



HAL
open science

Optimalité dans la production de parole: quel rôle sur la formation de trajectoires ?

Tsiky Rakotomalala, Pierre Baraduc, Pascal Perrier

► To cite this version:

Tsiky Rakotomalala, Pierre Baraduc, Pascal Perrier. Optimalité dans la production de parole: quel rôle sur la formation de trajectoires ?. JEP 2022 - 34e Journées d'Études sur la Parole, Jun 2022, Noirmoutier, France. hal-03870915

HAL Id: hal-03870915

<https://hal.science/hal-03870915v1>

Submitted on 24 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimalité dans la production de parole: quel rôle sur la formation de trajectoires ?

Tsiky Rakotomalala¹ Pierre Baraduc¹ Pascal Perrier¹

(1) Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

ny-tsiky.rakotomalala@grenoble-inp.fr, pierre.baraduc@grenoble-inp.fr,
pascal.perrier@grenoble-inp.fr

RÉSUMÉ

Le contrôle optimal, une théorie largement répandue en contrôle moteur, est encore rarement employé pour éclairer le contrôle neuromusculaire de la production de parole, probablement en raison de la complexité biomécanique du système orofacial. Dans ce papier, nous souhaitons confronter cette théorie avec l'hypothèse que l'un de nous a beaucoup étudiée, selon laquelle les trajectoires seraient majoritairement influencées par les contraintes biomécaniques des articulateurs de la parole. Pour cela, nous analysons la production de voyelles depuis une position de repos avec deux modèles de contrôle : le modèle GEPPETO pour lequel les trajectoires résultent de la spécification d'une posture finale et des caractéristiques biomécaniques du système de production de la parole, et un modèle de contrôle optimal où la formation de trajectoires dépend non seulement des contraintes biomécaniques, mais aussi d'un choix optimal des activations musculaires. Nos simulations permettent une comparaison directe des résultats dans les domaines acoustique, cinématique et moteur.

ABSTRACT

TRAJECTORY FORMATION IN SPEECH PRODUCTION : DOES OPTIMALITY MATTER ?

Optimal control theory is one of the leading theories in motor control, but it has rarely been used to study the control of articulatory models of speech production, probably due to the complexity of the biomechanics of the orofacial system. In this paper, we aim to confront this theory with the hypothesis extensively tested by one of us, that biomechanics have a decisive influence on the shape of trajectories. To this end, we compare vowel production from two different models : the GEPPETO model, in which actual trajectories result from the specification of sensory targets and the biomechanical characteristics of the speech production system, and an optimal control model in which trajectories are determined by the selection of optimal motor command patterns, on top of biomechanical constraints. A direct comparison of the results is presented in the acoustic domain, in the kinematic domain and in the motor domain.

MOTS-CLÉS : contrôle moteur de la parole, modèles internes, contrôle optimal, production de parole, formation de trajectoires.

KEYWORDS: speech motor control, internal models, optimal controller, speech production, trajectory formation.

1 Introduction

Le modèle GEPPETO (Patri *et al.*, 2015) est le seul modèle de contrôle moteur de la parole qui inclut un modèle du conduit vocal non pas géométrique, mais physique avec des caractéristiques biomécaniques (masse, élasticité, mécanismes de génération de force) et des aspects neurophysiologiques, avec la mise en place d'une rétroaction à court délai du type réflexe d'étirement musculaire. Les travaux sur ce modèle ont montré que cette biomécanique joue un rôle important dans la production de la parole avec en particulier un rôle majeur dans la formation des trajectoires (Perrier *et al.*, 2003). Toutefois, ce modèle présente une limite importante : alors que les retours sensoriels sont bien pris en compte au niveau du système nerveux central dans la planification (Patri *et al.*, 2019), la phase d'exécution s'exécute en mode purement "feedforward", sans aucune considération pour les retours sensoriels de haut niveau. Or il a été montré que les réponses à des perturbations intervenant pendant la production de la parole sont traitées pendant l'exécution du mouvement (par ex. Cai *et al.* 2011).

