

Titre de la thèse :

Identification et commande de la boucle sensorimotrice de production de la voix parlée

Laboratoire d'accueil :Institut FEMTO-ST, Dept. Automatique (AS2M) (<https://www.femto-st.fr>)**Spécialité du doctorat préparé :**

Automatique (application aux neurosciences)

Mots-clefs:

Automatique ; identification de système ; traitement du signal; instrumentation ; voix humaine ; neurosciences; temps-réel ; temps déterministe; FPGA

Descriptif détaillé de la thèse

Ce projet de thèse vise à développer de nouvelles techniques d'automatique (dispositifs expérimentaux et algorithmes) avec pour objectif d'accroître notre compréhension d'une question neuroscientifique : comment l'homme contrôle le son de sa voix?

D'un point de vue neuroscientifique, la parole chez l'homme est rendue possible par une boucle de rétroaction sensorimotrice (« vocal feedback ») dans laquelle les locuteurs parlent, entendent ce qu'ils viennent de prononcer, et se corrigent si nécessaire. Ce système est très rapide et en grande partie automatique : par exemple, quand on s'entend parler avec un délai de quelques centaines de millisecondes, on ne peut s'empêcher de bégayer (Stuart et al., 2002) ; quand on modifie la voix du locuteur dans son casque de façon à en augmenter la hauteur (pitch), celui-ci ajuste automatiquement le pitch de sa voix vers le bas pour compenser, avec un délai d'environ 300 ms (Burnett, Senner & Larson, 1997).

Ce phénomène fait l'objet de très nombreux travaux en neurosciences (y compris dans notre équipe : voir Aucouturier et al. 2016 ; Goupil et al. 2021). Ces travaux utilisent principalement la méthodologie « altered feedback » qui vise à introduire des perturbations de la voix perçue (ex. plus haute ou plus basse qu'elle n'a été produite) en quasi-temps-réel et en continu grâce à des algorithmes de traitement du signal, et d'en étudier l'effet sur la voix produite. Quoique ce système repose essentiellement sur une architecture boucle-fermée, il a principalement été étudié en boucle-ouverte (i.e. avec des manipulations de hauteur qui ne dépendent pas de la sortie observée du système), et la communauté semble être complètement passée à côté de l'aspect « automatique » du problème : un seul article, marginal mais par une des équipes leader de la discipline (Guenther lab, Boston University), a proposé de faire de l'identification de paramètres d'un contrôleur PID, mais l'a fait en boucle ouverte (i.e. avec des données déjà collectées) et pour le comportement moyen d'un groupe (Kearney et al., 2022). Or, dans la pratique, la réponse individuelle de participants est très variable, à la fois entre participants et dans le temps au regard d'un même participant. Comprendre cette variabilité peut donner lieu à de grandes avancées théoriques (ex. de question : les bons chanteurs sont-ils de bons parleurs ?) et applicatives (ex. les paramètres du système peuvent-ils être des biomarqueurs pour détecter ex. un Parkinson précoce ; Liu et al. 2012)

Le travail de thèse proposé consiste à développer et utiliser expérimentalement un nouveau dispositif de feedback vocal temps-réel (développé sous LabVIEW sur cible CompactRIO/FPGA), permettant d'exécuter des algorithmes en temps déterministe (cad sans jitter) aussi bien en «hard» qu'en «soft», afin d'implémenter la modélisation, l'identification et la commande du système au niveau de chaque participant individuellement. L'identification fournira des paramètres qui seront testés comme possibles biomarqueurs de différentes caractéristiques vocales des participants (ex. parole en langue maternelle vs en langue étrangère, en voix parlée vs chantée), et pourront également donner lieu à une application clinique dans le cadre de la maladie de

Parkinson. La commande, cad la construction en temps-réel d'un signal de traitement de la voix adapté à la production de façon à en contrôler les caractéristiques, permettra de créer un système cyber-physique surprenant (= faire chanter une personne à son insu, alors qu'elle croit maintenir une voix stable au casque), et ouvrir de possibles applications à des interventions thérapeutiques (ex. pour le bégaiement ou les troubles du stress post-traumatique).

La personne recrutée rejoindra le Département « Automatique et Systèmes Micro-Mécatroniques » de l'institut FEMTO-ST à Besançon, et sera intégrée au groupe Neuro (<https://neuro-team-femto.github.io>), un groupe de recherche développant de nouvelles approches inspirées de l'automatique et des systèmes dynamiques pour les neurosciences.

Références bibliographiques

Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, 113(4), 948-953.

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. *Journal of Voice*, 11(2), 202-211.

