

Modélisation bayésienne de la planification motrice des gestes de parole : Évaluation du rôle des différentes modalités sensorielles

Jean-François Patri, Julien Diard, Pascal Perrier

▶ **To cite this version:**

Jean-François Patri, Julien Diard, Pascal Perrier. Modélisation bayésienne de la planification motrice des gestes de parole : Évaluation du rôle des différentes modalités sensorielles. JEP-TALN-RECITAL 2016 - conférence conjointe 31e Journées d'Études sur la Parole, 23e Traitement Automatique des Langues Naturelles, 18e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, Jul 2016, Paris, France. pp.419-427. hal-01345199v2

HAL Id: hal-01345199
https://hal.science/hal-01345199v2

Submitted on 19 Jul 2016

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.

Modélisation bayésienne de la planification motrice des gestes de parole : Évaluation du rôle des différentes modalités sensorielles

Jean-François Patri[1,2,3,4] Julien Diard[3,4] Pascal Perrier[1,2]
(1) Univ. Grenoble Alpes, Gipsa-lab, F-38000 Grenoble, France
(2) CNRS, Gipsa-lab, F-38000 Grenoble, France
(3) Univ. Grenoble Alpes, LPNC, F-38000 Grenoble, France
(4) CNRS, LPNC, F-38000 Grenoble, France
Jean-Francois.Patri@gipsa-lab.grenoble-inp.fr,
Julien.Diard@upmf-grenoble.fr, Pascal.Perrier@gipsa-lab.grenoble-inp.fr

RÉSUMÉ

La prise en compte des informations auditives et proprioceptives dans le contrôle de la parole est mise en évidence par un nombre croissant de résultats expérimentaux. Cependant, les modèles de production imposent le plus souvent l'une ou l'autre des modalités, ou n'offrent pas de cadre formel pour évaluer leurs contributions respectives. Nous proposons d'explorer le rôle de ces modalités sensorielles dans la planification des gestes de parole à partir d'un modèle bayésien représentant la structure des connaissances mises en jeu dans cette tâche. Le modèle permet d'envisager trois mécanismes de planification, reposant sur la modalité auditive, proprioceptive ou sur les deux conjointement. Nous comparons des simulations obtenues par les deux premiers mécanismes de planification. Les résultats indiquent des réalisations articulatoires différentes mais donnant néanmoins des réalisations auditives qualitativement similaires dans leur variabilité.

ABSTRACT

Bayesian modeling of speech gesture motor planning: Evaluating the role of different sensory modalities

An increasing number of experimental results have identified a clear role of auditory and somatosensory information in speech motor control. However, most of the speech production models consider only one of these sensory modalities, or do not provide the possibility to formally evaluate the respective contribution of these modalities. We propose to explore the role of auditory and proprioceptive representations in speech gesture planning, based on a Bayesian model representing the structure of knowledge involved. The model allows to consider three planning mechanisms, based on the auditory or proprioceptive modality or the combination of both. We compare simulations obtained from the two first planning mechanisms. Results indicate differences in the generated articulatory patterns, giving rise however to qualitatively similar patterns of auditory variability.

MOTS-CLÉS : Contrôle moteur de la parole – Modélisation bayésienne – Multimodalité .

KEYWORDS: Speech motor control – Bayesian modeling – Multimodality.

1 Introduction

La remarquable capacité d'adaptation des gestes de parole à différents contextes et perturbations démontre que leur contrôle n'est pas le résultat d'un simple apprentissage stéréotypé mais met en jeu des mécanismes de planification, qui se révèlent en particulier par des stratégies d'anticipation (Noiray *et al.*, 2011). Quelles connaissances sont mises en jeu dans ces mécanismes et comment sont-elles structurées ? De nombreuses études expérimentales ont mis en évidence la prise en compte des information auditives et somato-sensorielles dans le contrôle et l'adaptation des gestes de parole (Lametti *et al.*, 2012). Le poids spécifique de ces retours sensoriels dans la la caractérisation des buts physiques associés aux unités phonologiques n'est pas connu et fait l'objet de nombreux débats, allant jusqu'à des enjeux importants en phonologie (Browman & Goldstein, 1992). Comment sont-ils pris en compte ? Ont-ils le même statut vis-à-vis du contrôle ? Sont-ils pris en compte de façon indépendante ou en combinaison ?