La littérature en contrôle moteur de la parole présente différentes méthodologies pour la prise en compte des retours sensoriels pendant la phase d'exécution des mouvements, en particulier :

- le modèle DIVA (Directions Into Velocities of Articulators, Guenther *et al.* 2006) utilise un système de contrôle hybride : le contrôle est essentiellement "feedforward" mais le modèle peut malgré tout s'adapter en cas de perturbations grâce à un système de contrôle "feedback" à court retard qui fonctionne en parallèle ;
- le modèle FACTS (Feedback-Aware Control of Tasks in Speech, Parrell *et al.* 2019) reprend le modèle géométrique du conduit vocal (Mermelstein, 1973) utilisé dans le modèle Task Dynamics (Saltzman, 1986), et l'associe avec le système de contrôle par retour d'état proposé par Houde & Nagarajan (2011). Les auteurs supposent qu'une estimation suffisamment précise de l'état est disponible pour pouvoir prédire le comportement futur de l'effecteur (ici les articulateurs du conduit vocal), et que le contrôleur dirige l'état de l'effecteur vers un état cible. Dans ce modèle, l'estimation est effectuée grâce à un filtre de Kalman, qui compare les retours sensoriels venant de l'effecteur à une prédiction interne de ces retours sensoriels, générée à partir d'une copie des commandes motrices (copie d'efférence). Ceci permet au système d'être robuste aux retards et au bruit dans la boucle "feedback".

Le modèle FACTS a montré que le contrôle par retour d'état produit des comportements articulatoires stables, ainsi que des réponses aux perturbations externes, spécifiques aux tâches, conformes aux observations expérimentales sur des locuteurs. Par ailleurs, dans les recherches sur le contrôle moteur du bras et de la main, les modèles exploitant une prédiction des retours sensoriels, de type SFC ou contrôle feedback optimal, ont déjà suscité un vif intérêt par leur capacité à expliquer de nombreuses observations comportementales à plusieurs niveaux d'analyse : cinématique (trajectoires de l'effecteur), mécanique (forces d'interaction), ou neurophysiologiques (électromyographie, électrophysiologie corticale) (Todorov & Jordan, 2002; Guigon *et al.*, 2007). Malgré ce cadre théorique puissant, le modèle FACTS ignore, et c'est selon nous une faiblesse importante, le rôle que la biomécanique joue dans la mise en forme des signaux articulatoires, et donc acoustiques, puisqu'il contrôle un modèle purement géométrique du conduit vocal.

Ainsi, l'objectif de cette étude est double. D'une part, nous voulons ajouter une prise en compte des retours sensoriels dans le modèle GEPPETO pendant la phase d'exécution motrice pour pouvoir gérer l'apparition de perturbations en cours de mouvement, tout en évaluant l'impact de l'imprécision des signaux sensoriels et moteurs dans la variabilité d'exécution. D'autre part nous souhaitons analyser l'intérêt d'introduire une notion d'optimalité dans le choix des commandes motrices, tout en prenant en compte les aspects biomécaniques. Nous avons donc implémenté un modèle de contrôle optimal dont l'effecteur est le modèle biomécanique de la langue utilisé dans GEPPETO, et l'avons utilisé

pour générer le mouvement de la langue depuis la position de repos (proche de la voyelle neutre /ə/) vers une voyelle, /i/, /a/, /ɔ/ ou /e/. Nous avons aussi généré ces transitions avec GEPPETO en ne prenant en compte que les valeurs des commandes optimales atteintes aux cibles vocaliques, et en faisant varier ces commandes à vitesse constante de leurs valeurs au repos vers les valeurs cibles. Les résultats des deux modèles ont été comparés, dans les domaines acoustique, cinématique et moteur.

2 Méthode

2.1 Le modèle GEPPETO

GEPPETO (acronyme de "GESTures shaped by the Physics and by a PERceptually-oriented Target Optimization") est un modèle qui permet de contrôler un modèle biomécanique bi-dimensionnel de la langue intégré dans le conduit vocal, dans le but de produire de la parole. Une séquence de parole y est considérée comme une séquence de phonèmes à laquelle on associe une séquence de cibles perceptives sous la forme des 3 premiers formants. Ces cibles acoustiques sont ensuite transformées en commandes posturales des articulateurs via une fonction d'association (résultant de l'apprentissage). La biomécanique de la langue est modélisée par la méthode des éléments finis (EF). Sept muscles responsables des principaux déplacements de la langue dans le plan sagittal sont modélisés (figure 1).

