

Proposition de contrat doctoral

Titre :

Rythme de la parole et gestes manuels en synthèse performative

Résumé du sujet :

Le but de cette thèse est de développer un cadre théorique et des expérimentations quant à l'utilisation du geste manuel pour le contrôle prosodique via des interfaces humain-machine, en synthèse performative. La synthèse vocale performative est un nouveau paradigme de recherche en interaction humain-machine, dans lequel une voix de synthèse est jouée comme un instrument en temps-réel à l'aide des membres (mains, pieds). Le contrôle du rythme de parole par les mains est un problème qui implique des unités rythmiques, des points de contrôle rythmique, les centres perceptifs des syllabes et des gestes de tapotement (tapping), voire des partitions gestuelles inspirées des phonologies autosegmentales ou articulatoires. Les unités rythmiques varient en fonction de la phonologie de la langue étudiée, ici le français, l'anglais et le chinois mandarin. Les enjeux de la thèse portent donc sur la modélisation des schémas de perception-action impliqués dans le contrôle rythmique, la modélisation des unités temporelles, la réalisation et l'évaluation d'un système de contrôle du rythme. Les applications visées sont :

1. l'apprentissage du contrôle naturel des contours intonatifs à l'aide de la chironomie pour l'acquisition de langues étrangères (anglais, français, mandarin) ;
2. l'apprentissage du contrôle chironomique des contours d'intonation de la langue maternelle, pour la suppléance vocale (larynx artificiel).

Contexte :

La voix n'est pas un "instrument" de musique, au sens d'un artefact mis en vibration par les membres ou par le souffle. Les organes vocaux sont internes, en grande partie invisibles, et contrôlés de façon complexe par plusieurs ensembles musculaires (respiration, phonation, articulation). Le contrôle vocal est donc, par nature intéroceptif, alors qu'il est davantage kinesthésique et extéroceptif pour les instruments de musique.

L'avènement de la synthèse numérique permet pour la première fois le rendu d'un son indéniablement vocal par un dispositif instrumental externe, mis à distance de l'appareil vocal. Les "instruments vocaux" sont "manœuvrés" par les mains, les pieds, à l'aide de capteurs ou d'interfaces humain-machine. Cette mise à distance pose la question du contrôle vocal dans des termes tout à fait différents de ceux du contrôle d'un instrument acoustique ou de la voix elle-même. Les instruments vocaux permettent actuellement un contrôle musical de la phonation : intonation, séquençage rythmique, qualité de voix, pour la voix chantée. Le contrôle très précis de l'articulation et du rythme en parole est encore problématique. Le propos de cette thèse est de traiter la question du contrôle gestuel du rythme prosodique et du séquençage articulatoire.

Objectifs et résultats attendus :

Cette thèse s'inscrit dans la ligne de recherche sur les instruments vocaux. Un instrument vocal est un synthétiseur en vocal temps réel à contrôle gestuel. La synthèse est réalisée par un programme pour produire les échantillons. Le contrôle gestuel utilise des interfaces pour capter les gestes. Les mouvements des articulateurs étant très rapides, il est difficile de les contrôler de façon directe par les gestes manuels et des méthodologies basées sur la représentation phonologique du rythme prosodique doivent être mises en place.

Le rythme est réalisé par des gestes des membres, mains ou pieds, en place des gestes articulatoires qui correspondent aux syllabes. Les circuits de perception-action ne sont plus les mêmes, ni les vitesses des organes mis en mouvement. Le contrôle du rythme prosodique en synthèse performative est donc un problème qui implique la définition d'unités rythmiques, de points de contrôle rythmique, de centres perceptifs des syllabes, de gestes de tapotement (tapping), voire de partitions gestuelles inspirées des phonologies autosegmentales ou articulatoires.

Des points de contrôles rythmiques doivent enrichir le signal vocal pour permettre d'en manipuler le déroulement temporel. Ces points doivent avoir du sens du point de vue de la phonologie de la

langue jouée, et de sa phonotactique. La perception du flux syllabique, avec ses centres perceptifs, est donc impliquée. Les gestes de contrôle, par appuis ou tapotage, impliquent des processus moteurs, à la fois analogues et différents de ceux des articulateurs. Les unités rythmiques varient en fonction de la phonologie des langues étudiées, ici le français, l'anglais et le chinois mandarin. Les enjeux de la thèse portent donc sur la modélisation des schémas de perception-action impliqués dans le contrôle rythmique, la modélisation des unités temporelles, la réalisation et l'évaluation d'un système de contrôle du rythme.

Les résultats attendus sont à la fois théoriques et pratiques :

- L'expérimentation perceptive permettra de mettre en relation les différentes unités temporelles;
- les théories phonologiques sur l'organisation du geste phonatoire seront mises à l'épreuve avec un nouveau paradigme expérimental;
- un nouveau synthétiseur sera réalisé;
- un ensemble de méthodes pour le contrôle gestuel de la synthèse, de nouveaux gestes et des interfaces adaptées seront développés et testés dans les tâches applicative visées, soit l'apprentissage du contrôle naturel des contours intonatifs à l'aide de la chironomie pour l'acquisition de langues étrangères (anglais, français, mandarin) et l'apprentissage du contrôle chironomique des contours d'intonation de la langue maternelle, pour la suppléance vocale (larynx artificiel).