Cependant, à l'exception de DIVA (Guenther *et al.*, 2006), la plupart des modèles de production définissent les buts de parole dans l'une ou l'autre seulement de ces modalités sensorielles (Saltzman & Munhall, 1989; Stevens, 1993) et n'offrent pas de cadre formel pour une évaluation de leurs contributions respectives. Ce travail propose une méthodologie de modélisation pour étudier ces questions dans la planification des gestes de parole à partir d'un modèle bayésien des connaissances mises en jeu dans cette tâche (Bessière *et al.*, 2013). Ce modèle, qui est une extension d'un modèle précédent (Patri *et al.*, 2015), permet de définir et évaluer trois mécanismes de planification, reposant soit sur la modalité auditive, sur la modalité proprioceptive ou sur les deux conjointement.

Nous considérons la question de la planification des commandes motrices envoyées aux muscles de la langue pour la production de phonèmes isolés. Les six muscles les plus importants pour expliquer les déformations de la langue dans le plan sagittal sont représentés dans un modèle biomécanique bi-dimensionnel de la langue (Perrier *et al.*, 2003). Le niveau d'activation de chaque muscle est contrôlé selon la théorie du point d'équilibre par une variable de contrôle λ qui spécifie une longueur seuil pour le muscle, au delà de laquelle une force musculaire active est générée (Feldman, 1986). Les caractéristiques spectrales du signal acoustique correspondant à chaque configuration de langue sont obtenues par le calcul de la fonction d'aire, puis par l'utilisation d'un modèle harmonique de la propagation des ondes dans le conduit vocal (Badin & Fant, 1984). Nous considérons les trois premiers formants comme la représentation auditive des signaux acoustiques à produire (Paliwal *et al.*, 1983). La nature précise des indices sensoriels caractérisant la proprioception de la langue est encore mal connue. Nous supposons qu'ils sont caractérisés par les longueurs des fibres des six muscles considérés. Nous supposons que le cerveau dispose de deux modèles internes lui permettant de prédire les images auditives et proprioceptives correspondant à chaque configuration de paramètres de contrôle λ (Callan *et al.*, 2004). Nous supposons également que le cerveau dispose d'une représentation des buts auditifs et proprioceptifs associés à chaque phonème et de leurs domaines de variabilité compatibles avec une production correcte du phonème (Hickok, 2014). Ceci caractérise les distributions des images auditives et proprioceptives attendues pour chaque phonème. Nous supposons finalement que le cerveau est en mesure de comparer les prédictions sensorielles effectuées par les modèles internes moteurs avec les attentes correspondant aux caractérisations sensorielles des phonèmes (Blakemore *et al.*, 2000).

La Section 2 présente la construction du modèle, par une traduction de ces hypothèses dans le formalisme bayésien. La Section 3 expose les résultats de simulations de production de phonèmes isolés selon la modalité sensorielle considérée.

2 Présentation du modèle

2.1 Description

Variables probabilistes Chaque grandeur intervenant dans la planification des commandes motrices est associée à une variable probabiliste. Ces variables sont les suivantes :

$\boldsymbol{M}$ est la variable motrice, définie comme une variable 6-dimensionnelle : $M \equiv (\lambda_1, \ldots, \lambda_6)$. Les λ_i sont les variables de contrôle pilotant la contraction des six muscles de la langue considérés par le modèle biomécanique.

$\boldsymbol{\Phi}$ est une variable discrète composée des phonèmes { /i/, /e/, /ɛ/, /a/, /oe/, /ɔ/, /k/ }[1].

$\boldsymbol{S_a^M}$ et $\boldsymbol{S_a^\Phi}$ représentent les informations auditives mises en jeu dans la planification. S_a^M est générée à partir de la variable motrice M via le modèle interne auditif. S_a^Φ correspond à la représentation des buts auditifs associés à chaque phonème. Ce sont des variables tridimensionnelles continues caractérisées par les valeurs des trois premiers formants du signal acoustique : $S_a^{M,\Phi} \equiv (F_1, F_2, F_3)$.

$\boldsymbol{S_p^M}$ et $\boldsymbol{S_p^\Phi}$ représentent les informations proprioceptives mises en jeu dans la planification. S_p^M et S_p^Φ sont respectivement associées à M et Φ via le modèle interne et la représentation des buts proprioceptifs de chaque phonème. On définit alors S_p^M et S_p^Φ comme deux variables continues 6-dimensionnelles $S_p^{M,\Phi} \equiv (L_1, \ldots, L_6)$, où L_i est la longueur spécifiée du muscle i.