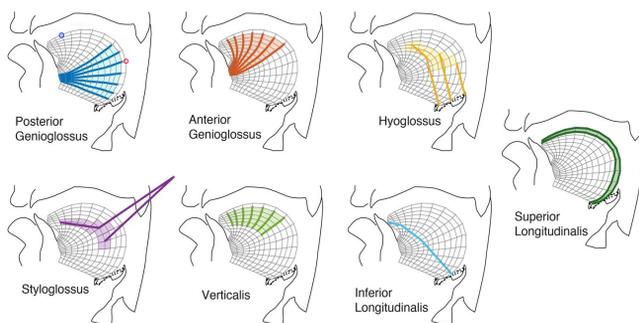


FIGURE 1 – Représentation des sept muscles pris en compte dans le modèle (Génioglosse Postérieur (GPP), Génioglosse Antérieur (GGA), Hyoglosse (HYO), Styloglosse (STYLO), Verticalis (VERT), Longitudinal Supérieur (SL), Longitudinal inférieur (LI)). Les lignes en gras représentent les macro-fibres, sur lesquelles la force musculaire globale est distribuée. Les quadrilatères en couleur sont des éléments sélectionnés dans la structure EF, dont la raideur augmente avec l'activation musculaire. Pour comparer les caractéristiques cinématiques des différents modèles de contrôle, nous avons choisi deux points de la langue caractérisant bien les propriétés phonétiques de l'articulation, un dans la zone palatale (point bleu sur la figure en haut à gauche), et un dans la zone pharyngienne (point rouge). Ces couleurs seront reprises dans les Figures 4, 5, et 6

Le mécanisme de génération de la force musculaire se base sur l'hypothèse du Point d'Equilibre (Feldman, 1986), selon laquelle les mouvements s'effectuent d'un point d'équilibre mécanique (c'est-à-dire une configuration où toutes les forces s'exerçant sur l'effecteur se combinent pour donner une résultante nulle) vers un autre. Dans le "modèle λ " associé à cette théorie, la force générée par un muscle dépend de la différence entre la longueur de ce muscle et un seuil d'activation λ qui est la variable de commande, implémentant ainsi un réflexe de bas niveau de type "réflexe d'étirement".

Pour une position de la langue donnée (mâchoire et ouverture labiale fixes), la fonction d'aire du conduit vocal est déduite de données anatomiques de référence (Perrier *et al.*, 1992). Les formants sont calculés avec un analogue acoustique harmonique du conduit vocal (Badin & Fant, 1984).

Pour la planification motrice, GEPPETO suppose que le système nerveux central stocke l'association entre les valeurs des variables de contrôle (λ) associées aux positions cibles et les fréquences des 3

premiers formants correspondants. Le modèle exploite cette association pour optimiser les commandes motrices réalisant les différents phonèmes de la séquence, sous une contrainte perceptive (Patri *et al.*, 2015). La figure 2 résume l'architecture complète du modèle.

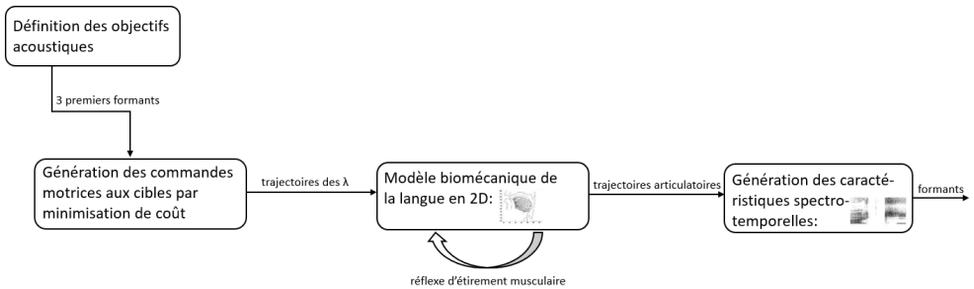


FIGURE 2 – L'architecture du modèle GEPPETO.

2.2 Le modèle de contrôle optimal

Comme GEPPETO, le modèle de contrôle optimal vise à atteindre des objectifs formantiques en utilisant le même modèle biomécanique. Toutefois, alors que GEPPETO sépare la phase de planification, dans laquelle un choix optimal des commandes cibles est opéré, de la phase d'exécution (purement feedforward), le modèle proposé ici ne sépare pas planification et exécution, et implémente l'optimalité au cours de l'exécution des mouvements, une fois spécifiées la cible formantique et la durée T du mouvement.

