjours aux Bell Laboratories, il acquiert rapidement une renommée internationale. Il œuvre dans la recherche scientifique au sein du CNRS, à l'Institut Electronique Fondamentale de Pierre Grivet de 1961 à 1971, aux Bell Laboratories dans le New-Jersey (USA), autour de Max Mathews et John Pierce entre 1964-1965 et 1967-1969, séjour pendant lequel il développe des travaux sur la synthèse des sons par ordinateur et leurs applications musicales (notamment la simulation des sons instrumentaux, les illusions sonores et paradoxes musicaux), à Orsay (1970-71), puis, à partir de 1972, au Centre Universitaire de Marseille-Luminy, à l'Ircam de 1975 à 1979 et enfin au LMA (Laboratoire de Mécanique et d'Acoustique) du CNRS à Marseille, institution dans laquelle il reste directeur de recherche émérite.

Invité dans de nombreux pays et institutions de recherche scientifique et musicale, comme le CCRMA de Stanford (après de son homologue chercheur-musicien John Chowning) en 1971, 1975, 1982, 1986, 1998, le studio électronique de Dartmouth College (avec Jon Appleton), et le Media Lab du MIT (USA) en 1987 et 1989 pour ses travaux autour du piano Disklavier Yamaha. Jean-Claude Risset fut maître de conférences en musique à l'Université d'Aix-Marseille entre 1971 et 1975, puis professeur entre 1979 et 1985, directeur du département « ordinateur » de l'Ircam entre 1975-1979, puis responsable entre 1993 et 1999 du DEA national « Acoustique, traitement du signal et informatique appliqués à la musique », dispensé à l'Ircam conjointement par l'Université de la Méditerranée et l'Université de Paris VI. Ses recherches scientifiques alimenteront incessamment son travail de musicien, et réciproquement.

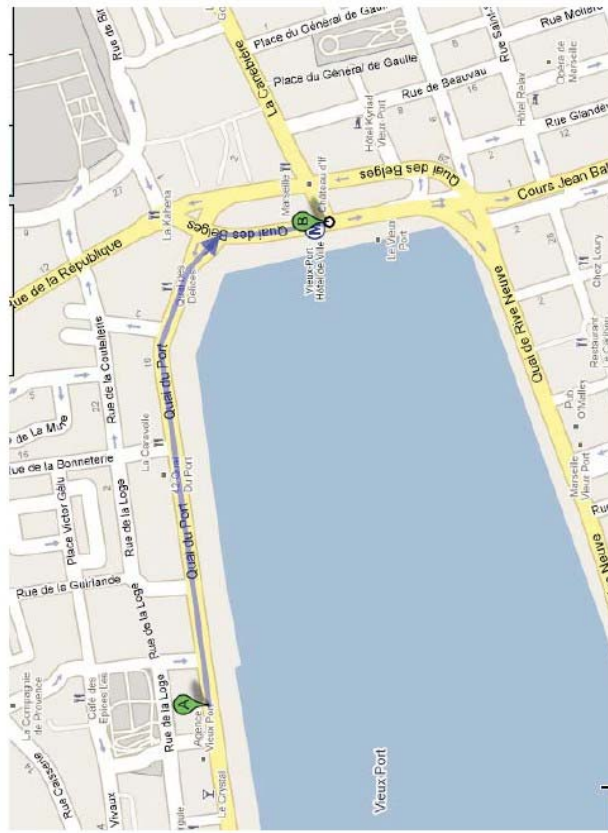
Son catalogue d'œuvres musicales, riche de 70 pièces, est composé de 15 œuvres pour « sons fixés sur support », à savoir des musiques électroniques réalisées aux Bell Laboratories, à l'Ircam, au LMA-CNRS, ou des musiques acousmatiques réalisées à l'INA-Grim, au GMEM..., de 19 œuvres instrumentales et de 35 œuvres mixtes (dont 6 avec électronique temps réel), catégorie qu'il défend tout particulièrement. Ces œuvres sont l'occasion de concrétiser l'idée de « composer le son lui-même », en plus de composer avec ces sons.

Murray Campbell (biographie issue de <http://www2.ph.ed.ac.uk/acoustics>)

Murray Campbell is Professor of Musical Acoustics in the School of Physics and Astronomy at the University of Edinburgh. His main research activity is concerned with different aspects of the acoustical and musical behaviour of wind instruments. For some years the non-linear dynamics of lip-vibrated wind instruments has been studied using artificial lip excitation and high-speed digital photography. Non-linear loss mechanisms in wind instruments have been explored using particle image velocimetry. Acoustic pulse reflectometry and related techniques have been developed as effective tools for the design and optimisation of wind instruments. Work is also in progress on acoustical studies of historic wind instruments, making use of the rich resources of the University of Edinburgh's Collection of Historic Musical Instruments.

Murray Campbell is Co-ordinator of the UK Musical Acoustics Network, Co-chair of the Technical Committee for Musical Acoustics of the European Acoustics Association, and Associate Editor for Acta Acustica united with Acustica. He is Musical Director of the Edinburgh Renaissance Band, the Scottish Gabrieli Ensemble and the Linton Singers.

Voici le plan pour aller du métro Vieux Port sur la ligne 1 jusqu'au Mar-seillois.



SESSION 1 – mercredi 25 novembre – 17h00	
CHARDON Gilles	Holographie acoustique en champ proche et compressed sensing
GOYET Louise	Le problème de la segmentation de la parole en mots en français : études électrophysiologiques
BOUCHARA Tifanie	Conception d'une lentille grossissante audiovisuelle pour l'exploration de bases de données multimédias
ESLING Philippe	Etude des mélanges sonores et orchestration
JODER Cyril	Étude des descripteurs acoustiques pour l'alignement temporel audio-sur-partition musicale
MOUSSALLAM Manuel	Traitements du son à partir de représentations parcimonieuses
LEMAN Adrien	Modèle basé sur le signal non-intrusif de l'évaluation de la qualité vocale, utilisant une méthode de classification automatique des bruits de fond
ELIE Benjamin	Caractérisation acoustique et vibratoire des instruments à cordes en vue de leur catégorisation
BROCOLINI Laurent	Discrimination d'environnements sonores urbains. Comparaison entre ACP et Carte de Kohonen
COUSINEAU Marion	Perception de séquences de hauteur et de sonie : normo-entendants, implantés cochléaires et simulations d'implant
MATHEY Marie	Dynamic of auditory processing
LIUNI Marco	Résolutions différents pour informations différents
SIRDEY Adrien	Perspectives offertes par les masques de Gabor dans la transposition des notes de musique

SESSION 2 – jeudi 26 novembre – 8h45	
FONTAN Lionel	Intelligibilité, perception et compréhension de la parole pathologique
MASSIDA Zoé	Voice perception in cochlear implanted patients
ANNABI-ELKADRI Nefissa	Sonogramme à multirésolution
DEGOTTE Gilles	Estimation du filtre du conduit-vocal adapté à un modèle d'excitation mixte pour la transformation et la synthèse de la voix
REGNIER Lise	Le vibrato, un descripteur pertinent pour la détection et l'indexation de la voix chantée ?
CHADEFAUX Delphine	Interaction doigt-corde
DESTEPHEN Clément	Synthèse sonore de véhicules en déplacement à vitesse variable
SCIABICA Jean-François	Modélisation psychoacoustique du bruit perçu dans un habitacle automobile
DELAITRE Pauline	Qu'entendez vous par « Zones Calmes » ?
LE BAGOUSSE Sarah	Méthodologie de test pour l'évaluation subjective de la qualité du son spatialisé
FUENTES Benoit	Une stratégie fondée sur l'ajout de bruit pour l'extension de bande en audio
DESSEIN Arnaud	Introduction aux outils de la géométrie de l'information pour la manipulation des flux audio
LOYAU Fanny	La gestion des communications multiples dans les systèmes aéronautiques complexes

SESSION 3 – jeudi 26 novembre – 11h	
DUBOIS Françoise	Détection de bruits émergents au sein d'habitués: mesure et modélisation
FRANCOIS Clément	Statistical learning of a sung language: behavioural and ERP studies to compare musicians and non musicians adults and normal readers and dyslexics children
BARTHET Mathieu	Reconnaissance automatique d'instruments: une application à l'indexation automatique de la musique
DEPREZ Romain	Vers l'adaptation automatique de contenus audio multicanaux
HUGUET Laurent	Caractérisation de phénomènes d'interactions source-filtre en voix parlée et chantée
SANNICANDRO Valerio	Space As Parameter – A Different Approach to "Écriture", From Acoustic to Music Notation
ABLITZER Frédéric	Statique de l'archet de violon : influence de la tension du crin sur la souplesse
EVENO Pauline	Instruments de musique à vent : recherche d'attributs acoustiques pertinents pour la facture instrumentale
CHAMBERS Claire	Biasing the perception of ambiguous pitch stimuli
DEVERGIE Aymeric	Influence de la lecture labiale sur la ségrégation de flux de paroles
FDILI ALAOUI Sarah	Reconnaissance du geste expert et contrôle de systèmes complexes : cas de l'interface entre le geste dansé et le rendu graphique par modèle physique
PARSEHIAN Gaetan	Conception d'un système de guidage auditif spatialisé pour non-voyant
SELDRAN Fabien	Simulation d'un implant EAS (Electric Acoustic Stimulation) par un vocodeur hybride

SESSION 4 – vendredi 27 novembre – 8h45	
PASCAL Frédéric & T ƆƎƆƆƆ. .i.ǻ ^	Enregistrement et transmission, vers une approche pour une meilleure compréhension des implants cochléaires
FOUCARD Rémi	Influence of musical expertise on tone and phoneme processing in Mandarin Chinese
GULLUNI Sébastien	Similarité multimodale pour la recherche de reprises d'un morceau
MAAZAOUI Mounira	Segmentation interactive de Musique Electroacoustique
OLIVERO Anaik	Méthodes de séparation aveugles de sources audio par réseau de microphones
MALLARONI Bastien	Analyse et catégorisation de sons par multiplicateurs temps-fréquence
BIANCO Tommaso	Etude du rayonnement de résonateurs de longueur finie
BOUTON Sophie	Coordination of facial and oral activity in trumpet performance
DEROCHE Mickaël	Perception catégorielle des sons de parole chez les enfants munis d'un implant cochléaire
DOUKHAN David	Données psychoacoustiques sur l'annulation d'interférence harmonique
SCHÖNSTEIN David	Modèles de l'expressivité pour la synthèse de récits courts
MACHART Pierre	Méthodes pour l'adaptation individuelle de l'HRTF pour le rendu via la synthèse binaurale
	Segmentation Morphologique

SESSION 5 – vendredi 27 novembre – 11h	
ALAZARD Charlotte	Rôle de la prosodie dans la structuration du discours
CHOBERT Julie	Multi-Feature MMN to Speech sounds in Dyslexic and Control children.
CARMIAUX Baptiste	Similarités Morphologiques entre Geste et Son et leur Détection Temporelle
FREITAS Caetano Marcelo	Applications des Algorithmes Génétiques en Synthèse et Transformation Sonore par Descripteurs
HENNEQUIN Romain	Introduction de modèles source/filtre dans la factorisation en matrices non-négatives de données audio
MANIATAKOS Fivos	Une architecture pour le contrôle temporel de réinjection stylistique dans le cadre de l'improvisation assistée par ordinateur
KARKAR Sami	Bifurcations des régimes périodiques dans les systèmes dynamiques non linéaires. Application aux instruments de musique auto-oscillants.
ROZE David	Simulation de la dynamique d'une poutre de Reissner par les séries de Volterra
MOREL Julien	Gène en situation de multi-exposition sonore à différents bruits industriels
BOUVET Lucie	Le traitement global et local dans la musique
LOYET Raphaël	Simplification perceptuelle des méthodes à base de rayons pour l'auralisation
MICHAUD Pierre-Yohan	Méthode adaptée aux tests perceptifs pour l'évaluation d'un grand nombre de stimuli audio

SESSION 6 – vendredi 27 novembre – 14h30	
ANDEOL Guillaume	Localisation sonore dans le bruit et capacité de résolution spectrale
KARYPIDIS Charalampos	Le paradigme « même-différent » et les asymétries en perception vocale: comparaison des modèles de discriminabilité et enjeux théoriques
MOFFAT Graeme	Preliminary evidence for recovery from auditory deprivation in asymmetric hearing loss after hearing aid fitting
MICHEL Christophe	Un modèle de cochlée in silico à visée thérapeutique
MERER Adrien	Caractérisation perceptive de trajectoires pour la synthèse sonore
BELLET Romain	Contrôle passif du bruit grâce à un absorbeur nonlinéaire
VERRON Charles	Synthèse de sons d'environnement spatialisés
COHEN Ivan	Simulation d'un étage d'amplificateur guitare pour différents modèles de lampes
ZANGO Yves	Conception de signaux de référence pour calibrer l'évaluation subjective des codeurs de la parole et du son
NECCIARI Thibaud	Mesure du masquage auditif temps-fréquence, application à l'analyse-synthèse des sons
FREY Aline	Les Unités Sémiotiques Temporelles : des figures sonores ?
REBILLAT Marc	Identification, modélisation et contrôle de Large Multi-Actuator Panels pour la création d'un rendu audio-visuel spatialisé

1. Ablitzer Frédéric :

Statique de l'archet de violon : influence de la tension du crin sur la souplesse.

Malgré une abondante littérature scientifique sur le violon, l'archet n'a que rarement été étudié en tant que tel. Pourtant, son rôle dans l'interaction entre le musicien et son instrument semble essentiel, comme en témoigne l'exigence particulière des violonistes professionnels quant au choix de leur archet. Le travail présenté vise à établir des corrélations entre les qualités de jeu, évaluées en termes de jouabilité et de sonorité, et les propriétés physiques de l'archet. Dans un premier temps, on s'intéresse au comportement statique de l'archet en tant que structure mécanique, tandis que les aspects dynamiques seront abordés ultérieurement. L'une des perspectives est d'apporter une aide à la facture instrumentale en mettant à disposition des archetiers des outils de caractérisation et de simulation dans le but de mieux maîtriser le réglage des archets.

D'un point de vue statique, la fonction de l'archet est de transmettre les efforts exercés par le bras du violoniste jusqu'à la corde, par l'intermédiaire d'une mèche tendue. Ainsi, le rôle premier de la bague est de maintenir la mèche sous tension, ce qui impose certaines contraintes de conception. En effet, l'allongement de l'archet au cours de son évolution implique une tension du crin de plus en plus élevée afin de conserver une raideur verticale suffisante ; la diminution progressive puis l'inversion de la courbure (cambre), initialement convexe, a permis à la bague de supporter cet accroissement de tension. Depuis l'apparition de l'archet moderne à la fin du XVIIIe siècle, le cambre de la bague constitue un paramètre de réglage supplémentaire. Un cambre très prononcé favorise une tension de mèche élevée en situation de jeu. Cependant, une augmentation de tension du crin entraîne également une augmentation de souplesse latérale pouvant s'avérer critique dans certains modes de jeu. Ainsi, le réglage d'un archet semble résulter en partie d'un équilibre entre raideur verticale et souplesse latérale, auquel l'archetier parvient en ajustant le cambre et la mysection le long de la bague en fonction de l'élasticité du bois.

2. Alazard Charlotte :

Rôle de la prosodie dans la structuration du discours - Proposition d'une méthodologie de l'enseignement de l'oral vers l'écrit en français langue étrangère.

Cette recherche a pour objectif de tester l'influence de la pratique linguistique à l'oral sur les *compétences en lecture* d'apprenants de français langue étrangère (Transfert de compétences de l'oral vers l'écrit). Nous avons formulé l'hypothèse selon laquelle un travail poussé de correction phonétique - mettant l'accent sur la prosodie - permettrait d'améliorer les performances en lecture des apprenants (*Décodage de l'écrit facilité*).

Pour tester cette hypothèse, nous avons mené une étude longitudinale sur huit semaines avec quatre apprenants anglophones (deux apprenants de niveau élémentaire et deux de niveau indépendant) répartis en deux groupes : une groupe de *contrôle* qui a suivi des cours de compréhension écrite "classique" dans une approche communicative, et un groupe *test* qui a suivi des séances de correction phonétique centrées sur la sensibilisation à la structure prosodique du français. Deux tests identiques ont été proposés aux apprenants, en début et en fin de formation, afin d'évaluer leur fluidité orale. Les corpus recueillis ont tout d'abord été écoutés par deux experts qui devaient repérer la distribution et les types de pauses (silencieuses ou hésitations vocales), les frontières de groupe rythmiques et les accents réalisés. Nous avons ensuite mesuré la durée des pauses silencieuses et remplies, la taille des GR, ainsi que la durée sur les accents perçus (analyse acoustique). Enfin, les corpus ont été écoutés par six francophones non linguistes qui devaient évaluer sur une échelle de 0 à 8 la fluidité verbale (test perceptif).

Nos premiers résultats montrent que les caractéristiques prosodiques du français sont mieux maîtrisées pour les sujets du groupe *test*, en fin de formation : la distribution des pauses est plus cohérente par rapport à la structure syntaxique et se trouve globalement à des jonctures prosodiques naturelles (moins d'hésitations vocales à l'intérieur des syntagmes prosodiques), ce qui améliore la fluidité. Cette progression a particulièrement été perçue chez l'apprenant débutant du groupe *test*.

Un travail centré sur la prosodie à l'oral permettrait à l'apprenant de se débarrasser de stratégies erronées d'accès au texte, et donc d'acquérir certains automatismes en langue-cible qui facilitent l'appréhension (et la compréhension) globale du texte.

3. Andéol Guillaume :

Localisation sonore dans le bruit et capacité de résolution spectrale.

