

EXPLORATION DU TRANSFERT DE MODALITÉ VOIX-GESTE PAR INTERFACES HUMAIN-MACHINE POUR LA SUPPLÉANCE VOCALE

Contexte

Ce stage s'intègre dans le cadre du projet ANR GEPETO [1] (GESTures and PEdagogogy of InTOnation), dont le but est d'étudier l'utilisation de gestes manuels par le biais d'interfaces humain-machine, pour la conception d'outils et méthodes permettant l'apprentissage du contrôle de l'intonation (mélodie) dans la parole. En particulier, ce stage se place dans le contexte de la suppléance vocale, dans le cas d'une altération de la vibration des plis vocaux ou d'une cordectomie partielle ou totale. Une solution pour palier à cette absence de source glottique est de capter le chuchotement du patient à l'aide d'un microphone en sortie du conduit vocal, et d'y réinjecter un signal de synthèse de la source vocale en temps-réel [2]. La voix ainsi reconstruite est alors diffusée en temps-réel sur un haut-parleur. Dans ce contexte, il est souhaitable de pouvoir contrôler l'intonation de la voix (en particulier les variations de hauteurs). Le but du projet GEPETO au GIPSA-lab est de proposer de nouvelles solutions de contrôle de l'intonation en temps-réel par le geste de la main, qui peut être captée par diverses interfaces (tablette graphique, accéléromètre, etc.), et d'étudier l'usage de tels systèmes dans des situations d'interactions orales.

Description

Travaux antérieurs : De précédentes études ont portées sur l'utilisation d'interfaces type tablette graphique pour le contrôle de l'intonation (en particulier la hauteur) de voix pré-enregistrées, dans le cadre de la voix chantée et parlée [3, 4, 5]. Il a été montré qu'une telle interface permet un contrôle de l'intonation aussi précis qu'avec la voix naturelle dans une tâche d'imitation, mais repose en grande partie sur la modalité visuelle inhérente à la tablette graphique [6]. Par ailleurs, celle-ci présente un certain encombrement dans le cadre de l'application visée ici, qui nécessite une interface facilement transportable et ergonomique. Une autre étude concernant l'utilisation d'un accéléromètre pour le contrôle d'un électrolarynx (source artificielle injectée dans le conduit vocal) a également été menée [7], mais la seule solution actuellement commercialisée, à notre connaissance, consiste en un simple capteur de pression [8]. Un précédent stage [9] a également permis de tester d'autres interfaces comme un Leap Motion [10] ou une tablette Sensel Morph [11], ainsi que différentes stratégies de contrôle associées à ces interfaces, pour le contrôle de l'intonation.

Objectifs : L'objectif ici est de comparer expérimentalement ces interfaces selon différents critères, notamment la précision de contrôle offerte par l'interface, ainsi que son ergonomie et sa facilité de prise en main. Pour cela, l'utilisation de ces interfaces sera dans un premier temps étudiée dans le cadre du contrôle de voix pré-enregistrées, par l'utilisation d'un vocodeur [12] permettant de transformer la hauteur de la voix en temps-réel dans l'environnement Max/MSP [13]. L'utilisateur pourra ainsi se concentrer sur la tâche de contrôle, indépendamment de l'articulation. Dans un second temps, ces interfaces seront utilisées conjointement avec un système de transformation temps-réel de voix chuchotée en voix parlée développé au laboratoire. Cela permettra d'étudier l'usage d'un tel système dans une application de suppléance vocale, et en particulier la coordination entre le contrôle manuel de l'intonation et le contrôle naturel de l'articulation. Pour cela, différentes tâches pourront être envisagées, comme l'imitation d'une phrase cible, ou bien la production d'un caractère intonatif type (e.g. affirmatif ou interrogatif). Une des principales questions qui se pose dans cette étude concerne la capacité des sujets à transférer leurs connaissances implicites des contours intonatifs, acquises naturellement lors de l'apprentissage du langage, vers cette autre modalité qu'est le contrôle gestuel. On cherchera également à évaluer la charge cognitive induite par les différentes tâches effectuées et interfaces utilisées. Nous utiliserons différents indicateurs physiologiques comme la variabilité du rythme cardiaque, l'activité électrodermale, les contractions musculaires du bras, les mouvements respiratoires thoraciques et abdominaux.