Pour déterminer les commandes motrices qui permettent de générer les mouvements souhaités du modèle biomécanique, un module appelé *contrôleur optimal* est utilisé pour calculer la trajectoire des λ au cours du temps en minimisant une fonction de coût. Celle-ci contient un terme lié à l'effort neuromusculaire et un terme correspondant à une pénalité sur la précision :

$$C = \int_0^T \|u(t)\|^2 dt + \alpha \|p(T) - p_{goal}\|^2$$

où u est un vecteur qui contient l'écart entre les commandes motrices actuelles et celles de la position de repos, α est un paramètre de compromis entre précision et effort, p_{goal} est un vecteur contenant l'objectif perceptif dans l'espace des 3 formants ainsi que la spécification d'une vitesse finale nulle, et $p(T)$ sa valeur réelle à la fin de la trajectoire. Pour ce problème d'optimisation contrainte, nous utilisons la méthode de l'adjoint (Bryson, 1999).

Les retours sensoriels étant retardés du fait du temps de transport sur les fibres nerveuses, mais aussi de leur traitement cortical, il est important d'avoir un estimateur de l'état du système pour éviter les problèmes d'instabilité. Il existe donc dans notre modèle de contrôle optimal un *estimateur optimal* qui génère une estimation de l'état du système à chaque pas de temps en se basant sur les retours sensoriels retardés et l'estimation d'un modèle interne entraîné à prédire l'état du système au prochain pas de temps en fonction de l'état actuel et d'une copie des commandes motrices (copie d'efférence). Pour (entre autres) accélérer les calculs, ce modèle prédictif est un modèle simplifié du modèle EF, appelé modèle biomécanique "réduit". Le contour supérieur de la langue, décrit par la position de 16 noeuds dans le plan sagittal, soit 32 variables, a été réduit via une analyse en composantes principales en un vecteur à 4 dimensions. La fonction d'évolution temporelle du système biomécanique a été apprise par un perceptron à deux couches de 28 neurones, entraîné sur la base de simulations du

modèle EF. Ce perceptron prédit l'accélération à un instant donné, étant données la position, la vitesse et les commandes motrices (λ) au pas de temps précédent. La précision de ce modèle réduit est supérieure à celle du modèle utilisé dans une étude préliminaire antérieure (Baraduc & Perrier, 2017). Un schéma qui résume l'architecture complète du modèle de contrôle optimal est donné à la Figure 3.

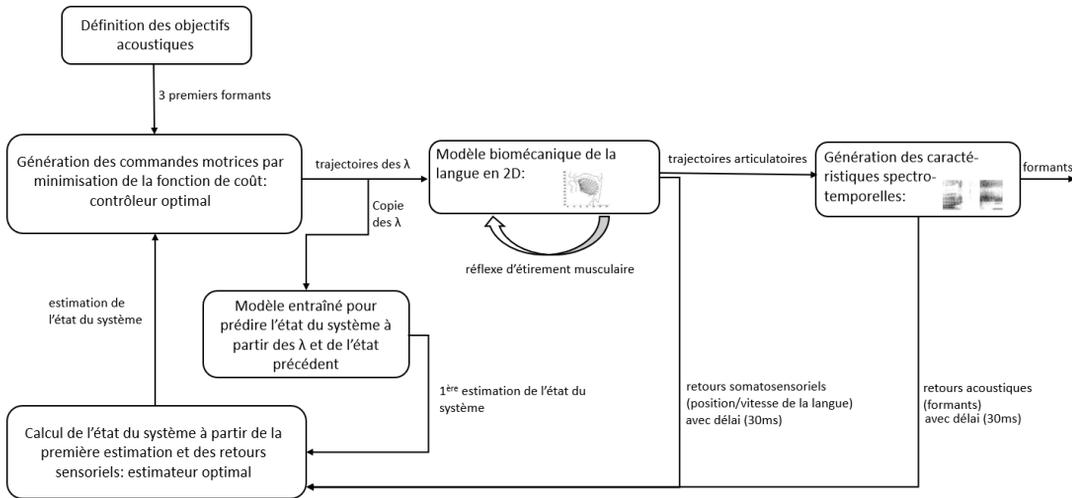


FIGURE 3 – Architecture du modèle de contrôle optimal proposé pour le mouvement depuis la position de repos vers une voyelle cible. Nous faisons l'hypothèse qu'un objectif perceptif (ici les trois premiers formants) doit être atteint à la fin de la trajectoire, lorsque la voyelle cible est atteinte. Les commandes motrices sont calculées à partir d'une minimisation de la fonction de coût (intégrant l'erreur à la cible sur les 3 premiers formants et l'effort neuromusculaire total) calculée à partir d'une estimation de l'état du système obtenue à partir d'un filtre de Kalman. Ces commandes sont envoyées au modèle biomécanique et génèrent un mouvement de la langue. Le contour supérieur de la langue, combiné avec les positions fixes des autres articulateurs issues de données de référence (Badin *et al.*, 2008), définit la géométrie du conduit vocal. On peut alors en calculer les formants suivant la méthode de Badin & Fant (1984). Ce retour sensoriel permet de mettre à jour l'estimation d'état et le coût.