La capacité à localiser une source sonore dans l'espace varie grandement d'un auditeur à l'autre. Cette variabilité inter individuelle pose question vis-à-vis de l'usage de la technologie son 3D. Le son 3D est une nouvelle interface hommes-machine spatialisant virtuellement les alarmes ou les communications radio pour en faciliter leur traitement perceptif. Elle est aujourd'hui proposée aux opérateurs des systèmes complexes, en particulier dans le milieu aéronautique (équipement des pilotes de chasse). Or ces environnements sont caractérisés par leur ambiance bruyante. Cela pousse à s'intéresser à la variabilité interindividuelle en localisation sonore dans le bruit et plus particulièrement, aux facteurs explicatifs d'une telle variabilité. Une fois ces facteurs identifiés l'objectif est de pouvoir proposer des solutions en terme d'entraînement pour amener le maximum d'opérateurs à une utilisation optimale du son 3D.

Après une précédente étude sur l'impact de l'innervation cochléaire (système olivo-cochléaire éfférent médian), nous nous sommes intéressés au rôle des processus d'analyse spectrale dans l'explication de la variabilité inter individuelle en localisation sonore dans le bruit. Nos premiers résultats mettent en évidence une corrélation entre capacité d'analyse spectrale et capacité de localisation sonore dans le bruit. Ces résultats complètent ceux de l'étude portant sur l'innervation cochléaire : en effet la relation entre capacité d'analyse spectrale et localisation sonore s'observe pour des niveaux de bruits légèrement inférieurs à ceux où la capacité de localisation est fortement liée à la « qualité » de l'innervation cochléaire. Cette observation pourrait s'expliquer à la lumière de la Reverse Hierarchy Theory (théorie de la hiérarchie inversée) qui souligne l'importance des processus perceptifs de bas niveau dans la perception en ambiance bruitée.

4. Annabi-Elkadri Nefissa :

Sonagramme multirésolution.

Les sonagrammes classiques montrent généralement l'intensité (dB) en fonction du temps et de la fréquence. Le niveau de pression acoustique en dB est approximativement proportionnelle au volume perçu par l'oreille (à des fréquences élevées). Un sonagramme classique fournit un bon affichage du point de vue psycho-acoustique pour la parole, à condition que la longueur de la fenêtre soit comparable au temps d'intégration de l'oreille. Cette longueur est généralement choisie entre 10ms et 20ms. Le sonagramme classique n'offre qu'un seul temps d'intégration qui est la longueur de la fenêtre. Il met en œuvre un filtre passe-bande uniforme, autrement dit, les échantillons spectraux sont régulièrement espacés et correspondent à des bandes passantes égales. Le choix de la longueur de la fenêtre détermine la résolution comme en temps-fréquence pour toutes les fréquences du sonagramme. Plus la fenêtre est étroite, plus la résolution en temps est bonne et plus la résolution en fréquence est mauvaise. Ce qui implique que la résolution d'affichage des formants, des voissements et des bruits de frictions, dans les basses fréquences est moins bonne que la résolution des explosions dans les hautes fréquences et inversement... Il faut donc faire le bon choix de fenêtrage par rapport à l'allure du signal. Nous proposons une méthode pour améliorer le sonagramme classique. Il s'agit de calculer un sonagramme à multi-résolution en combinant plusieurs FFT de différentes longueurs de telle manière que la précision temporelle soit plus élevée dans les hautes fréquences, et la résolution de fréquence plus élevée dans les basses fréquences, comme dans l'oreille. Dans chaque FFT, la résolution de fréquence doit être supérieure ou égale à celle de l'oreille dans la bande de fréquence qu'il couvre.

5. Barthet Mathieu :

Reconnaissance automatique d'instruments : une application à l'indexation automatique de la musique.

La diffusion croissante de musique sous format numérique suscite de nouveaux défis concernant la recherche d'informations musicales et la sélection de morceaux à partir du contenu même des signaux audiofréquences.

Le projet OMRASII (Online Music Recognition And Searching), mené au laboratoire Centre for Digital Music (Queen Mary University London), vise à développer de nouvelles méthodes d'échanges d'informations et d'indexation sur des grandes bases de données musicales (par exemple, sa propre collection de musique sur un Apple Ipod). L'une des perspectives de ce projet est de fournir de nouveaux outils de navigation dans les collections musicales, permettant aux utilisateurs de réaliser des requêtes de plus en plus précises et avancées (par exemple, trouver tous les morceaux contenant du chant et du saxophone).

L'un des aspects fondamentaux du son musical est le timbre, cet attribut de la perception qui nous permet de distinguer deux sons de qualités sonores différentes (ceux de deux instruments de musique différents, par exemple), ayant des hauteurs, sonies (intensité perçue), et durées identiques. Comment reproduire de manière artificielle ce que l'oreille d'un mélomane est capable d'effectuer à tout moment : reconnaître deux instruments (timbres) différents?

L'approche choisie pour aborder ce problème est pluridisciplinaire et mêle la psychoacoustique, le traitement du signal audio et l'informatique. Nous décrivons un système de reconnaissance automatique d'instruments de musique, élaboré pour des séquences mono-timbrales (pas de mélanges d'instruments). Ce dernier se compose de trois parties : la modélisation paramétrique des signaux (Linear Predictive Coding), leur caractérisation à l'aide de descripteurs de timbre représentant l'enveloppe spectrale (Linear Spectrum Frequencies), et l'utilisation d'une méthode d'apprentissage automatique (k-means).

Nous présenterons l'intégration de ce système sous la forme d'un module VAMP (<http://www.vamp-plugins.org/>), fonctionnant de manière interactive avec l'application Sonic Visualiser, développée à l'université Queen Mary (<http://www.sonicvisualiser.org/>). Cette application permet d'analyser et de visualiser le contenu des signaux sonores et musicaux.

6. Bellet Romain :

Contrôle passif du bruit grâce à un absorbeur non linéaire.

Ce résumé décrit un dispositif expérimental qui a été conçu pour reproduire le phénomène de « pompage énergétique » entre un milieu acoustique et un absorbeur non linéaire. Les applications potentielles visées concernent la réduction des bruits dans le domaine des basses fréquences, là où les matériaux poreux sont peu efficaces.

Le concept de pompage énergétique a été introduit par Vakakis et Gendelman en 2001, sur un système mécanique à deux degrés de liberté. L'idée est de calmer les vibrations d'une structure linéaire en lui adjoignant un absorbeur dynamique constitué d'une masse, d'un amortisseur et d'un ressort à raideur purement non-linéaire. En choisissant habilement le couplage et la non linéarité, il est possible de créer un « mode de vibration non linéaire » tel que l'énergie vibratoire de la structure à isoler est transférée vers l'absorbeur de façon irréversible, sans retour possible, c'est le phénomène de « pompage énergétique ». Sous certaines conditions, on peut obtenir une disparition rapide des oscillations dans la structure à isoler.

L'idée est ici d'appliquer ces concepts en acoustique. Nous avons mis au point un premier montage expérimental pour montrer la faisabilité du transfert irréversible d'énergie entre un milieu acoustique (linéaire) et une membrane visco-élastique (non linéaire). Le montage est constitué d'un tube, d'une boîte de couplage et d'une membrane à peine tendue. L'air contenue dans le tube joue le rôle du système linéaire à calmer, et la membrane est dimensionnée pour effectuer des oscillations de grandes amplitudes et se comporter ainsi comme un système presque purement non linéaire.

Nous présenterons donc le montage expérimental et les résultats des campagnes d'essais montrant les transferts d'énergies, les réponses fréquentielles du système ainsi que différentes analyses et illustrations de ces phénomènes.

7. Bianco Tommaso :

Coordination of facial and oral activity in trumpet performance.

Sound production in trumpet performance builds on the concurrency and synergy of various factors : the closing of the lips, the mouthpiece force (force between lips and mouthpiece), the lips' compression and tension, the tongue arch, the intraoral pressure and the air speed and support. The correct interplay of these factors results from the expert control of two coupled systems : the respiratory apparatus and the oral and facial neuromuscular control. Literature on trumpet performance, however, focussed separately on these two constituents. Pioneer work of Bouhuys [1], who related pitch and loudness to breathing patterns in wind instrument players, promoted further studies on respiratory and oral mechanics in lip reed instruments. Bouhuys himself [2] re-covered the subject with stress on mouth pressure and flow rate. Champigny [3] focussed on tonguing movement and mouthpiece force, whereas Fletcher and Tarnopolsky [4] identified features between blowing pressure values and sound output properties. Experimental values measured on real performers gave the basis for qualitative relations between variables [5], computational lips models [6] and physical artificial reproduction [7]. In the neuromuscular aspect, Isley [8], White and Basmajian [9] first used electromyography technology for the identification of the facial muscles that are more involved in the embouchure setting. Other works related muscular activity and tone production [10], or used the results for the identification of playing [11] and medical [12] disfunctions. However, to the author's knowledge, no work has been conducted on the interplay of both the respiratory and oral systems. In this direction we carried out measurements in collaboration with the Respiratory mechanics laboratory at CHRM/MT, McGill University. In the present account we will present the preliminary results of such measurements, issued from the combined observation of the radiated sound envelope, the intraoral pressure, the mouthpiece force, the envelope of EMG activity for Depressor Anguli Oris and Zygomaticus Major muscles.

- [1] Bouhuys, A. (1964). Lung volumes and breathing patterns in windinstrument players. *Journal of Applied Physiology Journal of Applied Physiology*, 967-975.
- [2] Bouhuys, A. (1965). Sound-Power Production in Wind Instruments. *Acoustical Society of America Journal* 37, 453-+.
- [3] Champigny, Bernard (1990). La Denture des cuivres. *Marsyas* 14, 55-63.
- [4] Fletcher, N. H. and Tarnopolsky, A. (1999). Blowing pressure, power, and spectrum in trumpet playing. *Journal of Acoustical Society of America* 105,

874-881.

- [5] Elliott, S. and Bowsher, J. (1982). Regeneration in brass wind instruments. *Journal of Sound and Vibration* 83, 181-217.
- [6] X. Rodet, One and two mass model oscillations for voice and instruments, in *Proceedings of the 995 International Music Conference, Banff*, pp. 207-214, Computer Music Association, 1995.
- [7] C. Vergez and X. Rodet, Model of the trumpet functioning : Real time simulation and experiments with an artificial mouth, in *Proceedings of the Int. Symposium Musical Acoustics (ISMA-97)*, Edinburgh, Scotland, pp. 425-432, Aug. 1997.
- [8] Isley, Charles L. (1972). A theory of brasswind embouchure based upon facial anatomy, electromyographic kinesiology, and brasswind embouchure pedagogy.
- [9] White, E.R. and Basmajian, J.V. (1973). Electromyography of Lip Muscles and Their Role in Trumpet Playing. *Journal of Applied Physiology* 35, 892-7.
- [10] Heuser, Frank and McNitt-Gray, Jill L. (1991). EMG potentials prior to tone commencement in trumpet players. *Medical Problems of Performing Artists* 6, 51.
- [11] Heuser, Frank and McNitt-Gray, Jill L. (1998). Enhancing and Validating Pedagogical Practice : The Use of Electromyography during Trumpet Instruction. *Science & Medicine* 13, 155.
- [12] Iltis, P.W. and Givens, M.W. (2005). EMG characterization of embouchure muscle activity : reliability and application to embouchure dystonia. *Medical Problems of Performing Artists* 20, 25-34.

8. Bouchara Tifanie :

Conception d'une lentille grossissante audiovisuelle pour l'exploration de bases de données multimédias.

Afin d'améliorer les interfaces audio et audiovisuelles, nous nous proposons de reprendre les méthodes graphiques utilisées dans le domaine de la visualisation et de les étendre au domaine sonore. En particulier, nous avons développé une lentille grossissante qui agit sur les deux modalités auditive et visuelle de façon simultanée et cohérente, et qui, pour la modalité visuelle, s'apparente aux distorsions de type vues en oeil-de-poisson. Pour l'aspect audio, il s'agit de spatialiser les sons puis de déformer ce rendu sonore, 2D ou 3D, en fonction du rendu graphique. Une seconde méthode, dite Pan & Zoom, sans distorsion graphique, a également été implémentée. Ces méthodes ont été testées dans une application de recherche de documents vidéos en contexte réaliste. La tâche proposée était d'explorer une collection non structurée, constituée de 100 extraits vidéos musicaux de 10s lus en boucle simultanément, et de retrouver une vidéo particulière parmi toutes. Afin d'évaluer l'apport de la dimension sonore dans ce type tâche, nous avons comparé deux conditions de modalités : une avec seulement le rendu graphique et une avec un rendu audiovisuel.

L'évaluation a montré l'importance de la modalité visuelle par rapport à la modalité sonore, même dans un contexte musical, et a mis en avant l'intérêt d'utiliser la modalité auditive quand le rendu visuel est dégradé. Par la suite, nous chercherons à étendre vers l'audio d'autres méthodes graphiques, basées non plus sur la distorsion de l'espace mais sur les transformations du signal lui-même. Notamment, nous aimerions mettre en avant les correspondances et les complémentarités qui peuvent se faire, et d'un point de vue algorithmique et d'un point de vue cognitif, entre les traitements de l'image et ceux du son.

9. Bouton Sophie :

Perception catégorielle des sons de parole chez les enfants munis d'un implant cochléaire.

Bien qu'il existe de rares études sur la perception des sons de parole chez les enfants implantés, celles-ci rapportent des données contradictoires (Pisoni, 1999 ; Chin et al., 2001 ; Geers, Nicholas and Davidson, 2003 ; Medina et Serniclaes, 2005). L'objectif de l'étude présentée est double. D'une part, il s'agit de préciser les compétences de perception des sons de parole d'enfants implantés francophones et plus particulièrement leurs compétences de catégorisation de paires minimales. D'autre part, il s'agit de déterminer si les particularités fonctionnelles de l'implant peuvent influencer la perception des sons de parole.

Les performances de 18 enfants sourds implantés avant l'âge de 3 ans (moyenne d'âge chronologique : 9,1) et bénéficiant de leur implant depuis 5 à 8,7 ans ont été comparées à celles des enfants normo-entendants de même âge audio-perceptif (calculé à partir de la date d'implantation de l'enfant implanté et correspondant à la durée d'utilisation de l'implant). Les enfants normo-entendants ont donc un âge chronologique correspondant à l'âge audio-perceptif des enfants implantés (moyenne : 6 ans). La perception catégorielle de traits articulatoires caractéristiques des consonnes (lieu, nasalité, mode et voisement) a été évaluée avec une tâche de discrimination de paires minimales utilisant des mots (exemple : est-ce que ces deux mots sont pareils ou pas pareils ? bouton / mouton) et une tâche d'identification de mots (exemple : à quelle image correspond le mot « bouche », suivi des images « bouche » et « mouche »).

On observe que les enfants implantés et normo-entendants de même âge audio-perceptif ont des performances de perception catégorielle équivalentes. Néanmoins, des scores de bonnes réponses significativement plus faibles chez les enfants implantés par rapport aux enfants normo-entendants montrent que les enfants implantés présentent un niveau de précision catégorielle plus faible que celui des enfants normo-entendants.

On observe également que les performances diffèrent selon le trait articulatoire considéré puisque les performances s'ordonnent de la façon suivante : lieu = nasalité > mode = voisement. Une interprétation possible des difficultés des enfants implantés à percevoir les phonèmes est que l'implant cochléaire présente certaines particularités de transmission des signaux acoustiques. Cette interprétation est envisagée en comparant les indices acoustiques majeurs caractérisant les traits articulatoires (enveloppe temporelle vs structure temporelle fine) et les indices acoustiques transmis par l'implant. Le pattern de résultat est consistant avec le fait que les signaux d'enveloppe temporelle sont mieux transmis par l'implant cochléaire que les signaux de structure temporelle fine.

10. Bouvet Lucie :

Le traitement global et local dans la musique.

L'observation de patients cérébro-lésés a mis en évidence un traitement spécifique pour différentes composantes musicales. En effet les intervalles (la distance fréquentielle entre deux notes) seraient traités par l'hémisphère gauche et ainsi leurs détections impliqueraient le traitement local. Le contour d'une mélodie (la direction ascendante et descendante donnée par les notes) serait quand à lui traité par l'hémisphère droit et impliquerait le traitement global (Peretz, 1990). Ces résultats sont corroborés par diverses études utilisant les techniques d'imagerie (e.g., Schiavetto et al., 1999 ; Platel et al., 1997). Cependant, certains auteurs remettent en cause cette vision du traitement musical en effectuant une analogie avec le traitement visuel (List et al., 2007, Sanders & Poeppel, 2007). De plus, la question du traitement de la dimension temporelle de la musique en suivant la même dichotomie que la dimension mélodique est aussi largement discutée (Liégeois-Chauvel et al., 1998 ; Di Pietro et al., 2004). Dans cette présentation, j'exposerai les différentes visions du traitement musical et que l'implication de ces différentes conceptions dans l'étude de population pathologique comme l'autisme.

11. Brocolini Laurent :

Discrimination d'environnements sonores urbains. Comparaison entre ACP et Carte de Kohonen.

Relier des caractéristiques perceptives aux paramètres acoustiques nécessite de réaliser des mesures sur le terrain dans différents environnements sonores. Ces mesures peuvent tenir compte d'indicateurs globaux et d'indices spécifiques aux différentes sources sonores, et généralement, les lieux d'études tels que les marchés, les parcs ou les rues sont choisis selon l'utilisation faite par des citadins. Mais qu'en est-il des différences le long d'une rue ? L'entrée d'un parc est elle similaire à la rue juxtaposée ? A quelle distance d'une rue l'environnement sonore d'un parc peut-il être considéré comme typique ? S'il est facile visuellement de faire la différence entre deux situations proches physiquement (un boulevard le long d'un parc par exemple) qu'en est-il d'un point de vue acoustique ? Les zones urbaines typiques (parc, rues, boulevards, ...) correspondent-elles à des environnements sonores homogènes ? Afin de caractériser différents espaces urbains et déterminer la distance optimale qui les sépare, une série d'enregistrements a été effectuée. Sur une distance globale d'environ 600 mètres, 70 enregistrements sonores de 15 minutes ont été réalisés tous les 7 ou 8 mètres. Les sources sonores présentes sur les enregistrements ont été identifiées par un expert. Un ensemble de 18 indicateurs a été extrait des enregistrements.