Les domaines des variables M, $S_{a,p}^M$ et $S_{a,p}^\Phi$ sont définis par les domaines de variation admissibles dans le modèle biomécanique.

$\boldsymbol{C_a}$ et $\boldsymbol{C_p}$ sont deux variables introduites pour contraindre la cohérence entre les valeurs prises respectivement par les variables S_a^M et S_a^Φ d'une part et S_p^M et S_p^Φ d'autre part. Ce sont des variables binaires égales à 1 ou 0 selon que les valeurs prises par S_a^M et S_a^Φ et S_p^M et S_p^Φ sont identiques ou pas. Lorsqu'elles sont fixées à 1 elles imposent la cohérence des prédictions motrices $S_{a,p}^M$ et des attentes $S_{a,p}^\Phi$ liées à la caractérisation des buts associés aux phonèmes.

Décomposition Le schéma de la Figure 1 décrit l'organisation des connaissances proposée dans notre modèle. Il permet de formuler la distribution de probabilité conjointe de l'ensemble des variables $P(M\ S_a^M\ S_p^M\ S_a^\Phi\ S_p^\Phi\ C_a\ C_p)$ par la décomposition suivante :

$$
\begin{aligned}
& P(M\ S_a^M\ S_p^M\ S_a^\Phi\ S_p^\Phi\ C_a\ C_p) \qquad (1)\\
&= \quad P(M)P(S_p^M \mid M)P(S_a^M \mid M)P(\Phi)P(S_p^\Phi \mid \Phi)P(S_a^\Phi \mid \Phi)P(C_a \mid S_a^M\ S_a^\Phi)P(C_p \mid S_p^M\ S_p^\Phi).
\end{aligned}
$$

Cette décomposition traduit, par ses hypothèses simplificatrices, les relations supposées entre variables. Par exemple, le terme $P(S_a \mid M)$ traduit une hypothèse d'indépendance entre S_a et S_p conditionnellement à la connaissance de M, qui est suggérée par l'hypothèse du modèle interne moteur, d'après lequel M seul suffit à caractériser S_a. En d'autres termes, S_p n'apporte aucune information supplémentaire à l'identification de S_a si M est connue.

[1] Cette liste est limitée aux phonèmes réalisables par le modèle biomécanique de la langue, sans usage spécifique des lèvres, non représentées dans cette version du modèle.

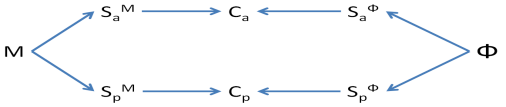


FIGURE 1 : Réseau représentant la décomposition de l'Équation (1).

Formes paramétriques Les termes de la décomposition de l'Équation (1) sont spécifiés par les formes paramétriques données ci-dessous.

$\boldsymbol{P(M)}$ représente la connaissance a priori sur la variable motrice. Nous supposons que toutes les configurations motrices utilisées pour la parole sont aisément réalisables et que de ce fait aucune n'est a priori privilégiée. $P(M)$ est donc identifiée à une distribution uniforme sur l'ensemble de son domaine de définition.

$\boldsymbol{P(S_a^M \mid M)}$ et $\boldsymbol{P(S_p^M \mid M)}$ représentent les connaissances des informations auditives et proprioceptives associées à la connaissance de M. Ces connaissances sont attribuées aux modèles internes prédisant les informations sensorielles à partir des commandes motrices. Notons $s_a^*(m)$ et $s_p^*(m)$ les valeurs prédites à partir de la valeur m de M. Les formes des distributions de probabilité associées sont définies par $P([S_{a,p}^M = s_{a,p}] \mid [M = m]) = \delta_{s^*_{a,p}(m)}(s_{a,p})$. δ_x correspond à la distribution Dirac centrée en x, traduisant le déterminisme, chez un locuteur donné, des relations entre les commandes motrices et le son produit. $P([S = s_{a,p}] \mid [M = m])$ est donc nulle tant que $s_{a,p} \neq s^*_{a,p}(m)$.