3 Résultats

Afin d'identifier les effets de la dynamique de l'articulateur, nous présentons les activations musculaires et les mouvements obtenus pour les deux points caractéristiques du contour lingual (cf. Figure 1) par les différents modèles étudiés, pour des transitions "repos[schwa]-voyelle" de durées "courtes" (90 ms dont 65 ms de temps d'activation et 25 ms de temps de maintien) et pour des transitions plus longues (150 ms dont 130 ms de temps d'activation et 20 ms de temps de maintien). Cette comparaison est faite entre les activations et les mouvements générés dans trois conditions : (1) contrôleur optimal appliqué au modèle biomécanique "réduit", (2) GEPPETO quand les commandes cibles sont les valeurs proposées par le contrôleur optimal à la fin du mouvement (ou la valeur minimum des commandes au cours de la trajectoire quand il y a une différence importante entre les deux), et (3) commandes inférées par le contrôleur optimal appliquées au modèle biomécanique EF. La comparaison des résultats (1) et (3) permet d'observer les effets de la réduction de modèle utilisée pour calculer les commandes optimales.

Dans les Figures 4, 5, et 6, les courbes en astérisques (*) correspondent à la condition 2, les courbes en traits discontinus à la condition 1, et les courbes en traits continus à la condition 3.

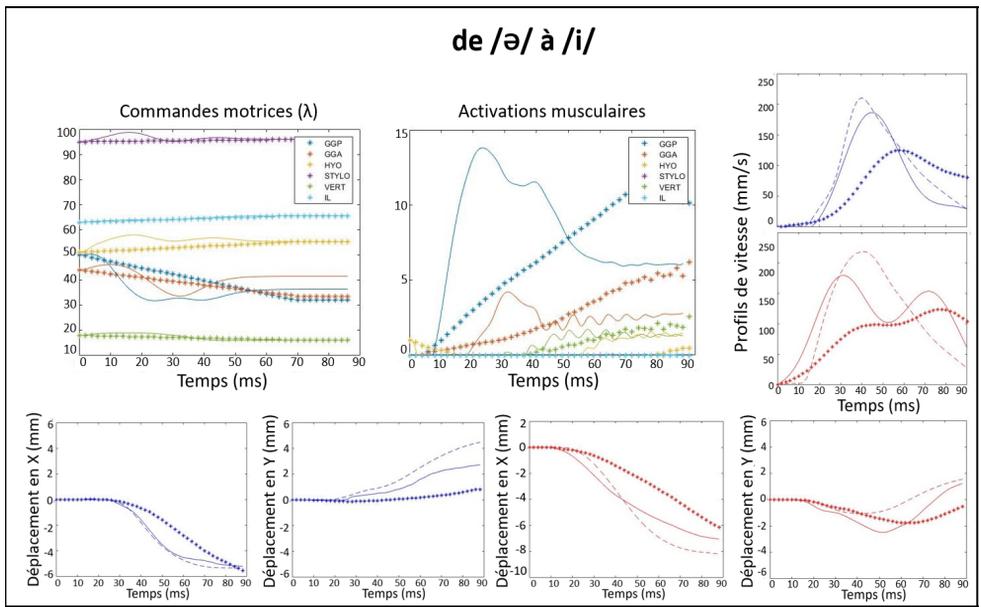
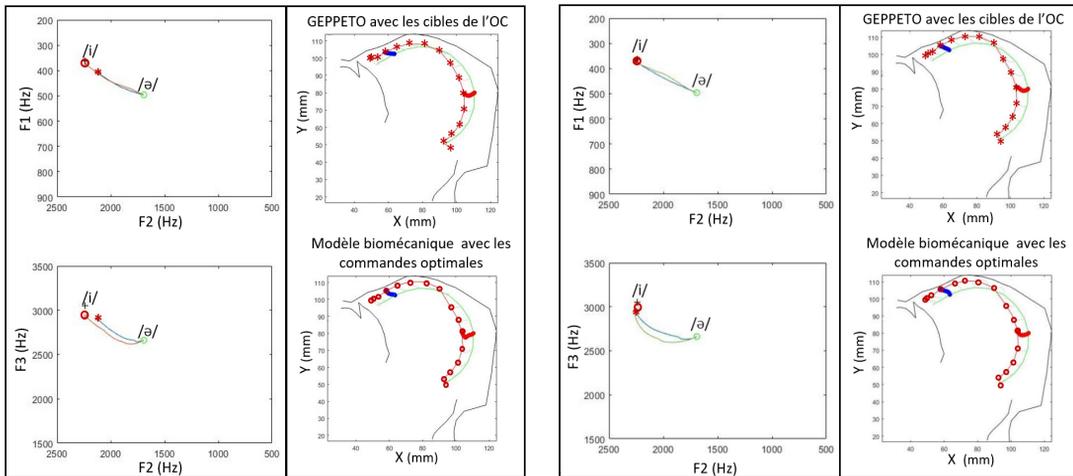


