ED SMAER

Sujet de thèses 2013

Laboratoire : **Institut Jean Le Rond d'Alembert, équipe LAM (Lutheries, acoustique, musique)**

Etablissement de rattachement : **UPMC**

Directeur de thèse et section CNU ou CNRS : **Olivier ADAM, PR section 61**

Codirection et section CNU et CNRS : **Boris DOVAL, MCF section 61**

Codirection et section CNU et CNRS : **Christophe d'ALESSANDRO, DR CNRS**

Titre de la thèse : **Synthèse et contrôle gestuel de voix chantée**

Collaborations dans le cadre de la thèse :

Rattachement à un programme :

Le sujet peut être publié sur le site web de l'ED SMAER : OUI ~~NON~~

Résumé du sujet :

Les systèmes de synthèse de voix chantée ne sont pas satisfaisant en terme de réalisme, l'auditeur détectant très rapidement le son produit par le système comme étant un son synthétique. Les avancées des recherches actuelles, autant sur la modélisation physique de la source glottique et sur les interactions entre la source glottique et le conduit vocal que sur le contrôle gestuel de la synthèse, permettent d'envisager une nouvelle approche des systèmes de synthèse.

L'approche envisagée consiste dans un premier temps à dresser un état de l'art des nombreuses variantes des modèles physiques de production de la source vocale ainsi que de la prise en compte de l'interaction source-filtre. L'objectif est ici d'établir un modèle physique de source et de conduit vocal orienté vers une bonne adéquation entre les caractéristiques dynamiques (notamment l'attaque et la tenue de sons droits) et acoustiques du son produit et celles des sons vocaux naturels.

Dans un deuxième temps, il s'agit de développer un système de synthèse de voix chantée contrôlé par le geste, et qui utilise à la fois ces modèles physiques associés à un modèle d'interactions source filtre, les connaissances acquises depuis quelques années sur les mécanismes laryngés, ainsi que les bases de données de chanteurs sur toute leur étendue. Les différentes étapes comprendront une implémentation temps réel de ces modèles physiques, une prise en compte des interactions source-filtre, une implémentation des données des phonétogrammes permettant la synthèse sur toute l'étendue, une étude des paramètres de contrôle de ces modèles permettant d'imiter les effets des gestes vocaux naturels, une étude des interfaces (tablette, gant, joystick, pad,...) et des gestes de contrôle.

Il s'agit donc d'un projet incluant une modélisation physique et acoustique du comportement des plis vocaux et du conduit vocal ainsi que de leur interaction, une implémentation temps réel de ces modèles, un développement de schémas permettant de contrôler les paramètres de ces modèles à partir d'interfaces de contrôle gestuel, et enfin des tests perceptifs permettant d'évaluer la qualité du système de synthèse.

Sujet développé (à présenter en 2 ou 3 pages maximum,
en précisant notamment le contexte, les objectifs, les résultats attendus)

Contexte :

Depuis plusieurs années, les recherches sur la voix chantée au LAM et au LIMSI se sont associées autour d'une part de l'étude de la source glottique (modèles de signaux, modèles physiques, mesures et estimation de paramètres, corrélations acoustiques), et d'autre part du contrôle gestuel de la synthèse notamment au travers de projets FEDER (2PIM, ORJO). Ceci a donné lieu au développement du système de synthèse de voix chantée Cantor Digitalis. Ces recherches ont montré d'une part l'intérêt du contrôle gestuel de la synthèse pour l'expressivité de la voix de synthèse, permettant l'utilisation de ce synthétiseur en contexte musical, et d'autre part que le contrôle de l'articulation (notamment des consonnes) reste un problème difficile.

Objectifs :

L'objectif est ici de concevoir un instrument de synthèse de voix chantée complet contrôlé par le geste en temps réel, et d'une qualité suffisante pour être utilisé en contexte musical.

L'une des questions qui se posent est de trouver un moyen de réduire la complexité du contrôle des nombreux paramètres des modèles de synthèse. Comment faire en sorte qu'un geste manuel rende compte de la dynamique propre de l'attaque du son ou de l'articulation d'une consonne ? Quels modèles intermédiaires faut-il mettre en jeu pour contrôler les paramètres de bas niveau ? L'idée est d'étudier dans quelle mesure les modèles physiques de source et de conduit vocal permettront de répondre à ces questions de façon plus efficace que ne le font les modèles de signaux. Par exemple les différentes phases de la production d'une plosive sourde (silence, explosion, friction, mouvement de formants) sont le résultat d'un geste articulatoire relativement simple (abaissement de la langue) qui peut se traduire facilement en un seul geste de contrôle de la géométrie du conduit, alors qu'un modèle de signal nécessite de simuler les différentes phases et d'en modéliser finement la synchronisation.

Une autre question qui se pose concerne le naturel de la synthèse. Quels sont les éléments qui font juger de la qualité « synthétique » des voix de synthèse ? Il semble qu'il ne s'agisse pas d'un facteur particulier mais d'une diversité d'éléments pouvant apparaître à des échelles de temps différentes et principalement liés au comportement dynamique de la voix : la trop grande régularité des sons droits, la dynamique des attaques, mais aussi les comportements globaux tels que les registres vocaux (tessitures et changements de timbre associés) ou les corrélations entre paramètres de haut niveau (corrélation entre intensité, fréquence fondamentale et autre paramètres de source par exemple).

Résultats :

Des résultats sont attendus en terme de compréhension de la production vocale, de modélisation de la source glottique, et des interactions entre la source glottique et le conduit vocal, ainsi que sur les possibilités de contrôler la synthèse par le geste. De nouvelles pistes de recherche découleront de la comparaison entre son de synthèse et son vocal. Parmi les applications envisagées figurent les applications musicales, où les questions de « lutherie » seront prépondérantes, en particulier les questions de réalisme et de qualité de la synthèse, de jouabilité et d'ergonomie, de répertoire. De ce point de vue, le système se comportera comme un « instrument vocal ». Des applications pédagogiques sont aussi envisagées, notamment auprès des chanteurs et professeurs de chant, où les notions importantes mais délicates associées à la production de la voix pourront être illustrées, grâce à la possibilité de « déconstruire » l'instrument en faisant écouter séparément les sons des différentes parties, ou l'influence de tel ou tel paramètre physique (abaissement du larynx, élargissement de la zone laryngée,...).