Afin de réaliser une classification des différents emplacements, basée sur ces indicateurs, et définir un maillage spatial pertinent, deux approches ont été utilisées. La première consistait en une analyse en composante principale alors que la seconde faisait appel à des modèles de réseaux de neurones, et plus précisément des cartes de Kohonen.

12. Caramiaux Baptiste :

Similarités Morphologiques entre Geste et Son et leur Détection Temporelle.

Dans le contrôle gestuel de la synthèse sonore, le paradigme dominant est la construction d'une fonction (ou mapping) qui, à des données gestuelles provenant de capteurs, associe des paramètres d'un moteur de synthèse sonore. C'est le processus de sonification. Au delà du mapping naïf associant des paramètres bruts (e.g accélération, fréquence ...), certains auteurs introduisent des paramètres de plus haut niveau afin de décrire le geste et le son permettant à l'interprète de transmettre une plus grande expressivité musicale (cf. [3]). De même, d'autres auteurs proposent de décrire le son et le geste en prenant en compte l'expression de l'émotion (cf. [1]).

Notre projet de recherche propose une approche différente pour le design d'un mapping adaptatif entre geste et son. Nous proposons une méthodologie en deux étapes. Tout d'abord nous analysons la relation geste - son lorsque le geste est effectué à l'écoute d'un son. Ensuite, nous inversons le problème afin de définir un mapping consistant et pertinent. Dans le travail présenté nous nous focaliserons sur la première étape.

Il s'agit d'étudier les similarités morphologiques entre les données gestuelles et sonores lorsque le geste est effectué à l'écoute d'un son. En premier lieu nous étudions la similarité entre une unité temporelle du geste et une unité temporelle du son. Les signaux sont fenêtrés puis représentés par une distribution de probabilités. Dans un espace adapté, la notion de similarité peut-être définie comme une métrique. Deuxièmement nous présentons une méthode de détection temporelle des structures présentes dans le signal gestuel et sonore. Une structure est un segment du signal dont le contenu d'information ne varie pas significativement. Ici nous nous plaçons dans le cadre de la géométrie d'information [2] dans lequel nous pouvons construire des algorithmes incrémentaux permettant une détection temps-réelle des structures musicales et gestuelles.

- [1] Antonio Camuri, Gualtiero Volpe, Giovanni De Poli, and Marc Leman. Communicating expressiveness and affect in multimodal interactive systems. *IEEE MultiMedia*, 12(1) :43-53, 2005.
- [2] Arshia Cont. Modeling Musical Anticipation : From the time of music to the music of time. PhD thesis, University of Paris 6 and University of California in San Diego, October 2008.
- [3] Marcelo Wanderley, guest editor. Mapping strategies in real-time computer music. *Organised Sound*, 7(02), 2002.

13. Chadefaux Delphine :

Interaction doigt-corde.

Jusqu'à présent, en acoustique musicale, les recherches se sont essentiellement concentrées sur la compréhension du fonctionnement des instruments et relativement peu sur la manière dont ils sont mis en vibrations. Pourtant, dans le cas des instruments à cordes pincées, il est connu que chaque instrumentiste produit un son qui lui est propre. Les conditions initiales données par le musicien se révèlent donc être déterminantes dans la signature acoustique de l'instrument.

Afin de comprendre ces paramètres propres aux musiciens ou à leur technique de jeu, un axe naturel d'étude est celui du mécanisme d'excitation d'une corde de harpe par l'instrumentiste. Une modélisation bidimensionnelle préliminaire de cette interaction a d'ores et déjà été mise en place [1]. Elle considère les paramètres de raideur, d'amortissement et de masse de la pulpe du doigt ainsi que la caractéristique de son frottement avec la corde. Ces éléments sont déterminés grâce aux mesures des déplacements du doigt et de la corde. Celles-ci sont effectuées avec une caméra rapide.

Ces travaux préparatoires ont révélé que certains paramètres mécaniques sont effectivement propres aux musiciens alors que d'autres s'avèrent caractéristiques de la technique de pincement employée. Les limites de ce modèle simple ont également été mises en évidence.

Un second axe qui sera développé au cours de la thèse est l'étude du signal vibratoire sur la table d'harmonie. En effet, l'analyse du transitoire d'attaque pourra être utile à l'estimation de nouveaux paramètres.

Une telle étude permettra de proposer une initialisation physiquement réaliste des modèles de synthèse sonore. Un autre intérêt est la conception d'un doigt « artificiel » qui permettra d'exciter de manière réaliste et répétitive les instruments.

- [1] K. Buys, Etude bidimensionnelle de l'interaction entre un doigt et une corde de harpe, Rapport de stage de master sous la direction de JL Le Carrou, Université Pierre et Marie Curie (Paris 6), (2009)

14. Chambers Claire :

Biasing the perception of ambiguous pitch stimuli.

We investigated hysteresis and context effects in Shepard tones perception (Shepard, 1964). Hysteresis is a memory-like phenomenon which occurs in systems whose present state depends on recent history. It has been previously associated with the perception of ambiguous stimuli. Shepard tones are complexes of sinusoidal components with an octave relationship, filtered by a fixed Gaussian envelope. When two such tones with different fundamental frequencies (F0) are successively presented, the dominant cue for judging pitch direction is the log-frequency proximity between components. At an F0-interval of a half-octave (6 semitones, or a tritone), the pitch direction heard is ambiguous. In the first condition, Shepard tone pairs were presented with a fixed F0-interval of 6 semitones. Subjects had to indicate whether they heard an upward or a downward pitch change. Perception was ambiguous across subjects, but remarkably stable for a single subject. Then, F0-intervals (from 1 to 11 semitones) were presented. Hysteresis was observed : the percept in a given trial was biased by the percept in the previous trial. Hysteresis was confirmed in additional conditions where F0-intervals were presented in increasing or decreasing series. Contrary to previous investigations, hysteresis is observed with an equivalent number of upward or downward responses. Thus, the biases observed cannot be attributed to response persistence and are likely to be perceptual in origin. We conclude that it is possible to produce strong hysteresis in pitch perception by using ambiguous stimuli. Our method creates possibilities for future investigation of the neural bases of pitch as activity may be recorded for different subjective percepts evoked by the same stimulus.

15. Chardon Gilles :

Holographie acoustique en champ proche et compressed sensing.

Une méthode classique pour étudier les propriétés vibratoires de structures (en particulier leurs déformées modales), est l'holographie acoustique en champ proche, qui consiste à analyser le champ acoustique rayonné par la structure, le rétropropager, puis en déduire la dynamique de la structure. Cette technique est rapide, peu chère, mais présente cependant quelques difficultés : la présence de bruit et le mauvais conditionnement de l'inversion des ondes oblige à mettre en oeuvre des techniques de régularisation, et les données à acquérir, stocker et traiter peuvent être de taille importante. Elle est toutefois très répandue, notamment utilisée pour le suivi de la restauration de clavecins, où elle sert à l'étude de la table d'harmonie.

Le *compressed sensing*, nouvelle méthode d'échantillonnage, tente justement de donner des solutions aux problèmes soulevés, en tirant parti de certaines connaissances a priori sur le signal recherché. Ces connaissances sont de type parcimonie, c'est-à-dire qu'on connaît une base où les quantités recherchées ont peu de coefficient non nuls. Ceci permet, même sans savoir lesquels de ces coefficients sont non nuls, d'obtenir des résultats corrects avec un nombre très faible de mesures. Ces méthodes ont de nombreuses applications (imagerie médicale, astronomie, photographie, etc.), mais encore peu dans le domaine de l'acoustique.

Cette étude montrera la pertinence d'une telle application, par la présentation du formalisme du *compressed sensing* associé à l'holographie, des simulations numériques, des indications sur les antennes particulières à construire pour appliquer ces techniques, et des résultats sur des données réelles. Des extensions seront également évoquées, comme l'application à des surfaces non planes, telles que des tables d'harmonie de violons.

16. Chobert Julie :

Multi-feature mmm to speech sounds in dyslexicans control children.

Developmental dyslexia is defined as an unexpected, specific, and persistent failure to acquire efficient reading skills despite conventional instruction, adequate intelligence, and socio-cultural opportunity. Importantly, the majority of dyslexic children also have other deficits such as attention disorders or arithmetic difficulties. In order to investigate how important dimension of speech such as phonology, frequency and duration are processed by children with limited attention abilities, we recorded their brain electrical activity. More specifically, we examined the Mismatch Negativity (MMN), a specific component of the auditory Event-Related Potentials (ERPs) elicited by any discriminable changes in auditory stimulation irrespective of the subject's attention. We compared MMNs elicited by deviants with two level of difficulty in two groups : Dyslexics and Reading-age matched controls. Results indicated differences between dyslexics and controls at the pre-attentive level. Dyslexic children showed phonological deficit but no frequency and duration deficits. This study is part of a project aimed at determining the influence of musical training (6 months) on sounds discrimination of and possibility on the remediation of dyslexia.

J. Chobert, C. François, C. Marie, D. Schön, J.L. Velay, M. Habib et M. Besson

17. Cohen Ivan :

Simulation d'un étage d'amplificateur guitare pour différents modèles de lampes.

Un certain nombre de simulateurs de circuits audio (amplificateurs guitare, synthétiseurs, matériel de studio...) ont été mis sur le marché, pour les musiciens qui veulent remplacer leur équipement par des équivalents numériques (hardware ou software), moins chers et plus pratiques à utiliser en studio ou sur scène. Néanmoins, le réalisme de ces simulations peut encore être amélioré. La modélisation des non linéarités responsables de la « signature sonore des circuits audio analogiques » est complexe et continue aujourd'hui à être un sujet d'étude.

La présentation portera sur la simulation des amplificateurs guitare à tubes pour des applications temps réel. Le comportement d'un étage à triode 12AX7 typique des préamplificateurs guitare est étudié, puis différents modèles de triodes qui prennent en compte un certain nombre de phénomènes considérés comme importants pour le réalisme sonore. Plus précisément, des comparaisons sont faites entre des modèles statiques et dynamiques de triodes. Dans chaque cas, l'étage électrique est modélisé par un système d'équations différentielles algébriques non linéaires, avec des schémas numériques standards et explicites pour les modèles statiques. Le modèle dynamique introduit des capacités parasites dont l'impact est audible à cause de l'effet Miller, et qui introduisent un problème raide dans les équations, ce qui nécessite l'emploi de schémas implicites. Des résultats sont obtenus et comparés à l'aide des plug-ins VST développés pour chaque modèle. Des exemples sonores sont disponibles sur <http://www.orosys.fr/cohen/aes127.htm>

18. Cousineau Marion :

Perception de séquences de hauteur et de sonie : normo-entendants, implantés cochléaires et simulations d'implant.

De nombreuses études ont montré que les capacités de discrimination d'attributs simples, tels que la hauteur et la sonie, sont dégradées chez les porteurs d'implants cochléaires, et qu'il existe une grande variabilité dans la population. La tâche de discrimination de séquence utilisée dans cette étude compense pour cette différence de capacité de discrimination, et permet donc de comparer l'efficacité du traitement de séquences pour différents individus, différents attributs et différentes populations.

La méthode utilisée ici a déjà permis de mettre en évidence un avantage pour le traitement de séquences de hauteur par rapport aux séquences de sonie chez le normo-entendant (Cousineau et al. 2009, JASA). Cette étude antérieure a aussi montré l'influence du type d'indices acoustiques de hauteur sur la performance : seuls les complexes harmoniques contenant des partiels qui sont résolus par le système auditif périphérique donnent lieu à une perception de séquence efficace. Ces résultats ont des implications sur la perception de séquences de hauteur chez les porteurs d'implant cochléaire dans la mesure où les implantés ne peuvent généralement pas résoudre les partiels des sons complexes.

Dans la présente étude, les données sur les normo-entendants répliquent, avec des sujets et des stimuli différents, l'avantage pour la perception de séquences de hauteur par rapport à la perception de séquences de sonie trouvée précédemment. Les implantés cochléaires, conformément aux prédictions, ne montrent pas d'avantage pour la perception de séquences de hauteur. Enfin, les simulations d'implants réalisées à l'aide de vocodeurs chez les normo-entendants montrent un pattern de résultat similaire à celui des implantés cochléaires. Ces derniers résultats confirment que le déficit observé chez les implantés cochléaires est bien dû au type d'indices de hauteur transmis et non à des facteurs spécifique à la population implantée (age, stimulation électrique, etc...). Cette étude montre que les seuils élevés ne sont pas seuls responsables de la mauvaise perception de mélodie observée chez les implantés cochléaires, et suggère qu'une amélioration de la transmission d'indices de hauteur de type « résolu » est un pré-requis à une perception de séquences de hauteur efficace pour les implantés cochléaires.

19. Degottex Gilles :

Estimation du filtre du conduit-vocal adapté à un modèle d'excitation mixte pour la transformation et la synthèse de la voix.

Actuellement, en transformation et synthèse de la voix, le filtre du conduit-vocal est habituellement supposé être excité par un spectre d'amplitude plat. Nous proposons d'utiliser un modèle de source mixte : un modèle de Liljencrants-Fant et un bruit Gaussien. L'estimation du conduit-vocal est donc adaptée à cette source en prenant en compte les amplitudes du modèle LF dans les basses fréquences et le bruit dans les hautes fréquences. Le modèle de production vocal résultant peut ensuite être utilisé pour contrôler indépendamment le conduit-vocal et la source dans le cadre de la transformation de la voix et pour l'apprentissage de leurs paramètres dans la synthèse vocale HMM.

20. Delaire Pauline :

Qu'entendez vous par « Zones Calmes » ?.

Présentation du projet de thèse sur la caractérisation des zones calmes définies par la directive européenne 2002/49/CE relative à l'évaluation et à la gestion du bruit dans l'environnement.

L'objectif de ce projet de recherche est de comprendre quelles sont les relations entre les usagers de la ville, les lieux qu'ils définissent comme « calme » et les politiques urbaines et environnementales. Il s'articule autour de trois axes de recherche. La première approche à la fois lexicographique et historique s'appuie sur les définitions du mot « calme » et des autres mots associés. L'objectif de cette approche est de comprendre comment les français se sont appropriés ce mot en lien avec l'évolution du développement urbain. Cette compréhension permettra dans un second temps de réaliser des enquêtes et des ateliers de concertation, afin de constituer des groupes de populations qui ont les mêmes représentations vis-à-vis des zones calmes. De plus, ce premier axe va permettre d'orienter des mesures physiques et acoustiques qui seront apportées pour évaluer les notions de calme.

21. Deroche Mickaël :

Données psychoacoustiques sur l'annulation d'interférence harmonique.

Un mécanisme d'annulation d'interférence harmonique pourrait, entre autres, contribuer à la ségrégation de voix concurrentes par différence de fréquence fondamentale (F_0). Quatre expériences explorent le fonctionnement d'un tel mécanisme en examinant la capacité des auditeurs à détecter une bande cible de bruit, large de 100-Hz, masquée par un complexe harmonique ou inharmonique. La première expérience montre que la détection du bruit est meilleure quand le bruit est masqué par le complexe harmonique que par le complexe inharmonique ; ceci étant vrai pour des bandes de bruit localisées jusque vers 3 kHz. Dans l'expérience 2, la différence des seuils de détection du bruit, masqué par les complexes harmonique et inharmonique, est similaire que la bande cible soit centrée sur un partiel du complexe ou centrée entre deux partiels. Les rapports de cible-surmasque, en sortie des filtres cochléaires, peuvent expliquer comment la position de la bande cible affecte les seuils. L'expérience 3 détermine la sensibilité du mécanisme à l'inharmonicité du complexe, estimée par des fonctions d'autocorrélation. Les seuils de détection ne commencent à augmenter que lorsqu'un décalage aléatoire de ± 12 Hz (pic d'autocorrélation de 0.9) est appliqué sur chaque partiel du complexe, par rapport à un complexe parfaitement harmonique. L'expérience 4 examine la largeur de bande utile au mécanisme d'annulation, en se servant de complexes inharmoniques avec une bande harmonique variable autour de la région de la cible. L'information harmonique est intégrée sur tout le spectre, mais l'inharmonicité des partiels proches de la région fréquentielle de la cible est plus bénéfique que l'inharmonicité des partiels distants. Le principal message est que le mécanisme d'annulation harmonique opère à travers plusieurs bandes fréquentielles et jusque vers 3 kHz.

22. Deprez Romain :

Vers l'adaptation automatique de contenus audio multicanaux.

A l'heure de l'introduction de systèmes grand public permettant d'aller plus loin que les systèmes stéréophoniques classiques, de nombreux défis vont devoir être résolus pour permettre une vraie démocratisation du son multicanal. Multiplicité des formats de diffusion, des systèmes de restitution, nouveaux modes de consommation... tous ces problèmes peuvent être ramenés à un seul : celui de l'adaptation automatique et optimale entre un format d'entrée et un système de restitution.

Pour ce faire, on peut distinguer deux étapes : d'un côté, il faut une connaissance précise du système de restitution disponible. C'est la calibration ou correction du système. Elle a pour objectif de compenser un mauvais placement des haut-parleurs ainsi qu'une acoustique éventuellement défavorable du lieu de restitution, mais peut-être étendue à l'ajout d'une réverbération spatiale. Les systèmes existants sont nombreux, mais peu d'entre eux ont réellement une base scientifique solide, notamment en ce qui concerne la correction de l'agencement spatial.

Ensuite, à partir de la connaissance du format d'entrée, il faut utiliser des méthodes bien adaptées afin de tirer parti au maximum du système cible. Cette étape sera appelée conversion du contenu. Une première solution permettant cette conversion est la "virtualisation" et fait appel aux technologies binaurales ou transaurales. Ces technologies sont intégrées dans nombre d'équipements actuels sous le terme de "virtual surround". Une autre solution est l'enrichissement spatial, qui est réalisé par les procédés dits d'"upmix". Ces deux types d'opérations peuvent éventuellement former des systèmes hybrides.