FIGURE 4 – Transitions "repos-*i*/" de 90 ms. En haut : Commandes motrices λ prises en compte (à gauche), activations musculaires (au milieu ; le supérieur longitudinalis n'a, pour des raisons de stabilité des simulations, pas été activé dans ces exemples) et profils de vitesse (vitesse tangentielle, à droite) résultants. En bas : Déplacements horizontaux et verticaux des deux points caractéristiques de la langue (cf. Fig. 1, même choix de couleur) depuis la position de repos. Conventions graphiques : cf. texte.



(a) En 90 ms

(b) En 150 ms

- + cibles spécifiées à l'OC
- position de repos
- Positions finales (modèle biomécanique et commandes optimales)
- * positions finales (GEPPETO)

FIGURE 5 – Trajectoires des points caractéristiques de la langue (cf. Fig. 1) et contours linguaux au repos et à la cible (panneaux (a) et (b), droite) et trajectoires dans les plans (F1,F2) (panneaux (a) et (b) en haut à gauche) et (F2,F3) (panneaux (a) et (b) en bas à gauche) pour $[\text{ə} \rightarrow \text{i}]$. Dans les plans (F1,F2) et (F2,F3), les trajectoires du modèle GEPPETO finissent par une astérisque, celles du modèle optimal par un cercle.

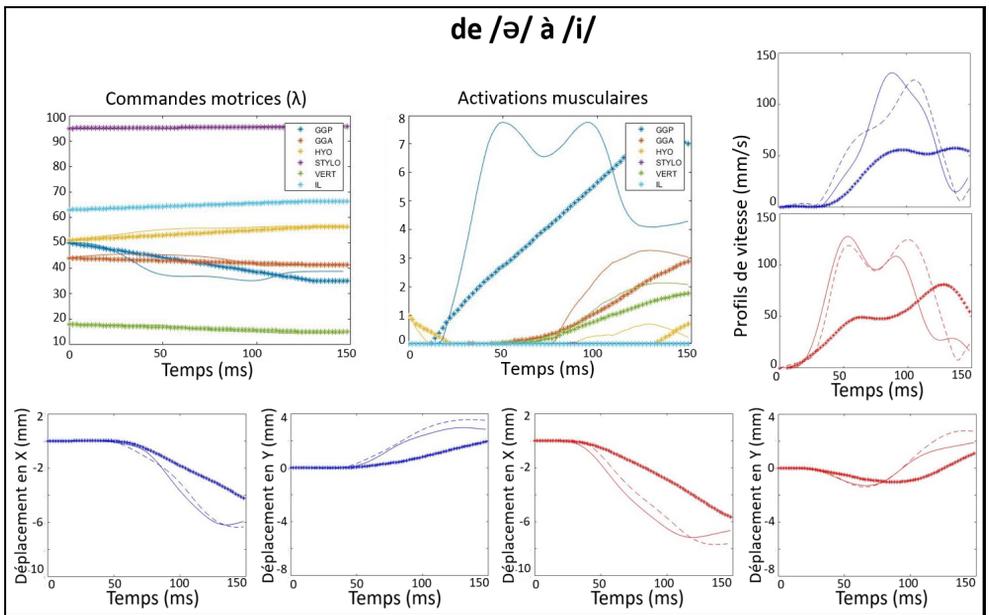


FIGURE 6 – Transitions "repos-/i/" de 150 ms. Même organisation et mêmes conventions graphiques qu'à la Figure 4.