Goupil, L., Johansson, P., Hall, L., & Aucouturier, J. J. (2021). Vocal signals only impact speakers' own emotions when they are self-attributed. *Consciousness and Cognition*, 88, 103072.

Kearney, E., Nieto-Castañón, A., Falsini, R., Daliri, A., Heller Murray, E. S., Smith, D. J., & Guenther, F. H. (2022). Quantitatively characterizing reflexive responses to pitch perturbations. *Frontiers in Human Neuroscience*, 16, 929687.

Liu, H., Wang, E. Q., Metman, L. V., & Larson, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PloS one*, 7(3), e33629.

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, 111(5), 2237-2241.

Profil demandé

Ce projet de thèse sera adapté à un.e étudiant.e titulaire d'un M2 recherche et/ou d'un diplôme d'ingénieur.e avec un bon background en automatique (identification, commande boucle fermée, instrumentation), et/ou en traitement du signal (systèmes linéaires, signal sonore, temps-réel).

Les compétences techniques attendues sont la programmation Python (ex. analyse de données en pandas/seaborn) et LabVIEW (instrumentation, temps-réel). Une connaissance de l'environnement cRIO ou d'autres systèmes embarqués pourra également être mise à profit. Enfin, une expérience ou un intérêt pour le logiciel libre, le partage et la documentation de code sous git/github seront également appréciés. Les profils ne réunissant pas tous ces éléments sont évidemment les bienvenus, du moment que les candidat.e.s peuvent démontrer leur appétit d'apprendre.

Le projet étant à vocation expérimentale (développement de dispositif, expérimentation sur la personne humaine), le ou la candidate idéale aura une appétence pour l'expérimentation et le développement méthodologique, démontrée par exemple par une précédente expérience en stage de master. Enfin, l'application visée étant neuroscientifique, un intérêt et/ou une formation complémentaire en sciences cognitives

ou neurosciences sera un avantage apprécié – quoique non exigé.

Financement : MESRI Etablissement

Dossier à envoyer avant le **16 Mai 2025** (candidatures analysées au fil de l'eau même avant cette date).

Entretiens prévus du 19 au 28 Mai 2025.

Début du contrat : 1er Octobre 2025

Salaire mensuel brut : 2200€ (à partir du 1er janvier 2026 : 2300€ brut)

Direction de la thèse

AUCOUTURIER Jean-Julien, Directeur de recherche CNRS, jj.aucouturier@proton.me
Neuro group (<https://neuro-team-femto.github.io>), Institut FEMTO-ST, Besançon

Encadrement de la thèse : co-directeur(s) et co-encadrant(s)

Co-encadrant : Patrick NECTOUX, Ingénieur de recherche CNRS

Neuro group (<https://neuro-team-femto.github.io>), Institut FEMTO-ST, Besançon

Applicants are invited to submit their application to the PhD supervisors.

Application must contain the following documents:

- CV
- lettre de motivation (apportant des éléments factuels en réponse aux demandes du paragraphe "profil demandé", voir ci-dessus)
- au moins 1 lettre de recommandation (ex. responsable de stage de M2)
- Relevé de notes et classement de M1, relevé de notes et classement (si disponible) de M2

Thesis title :

Identification and control of the sensorimotor loop of spoken voice production

Host Laboratory :

Dept. Of Automation & Robotics, FEMTO-ST Institute, Besançon

Speciality :

Control engineering (with application to neurosciences)

Keywords:

Control engineering; system identification ; audio signal processing; human voice ; neuroscience ; real-time; deterministic time ; FPGA.

Job description

This thesis proposal aims to develop new control engineering techniques (experimental apparatus and algorithms) to advance our understanding of a neuroscientific question: how humans control the sound of their voice.

From a neuroscientific point of view, human speech is made possible by a sensorimotor feedback loop ('vocal feedback') in which speakers speak, hear what they have just said and correct themselves if necessary. This system is fast and largely automatic: for example, when we hear ourselves speak with a delay of a few hundred milliseconds, we cannot help but stutter (Stuart et al., 2002); when we modify a speaker's voice in their headphones so as to increase its pitch, they automatically adjust the pitch of their voice downwards to compensate, with a delay of around 300 ms (Burnett, Senner & Larson, 1997).