La présentation aura pour sujet le travail effectué pendant ma première année de thèse. Une étude objective et subjective de l'efficacité d'un système de correction existant sera présentée. Elle montre que la correction n'est vraiment intéressante que pour les systèmes de qualité médiocre. Des pistes seront proposées pour répondre aux problèmes posés. Elles s'appuient sur une représentation de l'information spatiale par le format Higher Order Ambisonics (HOA), qui a le double avantage d'être hiérarchique et indépendant à la fois du format d'entrée et du système de restitution. HOA constitue un espace dual pertinent dans lequel l'information peut être analysée et traitée.

23. Dessein Arnaud :

Applications de la géométrie de l'information à l'analyse et au traitement des flux audio.

La géométrie de l'information est une branche récente des mathématiques, et plus particulièrement de l'inférence statistique, qui étudie les notions de probabilité et d'information par le biais de la géométrie différentielle. Cette présentation s'attachera à introduire les concepts et notion théoriques de la géométrie différentielle, notamment de la géométrie riemannienne, utiles au développement d'un cadre mathématique formel pour la manipulation des flux de données audio. L'idée est d'arriver à des structures de manipulation alternatives, respectant plus les natures temporelle et probabiliste des flux audio que les structures habituellement utilisées dans les nombreuses applications de l'analyse de contenus audio. Deux champs applicatifs de ce cadre formel de manipulation sont envisageables : (1) l'analyse de contenus audio pour des applications comme l'apprentissage automatique de structures, la segmentation automatique des flux, la reconnaissance automatique de scènes sonores ; et (2) la transformation des sons avec des applications dans la restauration d'enregistrements, l'encodage et la compression des données, ainsi que le développement de nouvelles voies de transformation des sons dans un schéma d'analyse-synthèse.

24. Destephen Clément :

Synthèse sonore de véhicules en déplacement à vitesse variable.

L'objectif de mon stage est d'améliorer une technique de simulation sonore de véhicule. Cette technique est basée sur l'analyse/synthèse, utilisant des enregistrements de bruits de véhicules. On part d'un enregistrement correspondant à une certaine montée en régime du moteur. L'étape d'analyse commence par l'extraction d'informations caractéristiques de la partie harmonique. Cette dernière est alors synthétisée puis soustraite à l'enregistrement original, en résulte alors la partie stochastique. Cette partie est également analysée (par bandes d'octaves). Dans un second temps, ces informations sont utilisées pour effectuer la synthèse de n'importe quelle montée en régime. L'objectif final étant de pouvoir effectuer cette synthèse en temps réel.

25. Devergie Aymeric :

Influence de la lecture labiale sur la ségrégation de flux de parole.

L'analyse de scènes auditives, décrite par Bregman (1990), désigne notre faculté à séparer et regrouper les différentes sources sonores qui nous entourent pour former des flux auditifs cohérents. Cette aptitude nous permet de comprendre le propos d'un locuteur dans un environnement bruyant. Dans ce contexte d'écoute concurrentiel, il a été démontré qu'en regardant le mouvement des lèvres, l'intelligibilité était améliorée (Sumbly et Pollack, 1954).

De nombreuses études, mettant en jeu des tâches d'identification de segments de parole (voyelles, syllabes), suggèrent que le bénéfice apporté par la lecture labiale repose sur des interactions phonétiques entre les voyelles visuelles et auditives (Massaro, 1983). Plus récemment, Kim et Davis (2001) ont démontré que la détection de la parole dans le bruit était facilitée quand l'auditeur pouvait voir les lèvres du locuteur. La détection étant un processus précédant l'identification, nous pouvons alors supposer que l'interaction audio-visuelle impliquée est un processus bas niveau d'un point de vue cognitif.

Afin de tester cette hypothèse, nous proposons des séquences de 6 voyelles dont la fréquence fondamentale est soit identique pour toutes les voyelles (perception d'un seul flux de 6 voyelles) ou bien alterne entre deux valeurs (perception de 2 demi flux de 3 voyelles). Cette ségrégation séquentielle auditive, reposant sur une différence de fréquence fondamentale, est réputée être un processus bas niveau. Simultanément à une voyelle audio sur deux, nous proposons différents mouvements de lèvres visuels articulant la voyelle audio correspondante. Les deux études qui ont été réalisées, dans lesquelles on mesure l'état de ségrégation sous influence de l'indice visuel, démontrent que si le mouvement des lèvres est congruent d'un point de vue phonétique avec le flux audio correspondant, celui-ci vient renforcer la ségrégation séquentielle. L'interaction entre les voyelles visuelles et auditives dans ce contexte de ségrégation séquentielle auditive est alors une interaction précoce et pré phonétique.

26. Doukhan David :

Modèles de l'expressivité pour la synthèse de récits courts.

Si les systèmes de synthèses actuels sont généralement suffisants pour lire des phrases de façon neutre, il sont très vite pénibles à écouter, en particulier pour des textes assez long (plusieurs paragraphes). Les systèmes de synthèse ne sont guère capables de rendre expressif une narration. De même, les capacités motrices des robots humanoïdes sont actuellement peu exploitées et développées pour l'expression par le geste et la posture.

Ce projet de recherche porte sur la synthèse expressive audiovisuelle de récits courts. Le projet comprend deux aspects principaux. Dans une phase d'analyse, il s'agit de traiter automatiquement des textes, courts récits de type « contes pour enfant », afin d'en extraire un contenu pragmatique, sémantique, dialogique, narratif et émotionnel. Ce contenu servira dans une seconde phase d'une part à la synthèse de prosodie expressive, et d'autre part à alimenter un modèle comportemental en termes de postures, de gestes et autres mouvements du robot humanoïde NAO.

27. Dubois Françoise :

Détection de bruits émergents au sein d'habitacles.

Les bruits parasites dans les habitacles d'automobiles ou dans les voitures de trains sont gênants pour les passagers. Ces bruits (appelés signal par la suite) sont plus ou moins masqués par le bruit de fond du véhicule (le masque). Le but final est de proposer un modèle de détection des signaux émergents, et de l'appliquer à quelques cas réels. Dans le cas des bruits d'habitacles, on se trouve dans des environnements complexes où il est possible que de nombreux types de masquages interviennent.

A partir d'une situation simple à savoir la détection de sons purs (sept fréquences différentes, multiples de 80 Hz) dans un bruit blanc, un bruit de train et un bruit d'automobile (enregistré en soufflerie), il a été choisi d'effectuer les expériences au casque, en diotique (même signal sur les deux oreilles) ainsi qu'en champ libre, face à une enceinte monophonique en chambre sourde. Nous discuterons les données expérimentales récoltées afin de comparer ces deux modes de présentation. La prochaine étape de mon travail concerne la détection d'un signal multifréquentiel. Un son complexe composé de sons purs détectables de manière identique peut être détecté alors que chaque composante est inaudible individuellement, même si les composantes excitent des canaux auditifs différents. Le seul de détection d'un signal multifréquentiel ne se déduit donc pas facilement des connaissances sur la détection d'un son pur. Divers paramètres seront à examiner : niveaux relatifs entre partiels, différence de fréquence entre partiels, nature du masque (large bande puis auquel viennent s'ajouter également des harmoniques).

28. Elie Benjamin :

Caractérisation acoustique et vibratoire des instruments à cordes en vue de leur catégorisation.

Les différents d'appréciation entre les instruments de la part des musiciens ou des luthiers recouvre des aspects de rendu sonore, d'expressivité, de facilité de jeu ou encore de qualité de fabrication (finition, robustesse,...). Cette multiplicité des facteurs qui concourent à l'élaboration d'un instrument réussi rend difficile la recherche des voies d'amélioration : l'instrument idéal, en tant que cible à atteindre, n'existe pas. En revanche, les aspects cités plus haut dépendent en partie de facteurs pouvant être déduits de la mesure même s'ils sont liés pour une autre part à un jugement subjectif.

L'objectif de la thèse préparée est de disposer *in fine* d'un ensemble de valeurs numériques caractéristiques découlant de la mesure et permettant de dégager des classes d'instruments.

La collaboration entretenu avec certains luthiers participant au projet nous a permis d'obtenir des résultats qui seront présentés lors des JJCAAS. A l'heure actuelle, nous disposons d'un ensemble de données issues de mesures vibratoires sur des guitares et des harpes, nous permettant de regrouper des instruments selon certains critères.

Afin de vérifier la pertinence des critères retenus, des signaux de synthèse hybrides, autorisant des modifications sélectives de certaines variables seront utilisés. Les résultats des essais effectués au laboratoire d'acoustique de l'ENST, dans le but de réaliser une synthèse fidèle de la sonorité de guitares équipées de chevalets instrumentés seront présentés.

29. Esling Philippe :

Etude des mélanges sonores et orchestration.

De manière globale, l'orchestration peut être vue comme l'art des mélanges instrumentaux. Celui-ci a fait l'objet de nombreux traités et études qui l'abordent comme un savoir empirique. L'orchestration, malgré son importance capitale pour les compositeurs, est une des seules composantes musicales à ne pas avoir été abordée d'un point de vue scientifique. Avec le projet ANR « Sample Orchestrator », l'aide à l'orchestration s'est constitué comme un nouvel axe de recherche à l'Ircam. Deux thèses ont été réalisées dans le cadre de ce projet, [Tardieu08] et [Carpentier08]. Dans un premier temps, à la question très générale de savoir comment produire un timbre donné avec un orchestre donné s'est substituée celle plus restreinte de chercher, dans un ensemble d'échantillons de sons instrumentaux, une ou plusieurs combinaisons d'éléments se rapprochant au mieux d'un son cible fourni par l'utilisateur. Ces premiers travaux ont donné lieu à un prototype, déjà testé en situation réelle à l'Ircam. Ce prototype s'intéressant uniquement à la partie entretendue des sons et pour un ensemble de descripteurs restreints, ne fait qu'ouvrir la voie à un champ de recherche encore très vaste et inexploré qui au delà de son utilisation pour l'orchestre pourrait permettre d'étudier de manière plus globale la compréhension et le contrôle des mélanges sonores.

30. Eveno Pauline

Instruments de musique à vent : recherche d'attributs acoustiques pertinents pour la facture instrumentale.

Cette thèse s'inscrit dans le cadre d'un projet plus général, PAFI (Plateforme d'Aide à la Facture Instrumentale) soutenu par l'Agence Nationale de la Recherche. Ce projet partenarial consiste à mettre en oeuvre une plateforme logique et matérielle d'aide à la conception et à la caractérisation des instruments de musique destinée aux artisans de la facture d'instruments.

Les éléments qui concourent à l'élaboration d'un instrument réussi sont multiples. Même si l'instrument idéal, en tant que cible à atteindre, n'existe pas et ne constitue pas une notion univoque sur laquelle tous, musiciens, facteurs, mélomanes pourraient s'accorder, il est possible de déduire de la mesure des attributs directement liés à ces éléments.

L'objectif du projet est de disposer in fine d'un ensemble de valeurs numériques caractéristique découlant de la mesure et permettant d'anticiper les qualités musicales de tel ou tel instrument. Ces attributs seront issus de mesures géométriques ou acoustiques (impédance à l'entrée de l'instrument calculée à l'aide d'un logiciel spécifique ou mesurée à partir d'un dispositif appelé "pont d'impédance") ou d'enregistrements sonores (réalisés avec des musiciens ou au moyen de "bouches artificielles" permettant une excitation contrôlée et reproductible de l'instrument). Cette méthodologie qui s'appuie à la fois sur de la recherche en modélisation physique des instruments, et en traitement du signal, sera appliquée à un large éventail d'instruments de la famille des vents (bois et cuivres), dans le but de les comparer, de faire émerger des catégories ou de classer un instrument parmi ses pairs. Des simulations par modèle physique autorisant des modifications sélectives de certaines variables (comme l'harmonicité des résonances du résonateur ou leur facteur de qualité par exemple) pourront être utilisés pour rechercher les seuils de pertinence des attributs retenus.

31. Fdili Alaoui Sarah :

Reconnaissance du geste expert et contrôle de systèmes complexes : cas de l'interface entre le geste dansé et le rendu graphique par modèle physique.

D'une part, les interactions gestuelles sont souvent très recherchées en communication Homme Machine car le geste est une modalité d'entrée porteuse de sens et d'expressivité comme cela est le cas dans les domaines de la création et en particulier la danse. D'autre part, les modèles complexes tels que ceux de la mécanique des fluides intéressent, en raison de leur réalisme et le leur analogie avec des phénomènes naturels, les domaines de la danse et de la musique. Ces deux observations nous conduisent à nous intéresser à la convergence de ces deux problématiques vers l'étude et la reconnaissance en temps réel du mouvement dansé et son utilisation à des fins de contrôle gestuel de modèles physiques complexes et leur rendu graphique interactif.

La thèse comporte un premier axe orienté analyse du geste dansé segmenté en trois parties : capture de mouvement, traitement préliminaire des données et reconnaissance du mouvement dansé par des méthodes statistiques. Elle comporte aussi un deuxième axe orienté contrôle de modèles complexes segmenté en deux parties : correspondance entre le mouvement dansé et le rendu visuel puis rendu graphique des modèles physiques en lien avec les paramètres extraits de l'analyse du geste.

L'enjeu de la thèse est d'une part de parvenir à un couplage fort entre les paramètres du geste liés à la qualité du mouvement et les descripteurs de haut niveau liés à l'intention et l'expressivité du danseur. Et d'autre part, l'enjeu est de parvenir à mettre en place une interface Homme Machine dont l'entrée serait le geste du danseur « utilisateur » de l'interface et la sortie le rendu graphique du modèle physique. Le travail ainsi réalisé dans le cadre assez spécifique du lien entre le geste dansé et les modèles physiques pourra avoir des applications dans d'autres domaines associant geste créateur et modèles complexes tels que la composition et la performance musicale, le design et la création graphique.

32. Fontan Lionel :

Intelligibilité, perception et compréhension de la parole pathologique.

L'intelligibilité de la parole est une mesure apparue à la fin des années 1920 (Fletcher et Steinberg, 1929), avec le besoin pour les ingénieurs d'évaluer différents systèmes de transmission téléphonique. L'intelligibilité était alors calculée comme le pourcentage de mots (ou pseudo-mots) correctement transmis par un locuteur X à un auditeur Y. Cette méthode a ensuite été transposée dans le domaine clinique (Tikofsky & Tikofsky, 1964) pour évaluer l'intelligibilité de personnes souffrant de dysfonctionnements au niveau de l'appareil phonatoire.

Les objectifs techniques de la télécommunication étant différents des objectifs communicationnels d'un être humain (resp. transmettre du code sans perte VS se faire comprendre), la question se pose du rapport entre mesures de l'intelligibilité et compréhension des auditeurs. Une étude récente (Hustad, 2008) a notamment montré que de bons scores d'intelligibilité n'impliquaient pas nécessairement une bonne compréhension chez les auditeurs, et inversement. En outre, mesurer l'intelligibilité de la parole à travers des tâches de retranscription de phonèmes implique de considérer pour acquis des hypothèses qui sont très loin de faire l'unanimité chez les chercheurs travaillant sur la perception de la parole, notamment :

- qu'un lien «simple et direct» existe entre perception et (re)production ;
- que l'identification des phonèmes est la condition nécessaire et suffisante pour la compréhension de messages parlés (cf. notamment les différentes hypothèses concernant les unités perceptives de base dans (Nguyen, 2005)).

Afin de dépasser toutes les questions relatives aux processus de perception, nous proposons de juger directement la compréhension des auditeurs à travers l'observation de leurs actions. Ce type de situation expérimentale a été utilisé en psychologie pour évaluer la compréhension de la parole chez le bonobo et l'enfant en bas âge (Savage-Rumbaugh et al., 1993 notamment), et repose sur l'hypothèse selon laquelle la réalisation de consignes orales implique leur compréhension. Ainsi l'énoncé «posez la montre sur le chapeau» sera considéré comme compris si l'auditeur effectue l'action correspondante.

Références

Fletcher, H., & Steinberg, J. C. (1929). Articulation Testing Methods. *Bell Syst. Tech. J.*, 8, 806- 854.
Hustad, K. C. (2008). The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria. *J Speech Lang Hear*

Res, 51(3), 562-573.

Nguyen, N. (2005). La perception de la parole. Dans *Phonologie et phonétique* : Forme et substance (pp. 425-447). Paris : Hermès.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language Comprehension in Ape and Child. *Monographs of the Society for Research in Child Development*, 58(3/4), 1-221.

Tikofsky, R. S., & Tikofsky, R. P. (1964). Intelligibility measures of dysarthric speech. *Journal of Speech and Hearing Research*, 7, 325-333.

33. Foucard Remi :

Similarité multimodale pour la recherche de reprises d'un morceau.

La similarité musicale reçoit un intérêt toujours grandissant, compte tenu du nombre d'applications potentielles qui peuvent s'appuyer sur ce concept. L'identification de reprises, qui consiste à rechercher différentes interprétations d'un même morceau, est un aspect spécifique de la similarité musicale.

Nous verrons au cours de cette présentation que l'on peut proposer de nouvelles manières d'exprimer la similarité en considérant séparément, pour chaque morceau, le mélange, une estimation de la mélodie principale et de l'accompagnement. Dans le contexte de la détection de reprises, nous verrons que les meilleures performances ont été obtenues en considérant le mélange et l'accompagnement estimé. Comme nous le démontrons dans les expériences que nous avons menées sur deux différentes bases de données, ce système permet à l'analyse de similarité de se concentrer davantage sur la progression d'accords en considérant l'accompagnement, tout en restant robuste aux possibles erreurs de séparation en considérant le mélange.

34. François Clément :

Statistical learning of a sung language : behavioural and ERP studies to compare musicians and non musicians adults and normal readers and dyslexics children.