$\boldsymbol{P(S_a^\Phi \mid \Phi)}$ et $\boldsymbol{P(S_p^\Phi \mid \Phi)}$ représentent les connaissances des buts sensoriels correspondant à un phonème donné. On les identifie aux distributions des signaux sensoriels observés expérimentalement pour chaque phonème. Les productions acoustiques de chaque phonème étant caractérisées par des ellipsoïdes de dispersion dans l'espace des trois premiers formants, nous identifions $P(S_a^\Phi \mid \Phi)$ à une loi normale spécifiée par les paramètres de l'ellipsoïde correspondant au phonème considéré. La caractérisation des signaux proprioceptifs correspondant aux phonèmes est plus délicate. Nous supposons que la caractérisation auditive des phonèmes se met d'abord en place, permettant de guider les premières productions du sujet. La prise en compte des informations proprioceptives associées à ces productions initiales $P(S_p^M \mid \Phi\ [C_a = 1])$ permettrait alors au sujet de construire peu à peu une caractérisation proprioceptive de ses phonèmes $P(S_p^\Phi \mid \Phi)$. En d'autres termes, cet apprentissage identifie $P(S_p^\Phi \mid \Phi)$ au résultat de l'inférence $P(S_p^M \mid \Phi\ [C_a = 1])$:

$$P(S_p^\Phi \mid \Phi) \equiv P(S_p^M \mid \Phi\ [C_a = 1]) \propto \sum_m P(S_p^M \mid [M = m]) P([S_a^\Phi = s_a^*(m)] \mid \Phi). \tag{2}$$

$\boldsymbol{P(C_a \mid S_a^M\ S_a^\Phi)}$ et $\boldsymbol{P(C_p \mid S_p^M\ S_p^\Phi)}$ correspondent enfin aux contraintes de cohérence :

$$P([C_{a,p} = 1] \mid [S_{a,p}^M = s^M]\ [S_{a,p}^\Phi = s^\Phi]) = \begin{cases} 1 & \text{si } s^M = s^\Phi \\ 0 & \text{sinon.} \end{cases} \tag{3}$$

2.2 Utilisation du modèle

Nous simulons le mécanisme par lequel le cerveau sélectionnerait, selon nos hypothèses, les commandes motrices à envoyer aux muscles de la langue pour produire un phonème. Nous étudions donc l'inférence de la variable M à partir du phonème que l'on cherche à produire. Le modèle nous permet d'aborder cette question de trois façons : a) en imposant avec $C_a = 1$ la cohérence des variables

auditives S_a^M et S_a^Φ, l'inférence de $P(M \mid \Phi\ [C_a = 1])$ planifie les commandes motrices pour la production d'un phonème reposant sur les connaissances auditives seules ; b) de même, en imposant $C_p = 1$, $P(M \mid \Phi\ [C_p = 1])$ planifie les commandes motrices reposant sur les connaissances proprioceptives seules ; c) finalement en imposant à la fois $C_a = 1$ et $C_p = 1$, $P(M \mid \Phi\ [C_p = 1][C_a = 1])$ planifie les commandes motrices à partir des connaissances auditives et proprioceptives, suggérant une caractérisation bimodale des buts moteurs de parole.

Ces trois inférences sont obtenues à partir de la distribution de probabilité conjointe $P(M\ S_a^M\ S_p^M\ S_a^\Phi\ S_p^\Phi\ C_a\ C_p)$ spécifiée par l'Équation (1). Pour $P(M \mid \Phi\ [C_a = 1])$ on a :

$$P(M|\Phi[C_a = 1]) = \frac{P(M\Phi[C_a = 1])}{P(\Phi[C_a = 1])} = \frac{\sum_{S_a^M, S_a^\Phi, S_p^M, S_p^\Phi, C_p} P(M S_a^M S_p^M S_a^\Phi S_p^\Phi [C_a = 1] C_p)}{\sum_{M, S_a^M, S_a^\Phi, S_p^M, S_p^\Phi, C_p} P(M S_a^M S_p^M S_a^\Phi S_p^\Phi [C_a = 1] C_p)}.$$

En remplaçant $P(M\ S_a^M\ S_p^M\ S_a^\Phi\ S_p^\Phi\ C_a\ C_p)$ par l'expression de l'Équation (1) et en effectuant les sommations pour le cas d'une valeur m de M on a :

$$P([M = m] \mid \Phi\ [C_a = 1]) \quad \propto \quad P([S_a^\Phi = s_a^*(m)] \mid \Phi), \tag{4}$$

où le symbole de proportionnalité $\propto$ tient compte du facteur de normalisation. De façon similaire on obtient :

$$P([M = m] \mid \Phi\ [C_p = 1]) \quad \propto \quad P([S_p^\Phi = s_p^*(m)] \mid \Phi), \tag{5}$$

$$P([M = m] \mid \Phi\ [C_a = 1]\ [C_p = 1]) \quad \propto \quad P([S_a^\Phi = s_a^*(m)] \mid \Phi) P([S_p^\Phi = s_p^*(m)] \mid \Phi). \tag{6}$$