Les Figures 4 et 6 illustrent les résultats obtenus à travers l'exemple des simulations de la transition "repos-/i/" de durées respectives 90 ms et 150 ms. Une première différence significative apparaît immédiatement : alors que la variation linéaire des commandes motrices du modèle GEPPETO génère des variations monotones et quasi-synchrones des activations musculaires, le modèle optimal active très vite et très fortement le ou les muscles agonistes les plus fortement impliqués dans la transition, dans le cas du /i/ le génioglosse postérieur (GGP). Cette stratégie a une conséquence cruciale pour la réalisation de la transition rapide, que nous pouvons observer sur la Figure 5 (Panneau (a), gauche) : alors que le modèle optimal permet l'atteinte des valeurs formantiques cibles en 90 ms, le modèle GEPPETO n'y arrive pas et génère une tendance à la neutralisation de la voyelle. Certes, le modèle GEPPETO pourrait vraisemblablement atteindre la cible dans le délai imparti en recrutant un niveau d'activation musculaire global plus élevé, comme l'ont proposé Patri *et al.* (2015), mais au prix d'un effort total et d'une co-contraction terminale supérieurs. Il est intéressant de noter que le modèle optimal propose la même stratégie pour la transition lente (dans la mesure où il cherche la stabilité à 130 ms), mais que le modèle GEPPETO a cette fois atteint parfaitement la cible dans l'intervalle de 150 ms. Ces différences de commande impactent les caractéristiques cinématiques du déplacement lingual (Figures 4 et 6) et donc la forme des trajectoires dans le plan sagittal (Figure 5, panneaux de droite) : dans le cas du /i/, le modèle optimal tend à induire dans le pharynx d'abord un mouvement antéro-postérieur puis un mouvement d'élévation, et dans la région palatale d'abord un mouvement d'élévation puis une antériorisation. Le modèle GEPPETO propose une variation plus distribuée dans les deux directions tout au cours du mouvement. Cette différence est plus marquée pour la transition rapide. Toutefois, il est important de souligner que cette différence stratégique est surtout marquée pour le /i/ qui nécessite une forte élévation de la langue vers l'avant, mais qu'elle l'est beaucoup moins les voyelles /e/ et /a/ par exemple.

4 Discussion et perspectives

Dans cette étude, nous avons développé un modèle de contrôle optimal appliqué à un modèle biomécanique de la langue lors de la production de parole. Notre but était de quantifier ce qu'apporte cette démarche, par rapport à une conception plus simple, implémentée dans le modèle GEPPETO, qui suppose une optimalité limitée à l'inférence des commandes motrices correspondant aux cibles (phonèmes) successives de la séquence étudiée. Cette évaluation est d'autant plus importante que, malgré l'efficacité prédictive de la théorie du contrôle optimal, elle a été critiquée sur son principe, entre autres par Friston (2011) dans sa théorie de l'inférence active qui suggère un *a priori* probabiliste sur le comportement plutôt qu'une minimisation de fonctions de coût.

Cette étude sur le rôle de l'optimalité dans le contrôle de l'exécution d'une tâche de parole est encore assez incomplète, puisqu'elle n'aborde pas la production de consonnes et se limite à la production de phonèmes isolés. Néanmoins, cet article est l'occasion de présenter notre démarche modélisatrice dans le contexte de nos travaux autour du modèle GEPPETO. Le modèle de contrôle optimal permet de prendre en compte les contraintes temporelles et les contraintes de précision sur les cibles, qui sont intégrées de manière plus globale dans GEPPETO. De manière intéressante, les variations temporelles des activations musculaires obtenues avec le modèle de contrôle optimal semblent plus réalistes que celles de GEPPETO : elles montrent la possibilité d'avoir des activations musculaires fortes dès le début du mouvement (cf. par exemple Tuller *et al.* 1982) qu'on ne peut pas obtenir simplement avec GEPPETO. Cependant, les profils de vitesses (Figure 4) montrent une différence significative selon que les commandes optimales sont appliquées au modèle réduit ou au modèle biomécanique. Ceci reflète les limites de l'utilisation d'un modèle réduit pour optimiser les commandes motrices ; ce modèle réduit est performant en général mais devient biaisé pour des dynamiques très rapides.