Méthodologie :

Les théories phonologiques et phonétiques de l'organisation temporelle des langues étudiées seront considérées dans le contexte du paradigme de la synthèse performative. L'étude des relations entre points de contrôle rythmique, centres perceptifs, gestes de tapotage et unités phonologiques implique la modélisation et l'expérimentation, avec des sujets réalisant des tâches de perception-action. La méthodologie relève ici de la psychologie et de la phonétique expérimentales : définition de corpus, mise en œuvre de protocoles de test, tests, analyses statistiques.

Un synthétiseur qui utilise les nouveaux paradigmes de contrôle rythmique sera développé. La méthodologie relève ici du traitement du signal audio et de la parole ainsi que de l'informatique, depuis la conception jusqu'à la programmation.

Ainsi un ensemble de méthodes pour le contrôle gestuel du rythme prosodique et du temps sera développé et testé dans les tâches applicatives visées. Ces méthodes comprennent à la fois les gestes et les interfaces de contrôle et relèvent de l'informatique dans le domaine des interfaces humain-machine.

Prérequis :

Ce sujet est à l'interface de la synthèse vocale et des interfaces humain-machine, de la prosodie, de la perception et de la performance musicale. Cela demande des connaissances générales en traitement du signal audio numérique et en informatique musicale ou en interface humain-machine. Une partie du travail portera sur le développement logiciel. Des connaissances sur la voix et la parole, en phonétique et phonologie, ainsi qu'en psychologie expérimentale ou sciences cognitives seront nécessaires.

Les candidatures avec une formation initiale en informatique et traitement du signal aussi bien que celles avec une formation initiale en linguistique, phonétique ou sciences cognitives seront considérées. La formation initiale sera éventuellement complétée dans les domaines qui seraient moins connus.

Encadrement :

Christophe d'Alessandro, DR CNRS, Responsable de l'équipe LAM
Institut Jean Le Rond d'Alembert, Sorbonne Université
christophe.dalessandro@sorbonne-universite.fr

Ce projet doctoral est dans le cadre du contrat ANR Gepeto, en collaboration avec le LPP (Sorbonne nouvelle), et le GIPSA-Lab, Université de Grenoble.

Début du contrat dès que possible (à partir d'octobre 2020)

Références :

- Delalez, S. et d'Alessandro, C. (2017). "Vokinesis: syllabic control points for performative singing synthesis", NIME'17 , pp. 198-203.
- X. Xiao, G. Locqueville, C. d'Alessandro, B. Doval, « T-Voks: Controlling Singing and Speaking Synthesis with the Theremin », Proceedings of the International Conference on New Interfaces for Musical Expression, NIME'19, June 3-6, 2019, Porto Alegre, Brazil, 110-115.
- Samuel Delalez, Christophe d'Alessandro « Adjusting the Frame: Biphasic Performative Control of Speech Rhythm », Proc. INTERSPEECH 2017, 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, August 18-25, 2017, DOI: 10.21437/Interspeech.2017, 864-868.
- Christophe d'Alessandro, Albert Rilliard, and Sylvain Le Beux « Chironomic stylization of intonation » J. Acoust. Soc. Am., 129(3), march 2011, 1594-1604
- Christophe d'Alessandro, Lionel Feugère, Sylvain Le Beux, Olivier Perrotin, and Albert Rilliard (2014) , « Drawing melodies : evaluation of chironomic singing synthesis » , J. Acoust. Soc. Am. 135 (6), 3601-3612.
- I. Chow, M. Belyk, V. Tran, and S. Brown. Syllable synchronisation and the P-center in cantonese. 49 :55–66, 2015.
- C. d'Alessandro, L. Feugère, S. Le Beux, and O. Perrotin. Drawing melodies : Evaluation of chironomic singing synthesis. J. Acoust. Soc. Am., 135(6) :3601–3612, March 2014.
- C. d'Alessandro, A. Rilliard, and S. Le Beux. Chironomic stylisation of intonation. J. Acoust. Soc. Am., 129(3) :1594–1604, March 2011.
- C. Fowler. "Perceptual centers" in speech production and perception. Perception & Psychophysics, 25 :375–388, 1979.
- P. Howell. Prediction of p-center location from the distribution of energy in the amplitude envelope. Perception and Psychophysics, 43 :90–93, 1988.
- P. F. MacNeilage. The frame/content theory of evolution of speech production. Behavioral and Brain Sciences, 21 :499–546, 1998.
- S.M. Marcus. Acoustic determinants of perceptual center (P-center). Perception and Psychophysics, 30 :247–256, 1981.
- J. Morton, S. Marcus, and C. Frankish. Perceptual centers (P-Centers). Psychological Review, 83(5) :405–408, 1976.
- B. Pompino-Marshall. On the psycho-acoustic nature of the p-center phenomenon. Journal of phonetics, 17 :175–192, 1989.
- K. Rapp-Holmgren. A study of syllable timing. STL-QPSR, 12(1) :014–019, 1971.
- B. H. Repp. Sensorimotor synchronization : A review of tapping literature. Psychon. Bull. Rev., 12(6) :969–992, 2005.
- B. H. Repp and Y. H. Su. Sensorimotor synchronisation : A review of recent research. Psychon. Bull. Rev., 20 :403–452, 2013.
- P. Wagner. The Rhythm of Language and Speech : Constraining Factors, Models, Metrics and Applications. Habilitation à diriger des recherches, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2008.