3 Résultats

3.1 Comparaison formelle des inférences auditive et proprioceptive

Avant d'exposer les résultats de l'implémentation du modèle, notons un premier résultat formel. Les Équations (2) et (5) donnent lieu à :

$$P([M = m] \mid \Phi\ [C_p = 1]) \quad \propto \quad \sum_{m'} P([S_p^M = s_p^*(m) \mid [M = m']) P([S_a^\Phi = s_a^*(m')] \mid \Phi). \tag{7}$$

En comparant à l'Équation (4), on constate que $P([M = m] \mid \Phi\ [C_p = 1]) \neq P([M = m] \mid \Phi\ [C_a = 1])$. Autrement dit, les branches auditive et proprioceptive ne proposent pas les mêmes inférences motrices. Cela peut paraitre étonnant étant donné que la caractérisation proprioceptive des phonèmes a été définie sur la base de leur caractérisation auditive. On peut remarquer que les inférences seraient identiques si chaque valeur de S_p^M était générée par une unique valeur de M. En effet dans ce cas la somme sur m' dans l'Équation (7) se réduirait à un unique terme : celui correspondant à la valeur m considérée pour M. Dans ce cas là on obtiendrait :

$$P([M = m] \mid \Phi\ [C_p = 1]) \propto P([S_a^\Phi = s_a^*(m)] \mid \Phi) \propto P([M = m] \mid \Phi\ [C_p = 1]). \tag{8}$$

Les branches proprioceptive et auditive deviendraient alors équivalentes. Cependant, étant donné qu'une même configuration articulatoire de la langue peut être obtenue à partir de différentes coactivations musculaires, l'équivalence précédente n'est pas assurée à priori. Ce serait également le cas

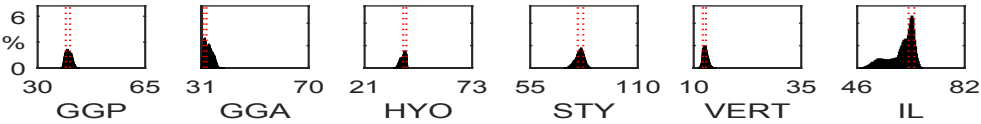


FIGURE 2 : Histogrammes des échantillons de S_p^Φ correspondant à l'inférence de l'Équation (2). Les lignes verticales indiquent la position des noyaux gaussiens retenus pour l'approximation de ces distributions.

si chaque valeur de S_p^M avait exactement le même nombre d'antécédents dans l'espace de M, mais aucune observation expérimentale ne laisse penser que cela puisse être le cas, et ceci est confirmé par notre modèle biomécanique.

Jusqu'à présent nous avons supposé une identification exacte de $P(S_p^\Phi \mid \Phi)$ avec l'inférence de $P(S_p^M \mid \Phi\,[C_a = 1])$ donnée par l'Équation (2). Cette identification formelle caractérise la distribution des valeurs de S_p^Φ dont les valeurs de S_a associées satisfont les cibles auditives. Bien que nous ayons choisi des cibles auditives simples, caractérisées par des gaussiennes, rien ne garantit que la distribution correspondante des valeurs de S_p^Φ soit également simple. La Figure 2 illustre les histogrammes des longueurs des muscles correspondant aux distribution marginales des échantillons de S_p obtenus par l'identification de l'Équation (2) obtenue par échantillonnage de type Monte-Carlo par chaines de Markov (MCMC) pour le phonème /e/. Bien que ces distributions marginales se caractérisent dans leur ensemble par des pics localisés, on observe également une structure plus complexe (en particulier pour le muscle Inferieur Longitudinalis dans l'exemple de la Figure 2). Cette complexité n'est pas étonnante étant donné les non-linéarités existant dans le lien entre l'articulatoire et l'acoustique. Cela suggère qu'une identification de $P(S_p^\Phi \mid \Phi)$ au résultat exact de l'inférence de l'Équation (2) demande une caractérisation proprioceptive complexe des phonèmes. Nous faisons l'hypothèse que cette complexité n'est pas stockée dans sa totalité dans la caractérisation des buts proprioceptifs associés à chaque phonème, et nous choisissons de définir $P(S_p^\Phi \mid \Phi)$ à partir d'une approximation par une mixture de quatre gaussiennes des échantillons de la Figure 2, dont on ne conserve au final que le mode gaussien de moindre variance. Les lignes verticales en pointillés sur la Figure 2 indiquent la position du noyaux gaussien retenu.