Selon le temps disponible, la question de l'apport d'un retour visuel (par exemple la visualisation des courbes produites en temps-réel) [6] sur la capacité de contrôle pourra également être étudiée.

Tâches

Les tâches à réaliser durant ce stage sont les suivantes :

- Se familiariser avec les différents systèmes de contrôle et synthèse/transformation de la voix utilisés dans le projet.
- Définir le protocole expérimental de l'étude : choix des interfaces, définitions des tâches cibles, données à récolter (contours intonatifs produits, données physiologiques, questionnaires, ...)
- Recruter des sujets et faire passer les tests
- Évaluer les résultats par l'analyse des données récoltées

Compétences requises

- Design d'expérience et analyse de données (statistiques)
- Connaissance de Python, Matlab, ou R nécessaire pour l'analyse des résultats
- Connaissances en traitement du signal souhaitable pour la compréhension du système et l'analyse des données
- Intérêt pour les interfaces humain-machine
- Notions de base en linguistique et en acoustique souhaitable
- Connaissance de l'environnement Max/MSP souhaitable pour le design de l'expérience et la compréhension des systèmes utilisés (utilisation des interfaces de contrôle et récolte des données)
- Grande motivation pour la méthodologie et l'expérimentation

Indemnité

L'indemnisation de stage est fixée par décret ministériel (environ 570 euros / mois).

Encadrement

Luc Ardaillon

Nathalie Henrich Bernardoni

Olivier Perrotin

+ 33 4 76 57 45 36

+ 33 4 76 57 45 36

luc.ardaillon@grenoble-inp.fr

nathalie.henrich@grenoble-inp.fr

olivier.perrotin@grenoble-inp.fr

Références

- [1] Projet ANR GEPETO. <https://anr.fr/Project-ANR-19-CE28-0018>.
- [2] Olivier Perrotin and Ian V McLoughlin. Glottal flow synthesis for whisper-to-speech conversion. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28 :889–900, 2020.
- [3] Christophe d'Alessandro, Albert Rilliard, and Sylvain Le Beux. Chironomic stylization of intonation. *The Journal of the Acoustical Society of America*, 129(3) :1594–1604, 2011.
- [4] Christophe d'Alessandro, Lionel Feugère, Sylvain Le Beux, Olivier Perrotin, and Albert Rilliard. Drawing melodies : Evaluation of chironomic singing synthesis. *The Journal of the Acoustical Society of America*, 135(6) :3601–3612, 2014.
- [5] Olivier Perrotin. *Chanter avec les mains : interfaces chironomiques pour les instruments de musique numériques*. PhD thesis, Université Paris Sud-Paris XI, 2015.
- [6] Olivier Perrotin and Christophe d'Alessandro. Seeing, listening, drawing : interferences between sensorimotor modalities in the use of a tablet musical interface. *ACM Transactions on Applied Perception (TAP)*, 14(2) :1–19, 2016.
- [7] Kenji Matsui, Kenta Kimura, Yoshihisa Nakatoh, and Yumiko O Kato. Development of electrolarynx with hands-free prosody control. In *Eighth ISCA Workshop on Speech Synthesis*, 2013.
- [8] Trutone Emote. <https://www.atosmedical.fr/product/trutone-emote/>.
- [9] Emilien Gros. Programmation d'interfaces sur maxmsp pour le contrôle gestuel de la voix synthétisée. *Rapport de stage - IUT Annecy*, 2021.
- [10] Leap Motion. <https://www.ultraLeap.com/product/leap-motion-controller/>.
- [11] Sensel Morph. <https://morph.sensel.com/>.
- [12] Masanori Morise, Fumiya Yokomori, and Kenji Ozawa. World : a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE TRANSACTIONS on Information and Systems*, 99(7) :1877–1884, 2016.
- [13] Max/MSP. <https://cycling74.com/products/max>.