When facing a new language for the very first time, one perceives it as a continuous stream of meaningless sounds. In order to learn a new language, it is necessary for the learner to succeed in segmenting this continuous stream into significant units : the words. This task remains very difficult because of the lack of acoustic cues at each words boundaries. Some behavioral studies have shown that human adults, human newborns and monkeys can segment a continuous stream of linguistic and non linguistic stimuli only based on implicit computations of transitional probabilities between adjacent syllables (Saffran et al., 1996a). In 2008, Schön et al. showed a faster learning when the language is sung compared to the learning of a spoken one. This facilitation is assumed to be due to a reinforcement of the linguistic structure by the musical one or by an enhancement of subject's motivation due to music (Schön et al., 2008). Nevertheless, the behavioral approach doesn't give any information about the neural processes implicated during learning and can underestimate what has been learned (Mc Laughlin et al., 2004). In the present experiment we coupled the Schön et al study with an electrophysiological approach. We recorded ERPs during the learning phase and during the two tests sessions : one concerning the linguistic dimension and the other concerning the musical dimension. We conducted this experiment on musicians and non musicians adults and on dyslexics and non dyslexics children. We analyzed the data of the tests sessions and showed in the linguistic dimension a larger negativity around 650 msec for familiar than for unfamiliar items. In the musical dimension, we found a different pattern with a negativity followed by a positivity. These components were also sensitive to the degree of familiarity of the tone sequences.

We will discuss the differences found between linguistic and musical sessions, then the differences found between the populations and finally the possible implications of this protocol as a diagnostic tool for dyslexia.

35. Freitas Caetano Marcelo :

Applications des Algorithmes Génétiques en Synthèse et Transformation Sonore par Descripteurs.

La synthèse sonore par descripteurs est une approche prometteuse pour obtenir des transformations perceptivement pertinentes sous l'hypothèse que les descripteurs capturent bien les dimensions perceptives prédominantes. Néanmoins, il est difficile d'obtenir une description exhaustive sur le plan perceptif, ce qui signifie qu'il existe des sons perceptivement différents qui ont les mêmes valeurs des descripteurs. Aussi, le simple appariement des valeurs prises par les descripteurs pour deux sons source et cible n'est pas suffisant pour une transformation correcte de la source vers la cible. On présente des techniques pour obtenir une représentation intermédiaire entre la source et la cible et on utilise un algorithme génétique pour apparier les valeurs des descripteurs. Cette méthode permet la manipulation indépendante des descripteurs de forme spectrale en gardant le nombre de pics désiré.

36. Frey Aline :

Les Unités Sémiotiques Temporelles : des figures sonores ?

Tout comme il existe des figures géométriques de base qui interviennent dans la perception visuelle, nous défendons l'idée selon laquelle il existe des figures temporelles dans la perception auditive. Nous nous intéressons plus particulièrement aux Unités Sémiotiques Temporelles (UST), qui sont définies comme des segments musicaux qui véhiculent une signification de part l'organisation de la matière sonore dans le temps. Ces segments musicaux sont regroupés en 19 catégories.

Dans des tâches de catégorisation, nous avons testé leur pertinence cognitive, ainsi que celle des Motifs Temporels Paramétrés (MTP), qui sont une modélisation des UST à l'aide de fonctions temporelles. La validité des UST comme support d'un contenu sémantique a aussi été testée dans une tâche d'amorçage, durant laquelle nous enregistrons l'activité électrique du cerveau (Potentiels Evoqués).

Nous savons que la perception auditive mobilise des connaissances pour structurer les éléments perçus, et que les connaissances de l'expert ne sont pas celles du novice. Nous explorons donc également l'influence de l'expertise à travers nos différentes expériences.

Nos principaux résultats ne montrent pas d'effet de l'expertise musicale dans les tâches de catégorisation des UST et des MTP. En outre, nos résultats valident la pertinence psychologique des UST qui sont bien perçues comme un tout, comme une forme sonore. Cette interprétation est confortée par l'analyse des données électrophysiologiques, qui montrent qu'un changement incongru de concept au sein d'une UST suscite des variations spécifiques au niveau des Potentiels Evoqués. Le travail de théorisation sur les UST mérite cependant d'être approfondi et certaines catégories d'UST doivent être mieux définies.

37. Fuentes Benoit :

Une stratégie fondée sur l'ajout de bruit pour l'extension de bande en audio.

La compression, la transmission ou le stockage d'un signal audio se traduit souvent par des pertes d'information. Parmi elles, on peut trouver une coupure des fréquences les plus aiguës (c'est le cas par exemple pour la transmission téléphonique pour laquelle on coupe les fréquences supérieures à 4000 Hz). Or, il est possible dans une certaine mesure de reconstruire de l'énergie dans les hautes fréquences à partir de la seule information contenue dans les basses fréquences. On parle alors d'enrichissement spectral aveugle (ou extension de bande). L'originalité de la démarche ici proposée est qu'elle repose sur de l'ajout de bruit coloré en hautes fréquences pour enrichir le signal, et cela grâce à un modèle paramétrique de signal audio.

38. Goyet Louise :

Le problème de la segmentation de la parole en mots en français : études électrophysiologiques.

Problématique : Acquérir une langue nécessite d'extraire la forme sonore des mots (le processus de segmentation) car peu de mots isolés sont entendus dans la parole. Comment les enfants arrivent-ils à extraire ces unités de sons et ce dès 8 mois ? Nazzi et al. (1998) ont proposé l'**hypothèse d'initialisation rythmique** qui sous-tend que la segmentation se met en place sur la base d'une sensibilité précoce aux propriétés rythmiques (ex. : unité de stress : anglais/hollandais ; syllabe : français ; mora : japonais) de la langue maternelle. Nazzi et al. (2006) ont montré chez les enfants français une segmentation en syllabes indépendantes dès 12 mois. Or, des données chez des enfants hollandais (Kooijman et al., 2005, 2007) suggèrent que des effets de segmentation plus précoces pourraient être obtenus avec une procédure expérimentale différentes : la méthode des potentiels évoqués (ERPs).

Objectifs : mise en évidence d'un effet de segmentation syllabique chez les enfants français de 12 mois au moyen des ERPs.

Étude 1

Méthode : 16 bébés français ont entendu 20 blocs expérimentaux. Chaque bloc contient une familiarisation (10 répétitions d'un mot bisyllabique, ex. : "guitare"), suivie d'un test (6 phrases contenant le mot familier et 6 autres un mot contrôle, ex. : "dorado"). Pour le test, des mesures d'amplitude moyenne ont été calculées sur des fenêtres temporelles de 150 ms, sur les syllabes initiales et finales des mots cibles et contrôles.

Résultats : Une ANOVA pour la syllabe initiale montre un effet de la familiarité dès 350-500 ms, $p = .01$. L'effet correspond à une négativité plus grande pour les items familiers, et est plus marqué sur les quadrants temporo-pariétal et fronto-central droit du scalp. Une analyse similaire pour la syllabe finale ne montre aucun effet. Supposant un possible masquage par l'effet de la première syllabe, nous avons répliqué cette étude en familiarisant les enfants avec la syllabe finale.

Étude 2

Méthode : Même procédure et analyses de l'étude 1.

Résultats : Une ANOVA pour la syllabe finale montre une interaction significative entre les facteurs quadrants et familiarité, $p = .045$. Effet similaire à l'étude 1, qui émerge dès 350-550 ms et est plus marqué sur les quadrants fronto-central gauche et droit du scalp.

Conclusion : L'ensemble de ces données permet de valider l'hypothèse d'initialisation rythmique.

39. Gulluni Sébastien :

Segmentation interactive de Musique Electroacoustique.

We present an interactive approach for the segmentation of Electro-Acoustic music by focusing on the local timbre variations. For this purpose, a preliminary low-level segmentation is achieved with an onset detector (detection of music events) and an agglomerative hierarchical clustering algorithm is used to cluster inter-onset segments. Subsequently, two alternative scenarios are compared in which the user feedback is used to update clustering. Evaluation is done on both synthetic and real world data. The results show that a simple interactive method can improve significantly the performances of the resulting segmentation.

In marked contrast to other more conventional musical forms, the composers of Electro-Acoustic Music work directly with the sound material using recording techniques. Apart from a very few exceptions, the composers have not created a symbolic representation of their pieces that could be assimilated to a score sheet. This renders the analysis and study of this type of music quite complex and totally user-centered, hence our work towards developing adaptive systems capable of analyzing and structuring Electro-Acoustic Music in a semi-automatic fashion using user relevance feedback, which to the best of our knowledge is completely novel.

In the present work, we focus on one of the most important structural aspects of contemporary music, i.e. Timbre. The importance of low level timbral features has already been evidenced in previous works on music segmentation and structuring where the aim is to decompose a conventional piece (generally pop/rock music) into high level mysections such as intro, verse, bridge, ... However our concern is significantly different from such proposals in the sense that our primary objective is to obtain a segmentation of the music into short-term segments of homogeneous timbre in a non-supervised fashion. Here, segments duration may vary from fractions of seconds to tens of seconds. For example, a sequence of a piano note, a synthetic sound, a trumpet phrase and again a piano note would be labelled : Timbre1, Timbre2, Timbre3, Timbre1. The diversity of timbres to be considered (natural environmental sounds, synthetic sounds and instrumental sounds) may open new fields of applications such as the analysis of audio documents which are not structured in the same way as conventional music (movie soundtrack, audiovisual documents or ambiances). We propose a simple yet novel timbre segmentation architecture based on an original interactive clustering method that allows us to achieve high accuracy by exploiting minimal user feedback. It is shown that very simple, hence tolerable feedback significantly improves the segmentation performance compared to a classic totally automatic approach.

40. Hennequin Romain :

Introduction de modèles source/filtre dans la factorisation en matrices non-négatives de données audio.

La factorisation en matrices non-négatives (NMF) est un puissant outil de réduction de rang de données de grande dimension. Cette technique a l'avantage de fournir une décomposition des données dont les éléments constitutifs sont porteurs de sens et ont une signification d'un point de vue perceptif pour des données audio.

Appliquée à des spectrogrammes de signaux audio, la NMF fournit une décomposition en une partie strictement fréquentielle (composée d'atomes fréquentiels) et une partie purement temporelle (les activations temporelles associées à chacun de ces atomes). Pour un signal de musique composé de notes, idéalement, chaque atome fréquentiel correspond à une note et l'activation temporelle de cette note correspond à la présence ou non de la note aux différents instants. Cette puissante propriété de la NMF a donné lieu à de nombreuses applications dans le domaine de la séparation de sources sonores et de la transcription automatique sur partition.

La NMF est particulièrement adaptée dans le cas de signaux dont les composantes (les notes) sont stationnaires, c'est-à-dire ne présentent pas d'importantes variations spectrales au cours du temps. Cependant de nombreux signaux musicaux peuvent présenter des composantes dont le contenu spectral peut varier fortement (la fréquence fondamentale restant la même), par exemple de la voix. La NMF n'est généralement pas capable de décomposer intelligemment de tels signaux : de nombreux atomes sont alors nécessaires pour un unique événement musical et la signification perceptive de chaque atome est alors perdue.

Ma présentation portera sur l'introduction de modèles source/filtre dans la NMF : ceux-ci permettent de prendre en compte des événements musicaux (notes) présentant un contenu spectral variant fortement dans le temps, tout en gardant la signification perceptive de la décomposition. Des exemples de telles décompositions source/filtre seront présentés.

41. Huguet Laurent :

Caractérisation de phénomènes d'interactions source-filtre en voix parlée et chantée.

Le modèle à la base de la production vocale humaine est un modèle linéaire source-filtre, qui a été exposé par Fant en 1960. La vibration des cordes vocales, ainsi que les éventuelles strictionnements ou occlusions du conduit vocal, génèrent des sources acoustiques. L'onde acoustique ainsi produite se propage dans le conduit vocal et est rayonnée aux lèvres vers le milieu extérieur.

Le conduit vocal agit comme un résonateur, en accentuant certaines fréquences. Il est modélisé par un ensemble de filtres dont les fréquences de résonance sont appelées formants.

Ce modèle s'appuie sur l'hypothèse très forte que les parties « source » et « filtre » sont indépendantes l'une de l'autre. Cette hypothèse d'indépendance entre les gestes phonatoires et articulatoires semble pourtant mise en défaut dans la pratique quotidienne des thérapeutes de la voix et des professeurs de chant. Les méthodes classiques de filtrage inverse n'apparaissent efficaces que dans des situations précises, comme par exemple en parole pour des voyelles ouvertes.

Pour mesurer les propriétés acoustiques du conduit vocal sans faire d'hypothèse à priori sur la source glottique, une méthode est développée depuis quelques années par une équipe australienne. Il s'agit d'une technique de mesure directe de l'impédance acoustique d'un conduit. Appliquée à la parole et au chant, cette technique consiste à injecter une source de débit acoustique calibrée dans la bouche d'un sujet pendant que celui-ci prononce des voyelles tenues. Le débit acoustique est donc connu et la pression acoustique est mesurée en parallèle au même endroit. L'impédance acoustique est déduite du rapport entre pression et débit acoustiques, et sa partie réelle présente des maxima aux fréquences de résonance du conduit vocal du sujet, qui sont alors détectées.

42. Joder Cyril :

Étude des descripteurs acoustiques pour l'alignement temporel audio-sur-partition musicale.

L'alignement temporel entre une interprétation musicale et sa partition consiste à associer, à chaque instant de l'exécution, la position correspondante dans la partition (ou réciproquement). Dans une perspective d'indexation, nous nous intéressons au problème où l'interprétation est un enregistrement audio, dans le cas hors ligne (où tout l'enregistrement est connu).

Les systèmes d'alignement peuvent être décomposés en deux niveaux : le bas niveau, qui correspond à l'analyse « instantanée » de l'audio, et le haut niveau qui modélise la dimension temporelle de la partition. La plupart des travaux traitant du suivi de partition mettent l'accent sur le modèle temporel utilisé, et le choix des descripteurs de bas niveau n'est pas toujours clairement motivé. Il n'existe pas à notre connaissance d'étude comparant l'efficacité des descripteurs utilisés pour cette tâche. De plus, l'évaluation des systèmes d'alignement a été jusqu'à présent limitée presque exclusivement à de la musique classique monophonique ou faiblement polyphonique. L'objectif de cet article est d'étudier spécifiquement l'influence de la paramétrisation de bas niveau pour deux stratégies d'alignement. Cette étude s'étend à de la musique populaire polyphonique multi-instrumentale avec présence éventuelle de percussions, ce qui représente le cas le plus général.

Différentes représentations spectrales de l'état de l'art sont employées dans un cadre formalisé avec deux stratégies d'alignement. Les résultats montrent que les descripteurs prenant en compte une dimension perceptive (échelle fréquentielle logarithmique) sont les plus pertinents ; ils sont notamment plus robustes pour le cas polyphonique. De plus, les performances relatives des différents descripteurs ne sont pas modifiées par la présence d'erreurs dans la partition. Enfin, la stratégie d'alignement à partir d'une resynthèse de la partition obtient des résultats globalement meilleurs que les modèles théoriques, même si sa complexité est supérieure.

43. Karkar Sami :

Bifurcations des régimes périodiques dans les systèmes dynamiques non linéaires. Application aux instruments de musique auto-oscillants.

Les systèmes dynamiques non linéaires présentent parfois un comportement auto-oscillant : bien que de l'énergie soit apportée au système de façon continue, des oscillations apparaissent. C'est en particulier le cas des instruments à vent et à corde frottée. Le passage d'un régime statique à un régime périodique est caractérisé par une bifurcation. D'autres bifurcations existent parfois et permettent de passer d'un régime périodique à d'autres régimes : périodique mais de période différente, quasi-périodique (parfois appelés multiphoniques), chaotique, etc... Nous nous intéressons en particulier à un modèle de clarinette et cherchons à déterminer dans un premier temps les seuils de naissance et d'extinction des auto-oscillations en fonction des différents paramètres de l'instrument et de son contrôle par l'instrumentiste.

Nous abordons le problème du point de vue des systèmes dynamiques non linéaires et de leurs bifurcations, contexte dans lequel de nombreux outils, tant théoriques que numériques, sont disponibles. Toutefois, la plupart de ces outils ne sont pas adaptés à des modèles non réguliers comme le sont les instruments à anche simple (non régularité de placage) ou à corde frottée (frottement de Coulomb).

La connaissance des nombreux régimes non standard qui peuvent apparaître au cours du jeu, ainsi que des jeux de paramètres qui favorisent ou au contraire excluent ces comportements constituent un atout précieux tant pour la facture instrumentale, l'apprentissage, la composition, ou encore la synthèse par modèles physiques.

44. Karypidis Charalampos :

Le paradigme "même différents" et les asymétries en perception vocale : comparaison des modèles de discriminabilité et enjeux théoriques.

La capacité de discriminer les sons dépend souvent de l'ordre de leur présentation. L'effet de l'ordre du stimulus aussi connu en tant que "phénomène d'asymétrie" peut être résumé dans la définition de Polka (2003) : "Asymétrie in vowel perception occur such that discrimination of a vowel change presented in one direction is easier compared to the same change presented in the reverse direction". Afin de saisir la nature de ces asymétries et identifier leur origine, il est nécessaire de définir les processus activés pendant le traitement des stimuli. Divers modèles explicatifs du paradigme "même-différent" en discrimination ont été formulés (perception catégorielle, prototypes et aimants perceptifs, périphéricité, dégradation/neutralisation des traces mémorielles, points d'ancrage perceptifs, trace/contexte, processus duels), chacun faisant différentes hypothèses sur le rôle des représentations mentales et prédisant différents résultats au niveau des résultats bruts et des temps de latence. Nous proposons ici de comparer ces modèles. Nous commençons par la présentation du modèle de la mémoire proposé par Haber & Hershenson (1980) qui suppose l'existence d'un dépôt auditif temporaire et d'une image mentale auditive, ainsi que des interactions explicites entre ces modules et la mémoire à court terme et à long terme. Ensuite, nous interprétons chaque modèle de discriminabilité au moyen des processus décrits par Haber & Hershenson. Ce travail peut avancer notre compréhension du phénomène en ce qu'il dresse un plan des théories et nous aide à élaborer des protocoles plus spécifiques pour étudier les divers paramètres déclenchant de effets d'ordre.