This phenomenon is the subject of a great deal of work in neuroscience (including in our team: see Aucouturier et al. 2016; Goupil et al. 2021). This work mainly uses the 'altered feedback' methodology, which introduces perturbations to the perceived voice (e.g. higher or lower than it was produced) in 'near-real time' using signal processing manipulations, and studies the effect on the voice produced. Although this system is essentially a closed-loop architecture, it has mainly been studied in open-loop (i.e. with pitch manipulations that do not depend on the observed output of the system), and the community seems to have completely missed the 'control engineering' aspect of the problem: only one paper, marginal though by one of the leading teams in the discipline (Guenther lab, Boston University), proposed to do parameter identification of a PID controller, but did so in open-loop (i.e. with data already collected) and for the average behaviour of a group (Kearney et al, 2022). In practice, however, the individual response of participants is highly variable, both across participants and across time within the same participant. Understanding this variability could give rise to major theoretical advances (e.g. the question: are good singers also good talkers?) and applications (e.g. could system parameters be biomarkers for detecting the behaviour of a group?

The proposed project involves developing a new real-time voice feedback device (developed under LabVIEW on a CompactRIO/FPGA target), allowing the implementation of signal processing algorithms in deterministic time (i.e., without jitter), both in hardware and software ; and to use it experimentally in order to model, identify and control the sensorimotor system at the level of each participant individually. System identification will provide parameters that will be tested as possible biomarkers of different vocal characteristics of the participants (e.g. native vs. foreign language speech, spoken vs.

sung voice), and may also give rise to a clinical application in the context of Parkinson's disease. Control, i.e. constructing a voice manipulation command signal adaptively to the real-time produced voice in order to control its characteristics, will make it possible to create a striking cyber-physical system (= making a person sing without their knowledge, when they believe they are maintaining a stable voice on headphones), and open up possible applications for therapeutic interventions (e.g. for stuttering, or traumatic stress disorders).

The successful applicant will be based in the Department of Automation and Robotics (AS2M), FEMTO-ST Institute in Besançon, and join the FEMTO Neuro group, a small group of researchers exploring control-engineering and dynamical-system approaches in neuroscience. (<https://neuro-team-femto.github.io>).

Bibliography

Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, 113(4), 948-953. https://neuro-team-femto.github.io/articles/2016/Aucouturier_PNAS_2016.pdf

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. *Journal of Voice*, 11(2), 202-211.

Goupil, L., Johansson, P., Hall, L., & Aucouturier, J. J. (2021). Vocal signals only impact speakers' own emotions when they are self-attributed. *Consciousness and Cognition*, 88, 103072. https://neuro-team-femto.github.io/articles/2021/Goupil_Consciousness_Cognition_2021.pdf

Kearney, E., Nieto-Castañón, A., Falsini, R., Daliri, A., Heller Murray, E. S., Smith, D. J., & Guenther, F. H. (2022). Quantitatively characterizing reflexive responses to pitch perturbations. *Frontiers in Human Neuroscience*, 16, 929687.

Liu, H., Wang, E. Q., Metman, L. V., & Larson, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PloS one*, 7(3), e33629.

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, 111(5), 2237-2241.

Applicant profile

This thesis project will be suitable for a student holding a Research M2 and/or an equivalent engineering degree with a good background in automation (identification, closed-loop control, instrumentation) and/or signal processing (linear systems, sound signals, real-time).

Expected technical skills include Python (e.g. data analysis in pandas/seaborn) and LabVIEW programming (instrumentation, real-time). Knowledge of the cRIO environment or other embedded systems (arduino, etc.) could also be useful. Finally, experience with open source software, and sharing and documenting code under git/github will also be appreciated. Profiles that do not meet all these criteria are of course welcome, as long as candidates can demonstrate an appetite for learning.

As the project is experimental in nature (device development, experimentation on humans), the ideal candidate will have an appetite for experimentation and methodological development, demonstrated for example by previous experience in a

Masters internship. Finally, as the intended application is neuroscientific, an interest in and/or additional training in cognitive science or neuroscience would be an advantage - although not a requirement.

Financement : MESRI Etablissement

Application to be sent before 16 May 2025 (applications analysed on an ongoing basis even before this date).

Interviews scheduled from 19 to 28 May 2025.

Start of contract: 1 October 2025

Gross monthly salary: €2200 (from 1 January 2026: €2300 gross)

Thesis Supervisor :

AUCOUTURIER Jean-Julien, Directeur de recherche CNRS, jj.aucouturier@proton.me
Neuro group (<https://neuro-team-femto.github.io>), Institut FEMTO-ST, Besançon

Thesis Co-supervisor :

Co-encadrant : Patrick NECTOUX, Ingénieur de recherche CNRS
Neuro group (<https://neuro-team-femto.github.io>), Institut FEMTO-ST, Besançon

Applicants are invited to submit their application to the PhD supervisors.

Application must contain the following documents:

- CV
- Cover letter
- At least 1 reference letter (ex. previous supervisor(s) from research internship)
- Masters (M1 and M2) grade transcript and ranking, as available