

ED SMAER

Sujet de thèse 2014

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

Etablissement de rattachement : **UPMC**

Directeur de thèse et section CNU ou CNRS : **Benoît FABRE, CNU 60**

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

Codirection et section CNU et CNRS : **Nathalie HENRICH, GIPSA-LAB, CNRS 34**

Titre de la thèse : **Suppléance vocale et voix augmentée par injection acoustique dans le conduit vocal**

Collaborations dans le cadre de la thèse :

Rattachement à un programme :

Le sujet peut être publié sur le site web de l'EDSMAE : **OUI**

Résumé du sujet :

Contexte : Comment mesurer les propriétés du système vocal en phonation ? Comment injecter dans la bouche une source acoustique qui se substitue à la source glottique naturelle et qui permette aux personnes privées de leur source vocale de s'exprimer ? L'état actuel des recherches sur la voix montre que les propriétés du conduit vocal sont assez bien identifiées mais qu'il est plus délicat de caractériser la source vocale. Par rapport aux différentes approches qui ont été mises en oeuvre, comme celle de Joe Wolfe à l'UNSW, nous développons depuis plusieurs années un dispositif original d'injection acoustique d'une source maîtrisée dans un conduit (Cédrik Erbsen, IJLRA et GIPSA-lab, soutenance en 2014), permettant non seulement de mesurer l'impédance du conduit vocal dans des conditions peu contraignantes mais aussi d'envisager une application de l'injection de source à la suppléance vocale.

Objectifs : Fort de cette expérience, les objectifs sont tout d'abord de développer un outil de suppléance vocal, qui permette d'injecter une source de débit glottique maîtrisée dans le conduit vocal pour permettre la parole ou le chant sans faire usage de ses plis vocaux. Ensuite il s'agit d'étudier à l'aide de cet outil les phénomènes d'interaction source-filtre du point de vue signal, et de proposer un modèle de synthèse de la source glottique qui intègre certains de ces aspects dans la modélisation du signal source injecté. Enfin, il s'agit d'étudier et de développer une solution de contrôle gestuel de la synthèse adaptée à l'usage courant.

Résultats : Ces travaux permettront une meilleure compréhension des interactions source-filtre, une meilleure modélisation de la source glottique pour des applications de synthèse de qualité vocale, et la mise au point d'un dispositif de suppléance vocale, de parole et de chant augmentés, contrôlé par le geste.

Sujet développé

Le contexte de la thèse est le suivant. Depuis plus d'un demi-siècle, la production vocale humaine est modélisée par la théorie source-filtre (Fant, 1960). Lors de la production de sons voisés, le mouvement vibratoire des plis vocaux génère une source de débit acoustique, qui charge le conduit vocal. Celui-ci se comporte comme un résonateur acoustique, caractérisé par sa fonction de transfert. L'onde acoustique ainsi filtrée est ensuite rayonnée par les lèvres vers le milieu extérieur. Une hypothèse très forte de cette théorie est l'absence d'interaction entre la source glottique et le filtre du conduit vocal. C'est cette hypothèse qui permet en particulier d'effectuer une séparation de la source et du filtre par prédiction linéaire, méthode utilisée notamment en analyse, en synthèse, en reconnaissance, en codage et en transmission de la parole. Des études récentes montrent néanmoins l'importance de prendre en compte ces interactions pour l'étude de la qualité vocale en parole et la modélisation d'expressions vocales diverses (parole, chant, cri, murmure). Cependant, l'étude de ces interactions est rendue délicate par le fait que l'instrument vocal fait partie intégrante de l'humain. Très rares sont les dispositifs ou méthodes expérimentales permettant d'estimer des paramètres de la source ou du filtre séparément, la plupart étant invasifs, et les protocoles associés souvent très lourds. De plus aucun dispositif actuel ne permet de mesurer l'onde de débit directement à la sortie des plis vocaux in vivo, la source vocale n'étant toujours estimée que de façon indirecte.

Les objectifs de cette thèse sont de :

- 1) développer un outil de suppléance vocal, qui permette d'injecter une source de débit glottique maîtrisée dans le conduit vocal pour permettre la parole ou le chant sans faire usage de ses plis vocaux
- 2) étudier à l'aide de cet outil les phénomènes d'interaction source-filtre du point de vue signal, et de proposer un modèle de synthèse de la source glottique qui intègre certains de ces aspects dans la modélisation du signal source injecté.

Pour réaliser le premier objectif, nous nous appuyerons sur des travaux de thèse récents, qui ont permis de maîtriser l'injection acoustique d'un signal synthétique dans une cavité. Nous développerons l'outil de suppléance vocal en adaptant une technologie existante sur le marché américain, l'UltraVoice. Nous améliorerons en particulier la qualité du signal injecté en sortie du dispositif, afin qu'il permette un rendu acoustique au plus proche de la parole humaine. Nous intégrerons également un contrôle chironimique des paramètres de synthèse du signal de source. Différentes modalités de contrôle seront testées (par tablette, par smartphone, par la captation de mouvements articulatoires comme l'ouverture-fermeture de la mandibule ou l'élévation et abaissement du larynx, par pression d'un ballon dans la main).

Le second objectif a des implications scientifiques plus fondamentales. L'idée est de comparer une production vocale réelle avec une production où la source glottique générée par le dispositif de suppléance vocale remplace la source glottique réelle. La source vocale réelle sera mesurée par électroglottographie, une technique non-invasive d'estimation du contact glottique. La source glottique équivalente à la source réelle sera estimée à partir d'un protocole d'analyse par la synthèse sur l'humain. Nous nous appuyerons dans un premier temps sur les modèles de source glottique existants, en optimisant les paramètres des modèles. Nous étudierons par la suite quels seraient les signaux de source qui permettent un rendu acoustique optimal pour la parole augmentée. La comparaison entre ces deux approches nous permettra de revisiter les modèles d'onde de débit glottique pour y intégrer certains aspects de l'interaction source-filtre.

Les résultats attendus sont les suivants :

- 1) une meilleure compréhension des interactions source-filtre
- 2) une meilleure modélisation de la source glottique pour des applications de synthèse de qualité vocale
- 3) la mise au point d'un dispositif de suppléance vocale, de parole et de chant augmentés, contrôlé par le geste

Parmi les retombées, outre un nouveau dispositif commercialisable concurrent des électrolarynx actuels, le dispositif développé a un intérêt pédagogique pour l'apprentissage des gestes articulatoires en voix parlée et chantée, en particulier pour les aspects de qualité vocale, d'ajustement phono-résonantiel et d'efficacité vocale. Il a aussi un intérêt musical puisque l'idée même d'injecter une source acoustique dans la bouche est née de la « TalkBox », interface utilisée par certains musiciens des années 70 pour donner une qualité « vocalique » à des sons de synthèse ou de guitare, qualité vocalique plus réaliste que les filtres wah-wah. Il peut ainsi être appliqué dans une optique de « karaoké augmenté », permettant à une personne d'articuler sur une source chantée donnée, ou même de chanter tout seul mais à plusieurs voix (en chantant et en actionnant le dispositif en même temps).