45. Le Bagousse Sarah :

Méthodologie de test pour l'évaluation subjective de la qualité du son spatialisé.

Aujourd'hui, l'enjeu des services audiovisuels (cinéma, télévision...) est de reproduire des scènes visuelles et sonores de la façon la plus réaliste possible. L'essor des différentes technologies, particulièrement la spatialisation, amène à modifier la pensée en terme de qualité. Avant leurs mises en diffusion sur l'ensemble des réseaux, l'appréciation de la qualité des contenus audio multicanaux est incontournable. Il est donc nécessaire de comprendre les mécanismes influant sur la perception du son et d'en ressortir les principaux axes de jugement de la qualité de l'audio en spatial. Les méthodologies actuelles de tests, notamment de type MUSHRA (« Multi Stimulus test with Hidden Reference and Anchor », voir ITU-R BS.1534), ne prennent pas en compte la dimension spatiale du son. Dès lors, il faut définir de nouveaux critères d'évaluation subjective de la qualité du son produit par les systèmes de restitution tel que le 5.1, l'ambisonie, le binaural... Un des travaux préliminaires est de recenser et d'analyser les caractéristiques spécifiques à la perception d'un son spatialisé. Des critères tels que le timbre, le volume, l'immersion, la différence de phase seront introduits afin de déterminer leurs degrés de pertinence dans l'évaluation de la qualité d'un son spatialisé. Les tests se feront sur des sujets dits « experts ». Les résultats obtenus seront analysés à la fois de manière linguistique et statistique. La méthode développée devra, dans l'absolu, s'affranchir du système de spatialisation utilisé et idéalement du mode d'écoute, au casque ou sur haut-parleur. Elle pourra donc s'appliquer à tous les tests sur l'audio spatialisé et se décliner sur l'ensemble des éléments de la chaîne audio (de la captation à la restitution en passant par les systèmes de compression et de transport ou autres éléments de traitement du son).

46. Leman Adrien :

Modèle basé sur le signal non-intrusif de l'évaluation de la qualité vocale, utilisant une méthode de classification automatique des bruits de fond.

Voici la description d'un modèle original d'évaluation de la qualité vocale en présence de différents types de bruit de fond, appliqué à tout les types de transmission téléphonique (RTC / RNIS / mobile / VoIP). Le modèle est construit à partir des tests subjectifs présentés dans l'article intitulé "Influence of informational content of background noise on speech quality evaluation for VoIP application", lors de la conférence Acoustics 2008. Les tests perceptifs présentés dans cet article montrent que le contenu informationnel du bruit de fond présent sur le signal vocal provoque une indigence sur la qualité vocale perçue. Les utilisateurs sont plus tolérants lors de l'évaluation de la qualité de la voix lorsque le bruit de fond est reconnu comme provenant de l'environnement du locuteur, plutôt qu'un bruit non reconnu de type électrique ou de souffle.

Quatre classes de bruit de fond sont alors déterminées suivant la gêne croissante occasionnée sur la qualité de la voix : respectivement, les classes de bruit "intelligible, d'environnement, de souffle, et de grésillement". Pour chacune de ces quatre classes, une relation est proposée entre la note de qualité vocale et la sonie du bruit de fond.

Le modèle proposé se décompose en quatre étapes : Dans un premier temps, l'outil "Voice Activity Detector" est utilisé pour séparer les zones de parole active et non active. Dans un deuxième temps, deux indicateurs basés sur le signal sont calculés sur le signal du bruit de fond. Ensuite, ces deux indicateurs sont renseignés à un arbre de décision afin de déterminer la classe du bruit de fond correspondante. Pour finir, la qualité vocale est calculée suivant une relation dépendante de la sonie du bruit de fond et de la classe du bruit de fond.

47. Liuni Marco :

Résolutions différentes pour informations différentes.

Les algorithmes pour l'analyse, synthèse et transformation des sons qui existent aujourd'hui nécessitent des fortes compétences sur le traitement du signal. Pour rendre l'accès à ces algorithmes performants plus facile, il serait souhaitable que les algorithmes s'adaptent eux même de façon automatique au signal et à la transformation souhaitée.

Un des paramètres clef pour presque tous les algorithmes qui analysent ou transforment les sons en utilisant la représentation temps-fréquence ou les modèles sinusoidaux est la taille de la fenêtre d'analyse : dans cette présentation on va montrer des exemples de la façon dont ce paramètre est lié à la résolution de l'analyse d'un signal, et quelles informations on peut privilégier en choisissant une taille plutôt que une autre.

Nous sommes actuellement en train d'étudier et développer des algorithmes pour l'adaptation automatique de la taille de la fenêtre d'analyse. Il semble possible que une stratégie appropriée soit l'utilisation et l'évaluation parallèle de plusieurs choix des paramètres et l'assemblage a posteriori du résultat final à partir des résultats individuels obtenus.

48. Loyau Fanny :

La gestion des communications multiples dans les systèmes aéronautiques complexes.

Les systèmes aéronautiques militaires atteignent aujourd'hui une grande complexité qui nécessite la prise en compte de facteurs humains et organisationnels pour permettre aux opérateurs d'atteindre un niveau de performance optimal.

Les missions actuelles impliquent systématiquement la participation de plusieurs intervenants. Ces derniers ont pour outil commun la communication verbale, cruciale pour la réussite d'une mission et le maintien d'un haut niveau de sécurité. Cependant, la communication verbale reste un outil "humain" : elle présente une part d'incertitude qui peut parfois conduire à des incompréhensions et à des dommages pour la mission, des incidents voire des accidents.

Afin d'améliorer la gestion des communications radio, une étude est menée sur les aéronés E3F – AWACS (Airborne Warning and Control System), particulièrement sujets à ce problème compte tenu du nombre élevé d'interlocuteurs susceptibles de communiquer ensemble. Les pratiques de gestion des communications radio semblent dépendre de l'expertise des membres de l'équipage. En effet, tous les instructeurs s'accordent pour dire qu'il est nécessaire de cumuler plusieurs années d'expérience en AWACS pour développer des compétences de gestion des communications.

Nos observations et la littérature scientifique nous conduisent à proposer plusieurs hypothèses rendant compte de cet effet de l'expertise : les experts disposent de plus de ressources attentionnelles pour la gestion des communications que les novices car ils routinisent les tâches techniques, de plus ils utilisent des stratégies de gestion des volumes des réseaux radio leur permettant de réduire la quantité de ressources attentionnelles utilisées pour gérer chaque communication, enfin ils développent une meilleure aptitude à diviser leur attention sur plusieurs communications simultanées.

Trois grands axes de recherche émergent donc : la routinisation et son apprentissage, le traitement des indices perceptifs et la division attentionnelle. L'étude de ces paramètres vise, à travers le développement d'outils FH, à adapter la sélection, la formation et l'entraînement des personnels aux compétences nécessaires à la gestion des communications dans les AWACS, et à plus long terme dans tous les systèmes complexes.

49. Loyet Raphaël :

Simplification perceptuelle des méthodes à base de rayons pour l'auralisation.

Le rendu sonore, i.e., auralisation en temps réel dans des volumes complexes reste à l'heure actuelle un défi. Les méthodes à base de rayons permettent uniquement de restituer les premières réflexions du son en temps réel. Dans cette présentation, une méthode de simplification à partir de données perceptives du nombre de chemins sonores à restituer est étudiée. Les chemins acoustiques entre une source et un auditeur situés dans un environnement sont classés en fonction de leur temps de parcours, leur direction d'arrivée vers l'auditeur et leur amplitude. La densité temporelle des rayons est aussi étudiée. L'environnement de démonstration SimAudio a été développé afin de réaliser des études psycho-acoustiques basées sur différentes hypothèses simplificatrices. Le résultat de chaque hypothèse peut être écouté et comparé à la solution complète comprenant toutes les réflexions du son.

50. Maazaoui Mounira :

Méthodes de séparation aveugle de sources audio par réseau de microphones.

Le principal objectif de cette thèse est de proposer de nouvelles solutions au problème de séparation aveugle de sources audio dans un milieu réel, en utilisant un réseau de microphones. Ces solutions seront appliquées à l'audition des robots dans le cadre du projet ROMEO. ROMEO est un projet labellisé par le pôle de compétitivité Cap Digital et financé par la Région Ile-de-France, la Direction Générale de la Compétitivité, de l'Industrie et des Services (DGCIS) et la Ville de Paris. Il vise à développer un robot humanoïde destiné à devenir un véritable assistant des personnes en perte d'autonomie. Nous présenterons dans ce qui suit les difficultés relatives à notre problématique, ensuite un état de l'art des principales méthodes de séparation aveugle de sources utilisant un réseau de microphones, et nous finirons par présenter nos axes de travail.

Notre but est donc de concevoir un système de séparation aveugle de sources audio en utilisant un réseau de microphones. Beaucoup de challenges sont à relever dans notre travail : Le système traitera des signaux réels puisque le robot évoluera dans un milieu réel (pièces à vivre, appartements, maisons dans notre cas), il faut donc tenir compte de la réverbération de la pièce, des éventuelles sources interférentes et du bruit ambiant, sans oublier la contrainte de l'exécution temps-réelle de l'algorithme.

Dans ce qui suit, nous présenterons quelques méthodes de séparation de sources utilisant les réseaux de microphones et issues de l'état de l'art. En premier lieu, nous avons exploré les méthodes de formation de voies : la formation de voies ou beamforming est un filtre spatial qui opère sur les sorties du réseau de microphones afin de former le diagramme polaire (le modèle de directivité) désiré. Le diagramme polaire désiré est celui qui annule les sources interférentes et qui ne laisse passer que les signaux désirés tout en minimisant le bruit capturé. Les méthodes de formation de voies se catégorisent en formation de voies indépendantes des données, formation de voies optimales et formations de voies adaptatives [2, 6]. Les méthodes de formations de voies nécessitent dans la plupart des cas la connaissance a priori des directions d'arrivées des sources, nous avons donc étudié quelques méthodes d'estimation des directions d'arrivées [2, 6]. Nous nous sommes intéressés ensuite à d'autres méthodes de séparation aveugle de sources. Ces méthodes utilisent l'analyse en composantes indépendantes [4], la décomposition modale empirique [1], le mélange de modèles où différents modèles sont utilisés pour décrire les sources [7] et la factorisation en matrices à coefficients positifs [9].

Certaines de ces méthodes ont été utilisées dans l'audition des robots qui consiste en la localisation et la séparation des sources dans l'environnement du robot [8]. Ensuite, nous avons abordé la représentation parcimonieuse des signaux audio [3] dans le but de développer des méthodes de séparation basées sur la parcimonie des sources dans un domaine donné.

Nos axes de travail porteront sur l'exploitation du réseau de microphones dans la séparation aveugle des sources d'une part en utilisant la formation de voies pour localiser et améliorer la détection du signal utile et d'autre part en essayant de trouver une représentation parcimonieuse des signaux sources en utilisant la structure particulière de notre réseau de capteurs (microphones autour de la tête du robot).

[1] A. Aissa-El-Bey, K. Abed-Meraim, and Y. Grenier. Underdetermined audio source separation using fast parametric decomposition. *Signal Processing and Its Applications*, 2007. ISSPA 2007. 9th International Symposium on, pages 1–4, Feb. 2007.

[2] Jacob Benesty, Jingdong Chen, and Yiteng Huang. *Microphone Array Signal Processing* (Springer Topics in Signal Processing). Springer, 1 edition, 4 2008.

[3] Emmanuel J. Candès. Compressive sampling. In *The international congress of mathematicians*, 2006.

[4] Pierre Comon. Independent component analysis, a new concept ? *Signal Processing*, 1994.

[6] Wolfgang Herboldt. *Sound capture for Human / Machine Interfaces*. Springer, 2005.

[7] N. Mitanoudis and T. Stathaki. Batch and online underdetermined source separation using laplacian mixture models. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(6) :1818–1832, Aug. 2007.

[8] Kazuhiro Nakadai, Tino Lourens, Hiroshi G. Okuno, and Hiroaki Kitano. Active audition for humanoid. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 832–839. AAAI Press / The MIT Press, 2000.

[9] Alexey Ozerov and Cédric Févotte. Multichannel nonnegative matrix factorization in onvolutive mixtures with application to blind audio source separation.

51. Machart Pierre : Segmentation Morphologique.

De nombreuses applications et travaux utilisant des sons enregistrés reposent sur la segmentation et la concaténation de fragments de flux audio. A l'occasion de collaborations avec des compositeurs et des artistes, nous avons pu constater qu'un événement musical récurrent ou qu'une forme sonore est souvent identifiable par l'évolution temporelle d'observations du signal. Nous souhaiterions contribuer au développement d'une méthode de segmentation innovante reposant sur l'évolution temporelle de descripteurs audio, et qui s'adapterait aux matériaux audio considérés, en interaction avec l'utilisateur.

En premier lieu, nous avons implémenté un prototype d'outil de segmentation semi-supervisée et interactive. Avec ce prototype, l'utilisateur fournit une annotation partielle du flux qu'il veut segmenter. Dans une boucle d'interaction, le système est alors capable de construire des modèles de classes morphologiques que l'utilisateur définit. Ces modèles sont ensuite utilisés pour proposer une segmentation exhaustive du flux, en généralisant les annotations de l'utilisateur.

Ces résultats reposent sur l'utilisation de Modèles Segmentaux, adaptés et implémentés pour des flux sonores représentés par un ensemble de descripteurs (MFCC). L'originalité de cette étude tient à l'utilisation de données réelles pour construire les classes morphologiques, issues de matériaux sonores divers et variés. Une méthode singulière pour construire le modèle global est alors définie en utilisant à la fois les paradigmes d'apprentissage et l'intégration de connaissances de l'utilisateur. L'approche globale de ce projet est validée par des expériences menées avec des flux de synthèse ou des sons réels (environnementaux ou des pièces musicales). Une validation plus qualitative et moins formelle tient aussi aux retours donnés par des compositeurs ayant travaillé avec nous tout au long de ce stage.

52. Mallaroni Bastien :

Etude du rayonnement de résonateurs de longueur finie.

Le but de cette thèse est de mieux comprendre la relation entre la géométrie d'un résonateur de longueur finie et son rayonnement. Pour cela nous étudions les fréquences de résonances complexes du système.

Dans un premier temps nous nous sommes intéressés au rayonnement d'un cylindre de longueur finie afin de disposer d'un cas de référence. Pour décrire la pression rayonnée nous utilisons une impédance de rayonnement intermodale décrite par Zorumski. De plus, nous avons développé une nouvelle formulation de la pression interne dépendant directement du champ externe à l'aide d'une méthode d'impédance ramenée. Nous avons ensuite cherché à obtenir les fréquences de résonances de notre système. Mais ces fréquences de résonances étant complexes (dû aux pertes par rayonnement) nous avons dû développer une nouvelle formulation de l'impédance de rayonnement de Zorumski pour qu'elle puisse s'appliquer à des fréquences complexes. A l'aide d'une méthode de Newton programmée dans un code Fortran nous avons obtenu les fréquences de résonances recherchées. Ces résultats ont été confrontés à ceux obtenus avec une méthode BEM développée dans un code Fortran ainsi qu'à des mesures expérimentales. La méthode BEM nous permet d'étudier des géométries de résonateurs complexes et ainsi de voir l'influence de la perce sur le rayonnement.

53. Maniatakos Fivos :

Une architecture pour le contrôle temporel de réinjection stylistique dans le cadre de l'improvisation assistée par l'ordinateur.

Le processus de la modélisation du style musical et de la réinjection stylistique basée sur la recombinaison des matériaux appris a déjà été élaboré dans des systèmes d'improvisation d'ordinateur (Omax, IRCAM) permettant un rendu sonore vaste et porteur d'invention. Nous sommes intéressés par l'étude des principes sous lesquels la réinjection stylistique peut être suffisamment contrôlée, afin que l'utilisateur de l'ordinateur puisse guider le processus de l'improvisation avec une certaine logique. Il s'agit donc de la conception des outils pour la modélisation et la représentation du style, ainsi que des algorithmes qui permettront la génération des flux musicaux selon les choix de l'utilisateur en temps réel et l'évolution d'une improvisation collective et mixte (ordinateurs et instruments conventionnels).

Dans cette étude, nous proposons une architecture modulaire pour l'improvisation assistée par l'ordinateur. En élaborant des structures originales émergeant de la théorie d'automates facteur oracle et des graphes de transition, notre approche permet de modéliser la distribution temporelle et l'interconnexion des patterns musicaux sous différents critères (hauteur, rythme, spectre, etc.). Nous posons la question de la complexité computationnelle pour la problématique du contrôle temporel, et de la projection des scénarios futurs, et proposons un algorithme qui réduit le problème à un problème de satisfaction des contraintes. Nous présentons quelques notions concernant l'implémentation de cette architecture sous la forme d'une librairie C++ indépendante, et d'un prototype GrAIPE (Graph Assisted Interactive Performing Environment) pour l'improvisation assistée par l'ordinateur dans Max-MSP.

54. Marie Céline :

Influence of musical expertise on tone and phoneme processing in Mandarin Chinese.

A same-different task was used to test the hypothesis that musical expertise improves the discrimination of tonal and phonological variations in a tone language Mandarin Chinese. Fourword sequences were presented to French musicians and non-musicians unfamiliar with the language and Event-Related brain potentials were recorded. Musicians detected both tonal and phonological variations more accurately than non-musicians. Moreover, tonal variations were associated with higher error rate than phonological variations and elicited an increased N2/N3 component taken to reflect the further computations necessary to process a linguistic contrast, tone, that is not relevant in French. This component developed 100 ms earlier in musicians than in nonmusicians. Finally, musicians also showed enhanced P3b components to both tonal and phonological variations. These results clearly show that musical expertise influences both perceptual processing and the categorization of linguistic contrasts in a foreign language. They show positive music-to-language transfer effects and open new perspectives for the learning of tone languages.