L'approfondissement à venir de ce travail passera par la simulation de consonnes, et, évidemment, la génération de séquences, dans lesquelles non seulement la cinématique des différentes parties de la langue, mais aussi les phénomènes de co-articulation seront au centre de nos observations. Dans ce contexte, le développement d'un modèle réduit qui puisse mieux prendre en compte des non-linéarités inhérentes aux contacts consonantiques entre la langue et les parois du conduit vocal est primordial. Pour cela nous envisageons de remplacer l'analyse en composantes principales, que nous avons utilisée ici pour effectuer la réduction de dimension, par des méthodes plus avancées d'apprentissage automatique comme les autoencodeurs. De même, la réduction de modèle devrait faire intervenir un modèle récurrent de type LSTM (Long Short Term Memory) pour mieux reproduire la dynamique de nos données. Pour finir, il faut noter qu'actuellement, le filtre de Kalman présent dans le modèle n'est pas exploité puisque nous avons simulé des trajectoires sans perturbations : les retours sensoriels étaient toujours cohérents avec les prédictions. Bien entendu, l'analyse des effets de différents types de perturbations sur le modèle fait également partie des perspectives de ce projet.

Remerciements

Ce travail a été financé par le Multidisciplinary Institute in Artificial Intelligence de l'Université Grenoble Alpes (MIAI@Grenoble Alpes, ANR-19-P3IA-0003), Chaire "Bayesian Cognition and Machine Learning for Speech Communication" (<http://www.gipsa-lab.fr/projet/MIAI-Speech/>)

Références

- BADIN P., ELISEI F., BAILLY G. & TARABALKA Y. (2008). An audiovisual talking head for augmented speech generation : models and animations based on a real speaker's articulatory data. In *International Conference on Articulated Motion and Deformable Objects*, p. 132–143 : Springer.
- BADIN P. & FANT G. (1984). Notes on vocal tract computation. *STL QPSR*, **2**(3), 53–108.
- BARADUC P. & PERRIER P. (2017). Motor control of the tongue during speech : Predictions of an optimization policy under sensorimotor noise. In *Society for Neuroscience Annual meeting (Neuroscience 2017)*, p. 408–03.
- BRYSON A. E. (1999). *Dynamic Optimization*. Addison Wesley.
- CAI S., GHOSH S. S., GUENTHER F. H. & PERKELL J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience*, **31**(45), 16483–16490.
- FELDMAN A. G. (1986). Once more on the equilibrium-point hypothesis (λ model) for motor control. *Journal of Motor Behavior*, **18**(1), 17–54.
- FRISTON K. (2011). What is optimal about motor control? *Neuron*, **72**(3), 488–498.
- GUENTHER F. H., GHOSH S. S. & TOURVILLE J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, **96**(3), 280–301.
- GUIGON E., BARADUC P. & DESMURGET M. (2007). Coding of movement- and force-related information in primate primary motor cortex : a computational approach. *European Journal of Neuroscience*, **26**(1), 250–260.
- HOUDE J. F. & NAGARAJAN S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, **5**, 82.
- MERMELSTEIN P. (1973). Articulatory model for the study of speech production. *The Journal of the Acoustical Society of America*, **53**(4), 1070–1082.
- PARRELL B., RAMANARAYANAN V., NAGARAJAN S. & HOUDE J. (2019). The facts model of speech motor control : Fusing state estimation and task-based control. *PLoS Comput. Biol.*, **15**(9), e1007321.
- PATRI J.-F., DIARD J. & PERRIER P. (2015). Optimal speech motor control and token-to-token variability : a bayesian modeling approach. *Biological Cybernetics*, **109**(6), 611–626.
- PATRI J.-F., DIARD J. & PERRIER P. (2019). Modeling sensory preference in speech motor planning : a bayesian modeling framework. *Frontiers in Psychology*, **10**, 2339.
- PERRIER P., BOË L. & SOCK R. (1992). Vocal tract area function estimation from midsagittal dimensions with ct scans and a vocal tract cast : Modeling the transition with two sets of coefficients. *Journal of Speech and Hearing Research*, **35**(1), 53–67.
- PERRIER P., PAYAN Y., ZANDIPOUR M. & PERKELL J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants : A modeling study. *The Journal of the Acoustical Society of America*, **114**(3), 1582–1599.
- SALTZMAN E. (1986). Task dynamic coordination of the speech articulators : A preliminary model. *Experimental Brain Research Series*, **15**, 129–144.
- TODOROV E. & JORDAN M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, **5**(11), 1226–1235.

TULLER B., KELSO J. & HARRIS K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology : Human Perception and Performance*, **8**(3), 460–472.