55. Massida Zoé :

Voice perception in cochlear implanted patients.

While cochlear implants allow good functional recovery of speech perception, performances in other auditory domains such as voice recognition remain not clearly studied. Rich information can be extracted from a voice, such as language, identity (i.e. age, gender, familiarity, etc.) or affective state. Our objective was to study more distinctly the paralinguistic voice perception in cochlear implanted patients.

We included one group of 40 normal hearing subjects (NHS) and one of 32 cochlear implanted patients (CIP) with different durations of implantation. NHS and CIP were tested with a voice detection task, a gender decision task and an emotion decision task. NHS were tested with original stimuli and with spectrally-degraded voices through vocoding with different number of channels. As expected, for all test normal hearing subjects showed good abilities which decrease as we reduced the number of frequency bands in vocoding. Patients who are newly implanted (1 day) had very poor performance in voice detection, gender and emotion decision task. While their performance seemed to improved with speech comprehension recovery, patients never reach the level observed in NHS at all the tasks even after more than 3 years post-implantation.

Current study is aimed at evaluating which one of pitch or timber is more perceived by cochlear users, and wich one is more usefull to do these tasks. We expect in an other experiment to link patients behavioral data with activity of voice area shown by Pascal Belin (Belin et al, 2000). Finally, our result could guide later therapeutic strategies.

56. Mathey Marie :

Dynamic of auditory processing.

Pilot experiments have shown that fast saccades can also be generated to auditory stimuli. Thus, when two sounds are presented simultaneously through loudspeakers on the left and right, subjects can efficiently saccade to the side where there is a voice. We will explore the conditions where this sort of performance is possible, and use a similar sort of temporal offset paradigm as the one used for vision to directly determine whether there is a processing speed advantage for biologically significant stimuli.

57. Merer Adrien :

Caractérisation perceptive de trajectoires pour la synthèse sonore.

Mon travail de recherche concerne le contrôle d'algorithmes de synthèse sonore à partir de paramètres de haut niveau. Des travaux réalisés dans l'équipe ont par exemple permis de concevoir un synthétiseur de sons d'impact qui donne à l'utilisateur la possibilité de passer continûment d'un matériau à un autre. Mon travail concerne plus particulièrement les informations sur le déplacement de sources acoustiques (trajectoire, dynamique,...) pour des sons monophoniques. La présentation s'attachera à décrire un protocole de caractérisation des différentes informations perçues par les auditeurs en vue de la résolution du problème inverse qui consiste à concevoir l'algorithme de synthèse permettant de contrôler ces différentes informations.

58. Michaud Pierre-Yohan :

Méthode adaptée aux tests perceptifs pour l'évaluation d'un grand nombre de stimuli audio.

Ces travaux se placent directement après ceux effectués par Lavandier (2005) sur les différences entre enceintes acoustiques qui ont permis de mettre en place un protocole de mesures compatibles avec les contraintes des approches physique et perceptive. Ce protocole fait appel à l'analyse des signaux et à une reproduction sous casque d'enregistrements des enceintes placées dans un environnement d'écoute normal. Il ne prend ainsi en compte qu'une partie des attributs perceptifs des enceintes, essentiellement ceux liés au timbre.

Dans un premier temps c'est l'approche psychoacoustique que nous cherchons à approfondir en se focalisant sur les méthodes utilisées pour les tests d'écoutes. Pour évaluer les différences entre enceintes et construire l'espace perceptif correspondant, Lavandier a utilisé deux méthodes : la comparaison par paires puis la classification libre. La première est la méthode la plus répandue et aussi celle qui est conseillée dans les normes sur les tests d'écoutes. En effet, elle permet d'obtenir un jugement direct de dissemblance entre deux sons présentés à l'auditeur. Le seul inconvénient est la durée du test pour un grand nombre de sons. Lavandier a donc utilisé la classification libre qui peut s'appliquer au tests d'écoute mais pour l'évaluation d'un panel plus important. La tâche pour cette méthode est de classer les stimuli en fonction de leur similarité en regroupant les sons similaires ensemble. Ainsi deux sons jugés différents apparaîtront dans deux groupes différents. Récemment, les travaux d' Aldrich et al. (2008) ont montré que la méthode de classification libre n'était pas adaptée à l'évaluation de la dissemblance car les auditeurs avaient tendance à catégoriser les stimuli qu'il entendaient à la place de juger de leur similarité.

La recherche d'une nouvelle méthode permettant l'évaluation de la similarité pour un grand nombre de stimuli a permis de trouver et d'adapter une méthode provenant de la perception visuelle (Rogowitz et al., 1998). Cette méthode permet d'évaluer la similarité d'un groupe de stimuli par rapport à une référence.

La méthode de comparaison à une référence a été adaptée à l'évaluation de stimuli audio en prenant en compte la mémoire auditive. Afin d'explorer la méthode et d'en connaître les limites et biais, il a été possible de simuler les jugements des auditeurs. Ainsi, plusieurs paramètres ont pu être caractérisés comme le nombre de stimuli ou d'auditeurs nécessaires à une estimation précise et fidèle des dissemblances. Le principe et les résultats de la simulation seront décrits. Enfin, la méthode a été appliquée à de tests d'écoute afin de retrouver les dimensions du timbre des enceintes acoustiques et ainsi comparer avec les résultats obtenus par Lavandier.

59. Michel Christophe :

Etude du rayonnement de résonateurs de longueur finie.

La modélisation de la cochlée est une alternative intéressante pour tester « in silico » des hypothèses difficiles à évaluer expérimentalement. Les modèles actuels (Zhang, 2001 ; Summer, 2002) sont généralement composés de quatre modules en cascade : i) la vitesse de vibration de l'étrier ii) la vitesse de vibration de la membrane basilaire iii) la dépolarisation des cellules ciliées internes (CCI) et la libération du neurotransmetteur (le glutamate) dans la fente synaptique iv) les trains de potentiels d'action (PA) dans les neurones. Dans le cas d'un fonctionnement normal de la cochlée, ces modèles développés initialement chez l'animal, reproduisent fidèlement les données électro-physiologiques in vivo (CAP, Peri-Stimulus Time Histogram, ...). En revanche, ils s'avèrent trop intégrés pour décrire certains fonctionnements pathologiques de la cochlée.

Le but de ce travail est de proposer une modélisation plus fine de la neurotransmission entre les CCI et les neurones afin d'étudier in silico des pathologies comme les acouphènes d'origine périphérique. Un modèle de bouton post-synaptique muni de différents récepteurs au glutamate (récepteurs AMPA et NMDA) a été développé et placé en cascade avec un modèle de neurone de type Hodgkin-Huxley. Ce modèle montre qu'une implication des récepteurs NMDA (inactifs en condition normale et actifs lors d'acouphènes) conduit à une augmentation drastique du taux de PA dans les neurones du nerf auditif. Ces résultats confirment les données publiées récemment chez le cochon d'Indes (augmentation de 250 % de taux de décharge) lors de prise de Salicylate (Ruel, Chabbert et al. 2008).

A terme, ce modèle de jonction CCI/neurone permettra d'étudier d'autres pathologies de la neurotransmission comme des surdités génétiques ainsi que l'effet de la stimulation électrique des neurones (intérêt dans le domaine de l'implant cochléaire).

Ruel, J., C. Chabbert, et al. (2008). "Salicylate enables cochlear arachidonic-acid-sensitive NMDA receptor responses." *J Neurosci* 28(29) :7313-23.
Summer, C. J., E. A. Lopez-Poveda, et al. (2002). "A revised model of the inner-hair cell and auditory-nerve complex." *J Acoust Soc Am* 111(5 Pt 1) : 2178-88.

Zhang, X., M. G. Heinz, et al. (2001). "A phenomenological model for the responses of auditory-nerve fibers." *J Acoust Soc Am* 109(2) : 648-70.

60. Moffat Graeme :

Preliminary evidence for recovery from auditory deprivation in asymmetric hearing loss after hearing aid fitting.

Asymmetric hearing impairment, whether induced by sensorineural hearing loss or unilateral hearing aid fitting, can give rise to a deficit in suprathereshold word recognition ability in the more impaired ear that has alternately been described as late-onset deprivation and a form of acclimatization (Silverman et al. 1984, Gelfand et al. 1987, Gatehouse 1987). The onset of this effect has been shown to be preventable in a clinical population with asymmetric sensorineural hearing deficits (Silverman et al. 2006) via hearing aid use, but only anecdotal reports of the reversal of late onset deprivation exist, and only in cases induced by monaural fitting. Here we report that this effect of late-onset deprivation in asymmetric sensorineural hearing loss is reliably (though perhaps not completely) reversible in a clinical population of 12 subjects, fitted with hearing aids in the more impaired ear after untreated hearing loss of greater than one year's duration, and re-measured at one to two months post-fitting. Word recognition threshold scores improved significantly in this group, and are contrasted with no significant improvements in the worse ears of binaurally fitted subjects with (N=7) and subjects without (N=9) prior asymmetry. The relatively rapid nature of the observed recovery from deprivation suggests an important role for central plasticity in between generalized auditory "feature selectors" (Gatehouse 1989) and inputs from the worse-performing ear, and these findings may warrant consideration in the fitting of hearing aids when treating asymmetric hearing loss.

61. Morel Julien :

Gène en situation de multi-exposition sonore à différents bruits industriels.

Une typologie perceptive constituée de six catégories de bruits industriels permanents stables a été établie dans une précédente étude. Les catégories ont été formées sur la base des ressemblances perçues entre les bruits. Au cours d'une autre étude, des indicateurs de gêne sonore ont été développés pour chaque catégorie perceptive. Une autre étape consiste à évaluer la gêne totale induite par des combinaisons de sources de bruit provenant de différentes catégories.

Cette présentation traite de la combinaison de bruits à large bande soit avec des bruits comportant un contenu en basses fréquences, soit avec des bruits ayant une composante spectrale à 100 Hz. La gêne totale est évaluée en laboratoire au cours de deux tests d'écoute intégrant une restitution stéréophonique des bruits (via deux haut-parleurs). Les combinaisons, formées de deux bruits, sont générées de la façon suivante : le niveau sonore d'une des sources varie, tandis que le niveau sonore de l'autre source est fixé.

De l'analyse des réponses aux deux tests, différents résultats ont été obtenus : (1) les bruits à large bande sont moins gênants par rapport aux bruits à fort contenu basses-fréquences, ou aux bruits comportant une composante spectrale à 100 Hz, pour un même niveau sonore. (2) L'effet, sur la gêne totale, du bruit dont le niveau sonore varie est significatif. Globalement, plus le niveau sonore du dit bruit augmente, plus la gêne totale se rapproche de la gêne spécifique due à ce bruit, dont le niveau sonore varie. (3) Cependant, le modèle de source dominante souvent performant pour prédire la gêne totale ne semble pas pertinent. Il est possible que cela soit dû au fait que certains bruits possèdent des pouvoirs d'inhibition différents, i.e. des capacités différentes à devenir la composante la plus gênante au sein d'une combinaison de bruits.

62. Moussallam Manuel :

Traitements du son à partir de représentations parcimonieuses.

L'outil des représentations parcimonieuses occupe depuis une dizaine d'années une place croissante dans le domaine de l'analyse-synthèse des signaux audio. Son postulat principal est qu'il existe une base dans laquelle un signal numérique peut être représenté de façon parcimonieuse, c'est à dire dans laquelle peu de vecteurs (par rapport à la dimension du signal) décrivent l'essentiel de son énergie. On appelle décomposition le processus de transformation d'un signal dans une telle base. Mais trouver une décomposition optimale (au sens de la parcimonie, du rapport signal/résiduel, ou des deux) se ramène souvent à un problème combinatoire de minimisation de norme quadratique. Des stratégies peuvent alors être mises en place (par exemple les algorithmes gloutons de type « Matching Pursuit ») pour obtenir des décompositions (parfois sous-optimales) à un coût raisonnable. La décomposition s'appuie alors sur un dictionnaire d'éléments appelés atomes qui peuvent être appris ou construits par dilatation et transposition de formes d'ondes localisées en temps et en fréquence telles que les ondelettes de Gabor ou les fenêtres MDCT. On dispose désormais de différents moyens de représenter efficacement un signal et il est également possible d'associer une certaine sémantique aux « objets » issus de la décomposition. La possibilité de réaliser nombre de traitements du son directement dans le domaine de la représentation apparaît alors comme un champ d'exploration tout à fait prometteur.

Dans cet exposé nous présenterons deux types de représentations et d'applications possibles : l'enrichissement spectral grâce à un dictionnaire appris d'atomes harmoniques, et la détection voix/musique grâce à un dictionnaire redondant de bases MDCT d'échelles différentes.

63. Necciari Thibaud :

Mesure du masquage auditif temps-fréquence, application à l'analyse-synthèse des sons.

La représentation mathématique d'un son cohérente avec la perception auditive de ce dernier est un enjeu de grande importance pour de nombreuses applications (e.g., réalité virtuelle, synthèse sonore, codage audio). Afin de répondre à une telle problématique, ce projet de thèse a pour objectifs de développer et d'évaluer de nouvelles représentations des signaux audio basées sur les théories mathématiques des représentations temps-fréquence (TF) et d'adapter ces représentations aux concepts de la psychoacoustique, en particulier les effets de masquage. Le masquage se définit comme l'élévation du seuil d'audibilité d'une source sonore (la « cible ») en présence d'une autre source (le « masque »).

Les modèles psychoacoustiques de masquage actuellement disponibles sont limités car ils considèrent les effets temporel et fréquentiel du masquage de façon indépendante et sont issus de mesures réalisées avec des stimuli mal localisés dans le plan TF. Ce constat nous a donc amenés à mesurer le masquage dans le plan TF à l'aide de stimuli brefs (9,6 ms) et à support fréquentiel compact (« Equivalent Rectangular Bandwidth » = 600 Hz) : des atomes TF. Les seuils de détection d'un atome cible ont été mesurés en fonction de la distance temporelle et/ou fréquentielle le séparant d'un atome masque. Lorsque seule la séparation fréquentielle ou temporelle masque-cible est manipulée, les résultats sont comparables à ceux obtenus avec des signaux de durée et de bande passante équivalentes. Cependant, les résultats obtenus pour différentes combinaisons TF sont nouveaux et constituent une base cruciale pour l'incorporation de données psychoacoustiques de masquage aux représentations TF des sons (étape en cours de développement).

Les résultats des mesures psychoacoustiques seront présentés ainsi que les résultats préliminaires de l'algorithme de traitement du signal visant à éliminer l'information inutile, d'un point de vue perceptif, dans la représentation temps-fréquence d'un son.

64. Olivero Anaïk :

Analyse et Catégorisation de sons par multiplicateurs temps-fréquence.

Les transformations temps-fréquence sont fréquemment utilisées pour fournir des représentations efficaces des signaux. L'efficacité est jugée en termes de parcimonie, c'est-à-dire de capacité à concentrer l'information sur un nombre contrôlé de coefficients significatifs, et d'interprétabilité. Cette seconde propriété signifie que les positions des coefficients significatifs dans le plan temps-fréquence, et leurs valeurs, contiennent des informations pertinentes sur la nature du signal analysé. Il est par conséquent pertinent de s'attendre à ce que des signaux proches se caractérisent par des représentations temps-fréquence proches, au sens où leur « morphologie » est similaire.

Parmi les approches usuelles pour la classification de signaux, on peut notamment mentionner les approches basées sur les comparaisons de spectres, par exemple des mesures de « distance » entre spectres calculées à partir d'une certaine « norme » du rapport des spectres. Le recours aux spectres suppose une hypothèse implicite de stationnarité, qui n'est généralement pas satisfaite. Nous montrons ici que des mesures similaires basées sur les « masques » estimés de multiplicateurs temps-fréquence sont plus pertinentes, y compris dans des situations où l'hypothèse de stationnarité n'est pas loin d'être satisfaite.

65. Parseihian Gaetan :

Conception d'un système de guidage auditif spatialisé pour non-voyant.

La présente étude s'inscrit dans le contexte du projet ANR NAVIG qui a pour objectif d'augmenter l'autonomie des non-voyants et des malvoyants dans une action primordiale, identifiée comme particulièrement problématique : la navigation.

Le système d'aide mis au point comprendra un GPS, un système de vision stéréoscopique et un système de restitution basé sur les techniques de spatialisation sonore binaurales. L'objet de ce travail de thèse est de mettre au point un moteur de synthèse binaurale qui augmentera la réalité avec des informations auditives permettant de localiser des cibles et d'atteindre une destination en évitant les obstacles.

Dans un premier temps, la mise en place du dispositif de restitution nécessite un travail d'adaptation de la technique binaurale à une écoute avec des HRTF (head related transfert fonction) non-individualisées. Nous avons mis au point un jeu de cache-cache sonore permettant un apprentissage accéléré de la localisation de sons avec des HRTF non-individualisées. Ce jeu utilise les facultés d'adaptation de notre cerveau grâce à un recouplement entre les informations auditives et les informations proprioceptives. Nous présenterons les améliorations que ce jeu peut apporter sur la localisation binaurale et discuterons de la pérennité de cet apprentissage. Dans un deuxième temps, le travail consiste à étudier la précision de l'action auditivement guidée en fonction du type de son utilisé et à élaborer un ensemble de règles de transmission entre informations visuelles et sonores qui dépendront du type de guidage envisagé. Nous présenterons quelques exemples de règles de guidage et discuterons de la précision de l'action qu'elles permettent de faire.

66. Pascal Frédéric :

Enregistrement et transmission, vers une approche pour une meilleure compréhension des implants cochléaires.

L'enregistrement sonore est un instrument d'observation à plus d'un titre : il permet évidemment de fixer une trace d'un phénomène acoustique mais il permet de procéder à son examen détaillé. Les propriétés de l'enregistrement sont pour une large part communes à celle de la diffusion. Il s'agit dans ces deux cas de la définition d'une image analogique. Nous trouvons dans les deux cas l'office de deux instances de membrane, soit en vérité deux procédures de transduction analogique. L'étape de captation est une transduction d'une variation de pression qui se produit en général dans un corps fluide vers la modulation en tension d'un signal électrique. L'étape de diffusion est la transduction de la modulation du signal électrique en une variation locale de pression du corps fluide (le phénomène électroacoustique). Ces considérations de proximité peuvent nous permettre d'envisager la description du fonctionnement de l'implant cochléaire d'une façon qui permettrait de concevoir le renouvellement de l'enquête concernant les points de déficience attestés de la prothèse. Il a été montré que l'usage de la simple simulation électrique de l'implant ne peut rendre compte des compétences des implantés eux-mêmes, ni de leur réception du signal acoustique du fait de l'absence de l'oreille externe et moyenne dans la procédure de réception du signal acoustique par l'implant. Un autre point critique est que l'intégration dynamique de la neuroprothèse par le sujet bénéficiaire ne peut être ni prise en compte ni documentée par ce type de procédure expérimentale. Les autres procédures expérimentales qui sont utilisées de façon plus centrale excluent la compréhension de la confrontation des sujets implantés à leur environnement ordinaire. Je tâcherai de montrer comment un complément des procédures habituelles pour l'analyse des performances des sujets implantés peut être envisagée à partir d'autres ressources. À ce titre les méthodes d'observation des phénomènes sonores qui ont été développées pour la composition de la musique concrète et de la musique électroacoustique peuvent être utilisées pour comparer les différentes réceptions des signaux acoustiques de façon entièrement quantifiable. Elles sont en même temps, suivant la tradition de Helmholtz, le support pour la meilleure compréhension de nos capacités de discrimination.

67. Rebillat Marc :

Identification, modélisation et contrôle de Large Multi-Actuator Panels pour la création d'un contenu audio-visuel spatialisé.

Dans les applications audio-visuelles, les systèmes de rendu audio (haut-parleurs) et visuels (écrans) sont presque toujours découplés. Cela impose alors des contraintes, en particulier sur le placement des haut-parleurs, qui nuisent à la qualité spatiale du rendu sonore. Une solution pour s'affranchir de ce problème est d'utiliser un nouveau type de transducteurs, nommé Large Multi-actuator Panels (LaMAPs) ou la reconstitution du champ sonore est réalisée directement par l'écran.

Un LaMAP est constitué d'une plaque de grandes dimensions faite d'un matériau sandwich avec un cœur en nid d'abeille. La face avant de cette plaque est couverte d'une peinture permettant une projection visuelle 3D stéréoscopique. Sur sa face arrière sont collés des excitateurs électrodynamiques. Ces excitateurs peuvent être pilotés par Wave Field Synthesis (WFS) pour créer un rendu sonore spatialisé. Un prototype de 2.6 m par 2 m a été construit au LIMSI et la qualité du rendu audio-visuel obtenu a été validée perceptivement. Le rendu proposé par le prototype est très précis en azimut, mais des erreurs perceptives sur les distances ont été observées.

La WFS est une application du principe de Huygens-Fresnel et suppose des sources acoustiques ponctuelles et omnidirectionnelles. En pratique, ce n'est pas le cas, et des méthodes d'égalisations sont alors utilisées. Pour piloter le prototype, une égalisation naïve ne prenant pas en compte les spécificités des LaMAPs a été appliquée individuellement à chacun des excitateurs. Il est probable qu'une amélioration de la technique d'égalisation utilisée, en prenant en compte les spécificités des LaMAPs, améliore perceptivement le rendu audio-visuel. Un modèle théorique adapté est utilisé pour représenter le comportement vibro-acoustique de la plaque. Les paramètres mécaniques du matériau étudié sont identifiés par une méthode vibratoire non-destructive. Les spécificités du comportement vibro-acoustique des LaMAPs sont ainsi connues et une stratégie de contrôle adéquate peut alors être envisagée. L'apport perceptif de ce contrôle sera évalué par la suite.

68. Regnier Lise :

Le vibrato, un descripteur pertinent pour la détection et l'indexation de la voix chantée ?

Mon travail de thèse vise à caractériser la voix dans un flux audio afin d'apporter des éléments pertinents pour l'indexation de la musique populaire. Le travail mené jusqu' alors se base sur les caractéristiques de la voix chantée suivantes : la présence quasi systématique de vibrato sur les notes tenues, l'utilisation du portamento pour lier les notes distantes et le caractère harmonique de la voix. La première étape de ce travail consiste à détecter la présence et l'absence de chant en présence d'un accompagnement instrumental. On choisit de détecter le chant en identifiant les partiels produits par la voix. Dans notre approche, les partiels sont modélisés par une fréquence variant lentement dans le temps (portamento) plus une modulation de fréquence quasi périodique (vibrato). Les partiels pour lesquels les caractéristiques du portamento ou du vibrato correspondent à celles d'une voix chantée sont identifiés comme étant du chant. Par ailleurs, l'identification des partiels émis par la voix nous permet, par re-synthèse, d'isoler la voix de son accompagnement. Pour améliorer la qualité de la séparation, le caractère harmonique de la voix est pris en compte en détectant, non pas des partiels isolés, mais des groupes de partiels harmoniques ayant les propriétés de la voix chantée. Ces groupes sont formés à l'aide d'algorithmes de clustering qui s'appuient sur une distance entre partiels construits de telle sorte que la distance entre deux partiels est proche de zéro lorsque les caractéristiques de leur vibrato ou de leur portamento sont similaires et qu'ils sont en relation harmonique. Par ailleurs, des modèles de chanteur sont construits à partir des caractéristiques de leur vibrato afin de résoudre des tâches d'identification et de classification de chanteurs. Dans ce travail, les résultats trouvés à l'aide de ce descripteur, le vibrato, sont comparés à ceux obtenus en utilisant des descripteurs plus classiques tels que les descripteurs de timbre.

69. Roze David :

Simulation de la dynamique d'une poutre de Reissner par les séries de Volterra.

La recherche en robotique utilise depuis de nombreuses années les géométries différentielle et algébrique pour simuler les mouvements de corps rigides. En effet, certains groupes et algèbres de Lie sont particulièrement bien adaptés à la réécriture des équations de la mécanique.

Le premier objectif de notre travail est d'écrire l'équation d'équilibre d'une poutre de Reissner dans une algèbre de Lie. Ce modèle permet de prendre en compte les couplages qui interviennent dans la dynamique d'une corde (torsion, flexions, cisaillement transverse...) et permettra de simuler par exemple, l'action d'un archer sur une corde de violon. Une équation concise de ce modèle sera donc le résultat d'un travail de réécriture nécessaire à la simulation de la dynamique non-linéaire de cette poutre.

Cette simulation sera réalisée grâce aux séries de Volterra. Cet outil permet de simuler des nonlinéarités polynomiales (i.e. faibles), c'est la raison de l'important travail de réécriture réalisé dans la première partie. Les noyaux de Volterra du système obtenu dans un premier temps, seront simulés par identification à des sommes et des produits instantanés de sorties de filtre.

70. Sannicandro Valerio :

Space As Parameter – A Different Approach to “Écriture”,
From Acoustic to Music Notation.

According which criteria can we affirm that a composition element (space) has reached the status of a musical parameter inside a musical work and, generally speaking, in the musical writing? To answer this question it is imperative to identify firstly those paths and possibilities that both musical research and technology might be useful for the composer and those influences that they could mean on musical language.

Consequently, our enquiry will see how musical material organisation can be structured starting from space/spatiality (seen not as a metaphor but as a true acoustic entity), an approach that will allow us to understand as well the aesthetic reasons that have been the background for this research. Finally, this study will show in which way space, an essential element for the music of our time, can be considered from an “inner” point of view instead than in its merely topological or architectural nature.

This perspective will thus lead our research into the domain of room acoustics (in particular to the resonance modes) and extend our exploration into a field of sensorial perception of timbres.

71. Schönstein David :

Méthodes pour l’adaptation individuelle de l’HRTF pour le rendu
via la synthèse binaurale.

La synthèse binaurale est un procédé de spatialisation des sons qui est basé sur le filtrage des signaux par des filtres reproduisant les propriétés de la propagation acoustique entre une source située en un point donné de l’espace et les oreilles d’un auditeur. Ces filtres sont définis par les HRTF (pour Head Related Transfer Function) qui correspondent aux fonctions de transfert entre la source sonore et les oreilles de l’auditeur.

Une synthèse binaurale de qualité, c’est-à-dire offrant une spatialisation complète et réaliste de la sphère 3D autour de l’auditeur, nécessite l’utilisation des HRTF individuelles, c’est-à-dire qui ont été mesurées sur l’auditeur considéré afin de prendre en compte les spécificités de sa morphologie (notamment : tête, pavillon des oreilles, haut du torse). Or, les mesures de HRTF requièrent un appareillage relativement lourd et onéreux, et impliquent de plus une procédure fastidieuse pour le sujet (séance de mesure durant 1 à 2 heures avec la tête bloquée pour un sujet). Ces contraintes rendent difficile, voire impossible la mise en oeuvre de mesures de HRTF dans le cadre d’applications grand public.

Le sujet de cette thèse concerne l’étude et le développement de plusieurs modèles pour l’adaptation de l’HRTF utilisée dans le rendu pour un individu donné, sans utiliser de mesure acoustique. Les méthodes concernées par cette étude sont catégorisées en deux groupes : Sélection de la meilleure HRTF parmi une base de données existante via des tests d’écoute. Sélection/génération d’une HRTF optimale utilisant des données morphologiques acquises via des photos numériques.

Ce travail comprend des phases acoustiques, audionumériques, et psychoacoustiques, dans lesquelles il y a les analyses des bases de HRTF, la génération des HRTF via les modèles morphologique (utilisant la méthode d’élément frontière), et des tests perceptifs pour l’évaluation des HRTF.

72. Sciabica Jean-François :

Modélisation psychoacoustique du bruit perçu dans un habitacle automobile.

Le bruit perçu à l'intérieur d'un habitacle automobile peut s'apprécier comme un « objet sonore » à part entière, une matière sonore porteuse d'informations diverses (qualité perçue, sportivité, plaisir) et utile (« santé » du véhicule, puissance à l'accélération). Ce bruit provient de trois sources sonores bien identifiées : le bruit moteur, le bruit aérodynamique et le bruit de roulement. Il s'agit d'un signal riche en composantes harmoniques dont les fréquences et amplitudes varient fortement sur une montée de régime.

Afin de clarifier l'impact des harmoniques moteur sur la perception globale du bruit habitacle, nous travaillons à réduire les composantes « signal » au strict nécessaire pour décrire les attributs perceptifs principaux du bruit habitacle (niveau global, bourdonnement, richesse harmonique). Pour cela, nous appliquons au bruit perçu dans l'habitacle une représentation issues des modèles auditifs pour refléter le ressenti par le conducteur. Le calcul des attributs perceptifs classiques du bruit perçu dans l'habitacle (bourdonnement et rugosité) se voit également facilité par ces représentations auditives en intégrant les phénomènes de masquage entre les harmoniques moteur et le bruit aérodynamique et de roulement, première étape de la réduction de l'information sonore.

Par ailleurs, nous pensons décrire les changements de timbre du bruit moteur entendus au cours d'une montée de régime par ces représentations auditives. Ces changements de coloration sont subtils mais perceptibles et contribuent au dynamisme perçu d'une automobile. Là où une représentation temps-fréquence est inefficace pour expliquer ces phénomènes, ces représentations auditives rendent possible la localisation de ces changements de coloration au cours de la montée de régime ainsi que l'identification des harmoniques moteur contribuant à ce phénomène.

73. Seldran Fabien :

Simulation d'un implant EAS (Electric Acoustic Stimulation) par un vocodeur hybride.

L'implant EAS (Electric Acoustic Stimulation) est destiné à des malentendants qui ont des restes auditifs exploitables dans les basses fréquences et une perte profonde dans les hautes fréquences. L'EAS va stimuler les fréquences graves acoustiquement par une aide auditive ; tandis que les hautes fréquences sont stimulées à l'aide d'un implant cochléaire (IC), qui envoie un signal électrique directement sur le nerf auditif.

Pour la partie IC de l'EAS, le porte-électrode (PE) utilisé est plus court qu'un PE standard mais le nombre d'électrodes sur le PE est inchangé. Les électrodes sont donc plus proches les unes des autres, ce qui augmente les interactions par rapport à un PE standard. Pour le codage du son par l'IC, il existe différentes stratégies qui varient essentiellement par leur vitesse de stimulation et le nombre d'électrodes stimulées à chaque instant. Le but de notre étude est d'évaluer s'il existe une stratégie de codage qui serait mieux adaptée à l'EAS parmi les stratégies CJS, ACE et SPEAK, c'est-à-dire une stratégie qui limite les interactions inter-électrodes sans dégrader l'intelligibilité du patient.

Nous avons conçu un vocodeur hybride (acoustique + électrique simulé) qui nous a permis d'évaluer, chez une population de 15 normo-entendants, l'intelligibilité en fonction de la stratégie de codage implémentée. Pour cette simulation, l'audition résiduelle de l'implanté est reproduite par un filtrage passe-bas du signal jusqu'à 650Hz et le signal « électrique » est codé entre 650 et 8000 Hz. Nous avons également reproduit un chevauchement fréquentiel entre les 2 stimulations et dans ce cas, le signal « électrique » est encodé entre 500 et 8000 Hz. Les tests ont été réalisés dans le silence et dans le bruit type cocktail party avec des rapports signal / bruit de -6, 0 ou +6 dB.

Nos résultats montrent des performances significativement meilleures lorsqu'il n'y a pas de chevauchement fréquentiel entre stimulation acoustique et stimulation électrique. Pour le nombre de maxima, plus le nombre augmente et plus l'intelligibilité augmente. Enfin, les tests statistiques n'ont pas mis en évidence un effet significatif de la vitesse de stimulation. Afin de limiter les interactions entre les électrodes sans dégrader l'intelligibilité, nos résultats suggèrent d'utiliser une stratégie de codage sans chevauchement fréquentiel, sans extraction de pic et à cadence lente. Une étude complémentaire chez des sujets implantés permettrait de valider cette hypothèse.

74. Sirdey Adrien :

Perspectives offertes par les masques de Gabor dans la transposition des notes de musique.

Les masques de Gabor sont des outils permettant la transformation de l'image temps-fréquence d'un son en un autre, qui correspond à un son différent. Nous étudions ici l'applicabilité de cette propriété à l'amélioration perceptive du vocodeur de phase de notes de musique jouées au piano. Nous montrons tout d'abord que les masques de Gabor permettent une analyse des lacunes du vocodeur de phase, et nous tentons ensuite de définir des masques paramétriques plus simples, permettant d'améliorer le rendu de la transposition par vocodeur de phase.

75. Verron Charles :

Synthèse de sons d'environnement spatialisés.

Ce travail de thèse porte sur deux thématiques complémentaires : la synthèse des sons d'environnement et la reproduction de l'espace acoustique. Les sons d'environnement désignent l'ensemble des sons que nous entendons chaque jour, hormis les sons musicaux et la parole. Il peut s'agir d'un son de vagues, des crépitements d'un feu, d'un éclat de verre, du vent qui souffle... Afin de plonger l'auditeur dans un environnement virtuel convaincant il est nécessaire de synthétiser le timbre de ces sources sonores, mais aussi leurs caractéristiques spatiales (position, largeur apparente, directivité, doppler) et celles de l'environnement acoustique (réverbération, occlusion). C'est le double objectif de ce travail de thèse, dont les applications touchent la création et l'enrichissement de contenus sonores : la sonification des jeux vidéo, le design sonore, la réalité virtuelle et la musique.

76. Zango Yves :

Conception de signaux de référence pour calibrer l'évaluation subjective des codeurs de la parole et du son.

L'évolution des performances des systèmes de télécommunications s'est traduite par un accroissement de la bande des signaux transmis ou stockés. Ainsi, ces systèmes supportent actuellement des signaux à bande bien plus large que celle des signaux bande étroite. Cette extension de bande a eu pour effet une meilleure perception de la qualité sonore des utilisateurs. Aujourd'hui, les opérateurs doivent évaluer et contrôler en permanence la qualité audio de leurs services afin de rester compétitifs. De par la nature même du signal de parole et de sa perception, cette évaluation nécessite la mise en oeuvre de tests subjectifs à grande échelle. Cependant, pour assurer une fiabilité des résultats de ces évaluations l'utilisation de signaux de référence s'impose.

Actuellement, le système de référence employé pour les tests de signaux de qualité téléphonique est le MNRU (Modulated Noise Reference Unit) qui fut conçu pour modéliser les défauts introduits par les codeurs audio dans les années 80. Quant au système de références utilisé pour les tests des signaux de haute qualité, il consiste en un ensemble de versions limitées en bande du signal original. Depuis, l'essor des techniques de codage s'est traduit par de nouveaux codeurs audio qui génèrent d'autres défauts, de caractéristiques diverses et variées.

La conception d'un nouvel appareil prenant en compte ces défauts est donc indispensable. Une première thèse réalisée au sein du laboratoire TECH/SSTP de France Télécom a mis en exergue le fait que les défauts introduits par les nouveaux codeurs peuvent être projetés sur un espace perceptif à 4 dimensions. De cet espace perceptif, des études visant à établir une corrélation entre les défauts et les caractéristiques des codeurs ont conduit à une première conception de signaux d'ancrage, qui n'est actuellement pas aboutie puisque seules 2 dimensions apparaissent correctement identifiées. A court terme, notre travail consiste à (i) valider l'espace perceptif à 4 dimensions pour les locuteurs homme et femme, (ii) vérifier la véracité des deux premières dimensions et les signaux de référence associés et (iii) caractériser au mieux les autres dimensions et concevoir des signaux de référence les plus adéquats avec ces dimensions. Ce travail sera suivi d'une phase de validation du modèle de référence proposé.