3.2 Comparaison des simulations des differentes inférences

Les trois inférences du modèle fournissent des distributions caractérisant les valeurs des commandes motrices (variable M) permettant de produire chaque phonème. Nous décrirons et comparerons dans cet article uniquement les résultats obtenus par les planifications dans l'espace auditif seul et dans l'espace proprioceptif seul. Ces distributions ont été approximées par un échantillonnage de type MCMC dans l'espace des commandes motrices. Pour comparer les deux stratégies de planification, nous étudions les résultats auditifs et proprioceptifs correspondants aux commandes motrices obtenues par chacune des inférences.

On s'attend à ce que la planification selon la modalité auditive donne lieu à une plus grande variabilité des configurations articulatoires, étant donné que la complexité des régions proprioceptives associées aux régions cibles dans l'espace acoustique nous a amenés à ne sélectionner qu'une zone restreinte dans chacune d'elles pour spécifier sous forme gaussienne les cibles à atteindre lors d'une planification

proprioceptive. Les images de droite de la Figure 3 présentent les résultats proprioceptifs [2] résultant de 100 échantillons de commandes motrices planifiées [3], selon la modalité proprioceptive (bas) et la modalité auditive (haut). Les ellipses grises représentent les ellipses de dispersion des données proprioceptives obtenues. Les ellipses noires représentent les cibles gaussiennes retenues pour la définition des cibles proprioceptives pour chaque phonème. Conformément à nos attentes, on constate que la planification dans l'espace proprioceptif donne des résultats dans cet espace conformes à la cible planifiée, alors que la planification dans l'espace auditif donne lieu à des configurations articulatoires nettement plus variables.

Dans le cadre de notre modélisation, la planification proprioceptive donne donc lieu à des formes du conduit vocal moins variables et plus prototypiques que la planification auditive. Du fait de la réduction de la région cible dans l'espace proprioceptif, que nous avons été amenés à proposer à cause de la complexité géométrique de la projection dans cet espace des régions cibles auditives, la planification proprioceptive n'autorise pas le recours à toutes les possibilités d'équivalence motrice qu'offre la planification auditive, à la fois dans le domaine des activations musculaires et dans le domaine des formes de langue. Dans ce contexte, la question qui se pose est celle de savoir si les deux modes de planification se différencient aussi par des variabilités différentes dans le domaine auditif. La réponse est donnée par les images de gauche de la Figure 3. Celles-ci présentent les résultats auditifs obtenus à partir des mêmes échantillons de commandes motrices utilisées pour générer les résultats proprioceptifs illustrés sur les images de droite. A nouveau, les ellipses grises représentent les ellipses de dispersion des données auditives obtenues. Les ellipses noires représentent les distributions gaussiennes caractérisant les cibles auditives pour chaque phonème. Nous constatons, et ce n'est pas une surprise, que les résultats obtenus par planification auditive sont en cohérence avec la cible planifiée. En revanche, nous observons de plus que les résultats obtenus par planification proprioceptive présentent globalement une variabilité similaire à celle obtenue par planification auditive [4], alors même que les formes de langue sont moins variables en planification proprioceptive. Cela indique que, même si l'ensemble des possibilités d'équivalence motrice n'est pas exploité par la planification proprioceptive, la variabilité des formes articulatoires autorisée par cette planification permet de générer une variabilité acoustique similaire à celle de la planification auditive. La plus grande variabilité proprioceptive liée à la planification auditive suggère une meilleure exploitation des possibilités d'équivalence motrice.

4 Discussion et conclusion

La comparaison des planifications vers des cibles gaussiennes définies dans l'espace proprioceptif et dans l'espace auditif a permis de mettre en évidence deux points importants. Tout d'abord l'approximation par un noyau gaussien des régions proprioceptives associées aux cibles auditives, elles

[2] Le sous espace des deux premières composantes principales des longueurs de muscles, obtenues sur l'ensemble des réalisations des phonèmes, est choisi pour visualisation. Cette caractérisation est pertinente étant donné que la dispersion des données dans l'espace des 6 longueurs de muscles est expliquée à plus de 90% par les trois premières composantes principales. On reconnait d'ailleurs dans les deux premières composantes extraites les dimensions haut/bas et avant/arrière à partir desquelles on positionne les voyelles dans l'espace du conduit vocal (Harshman *et al.*, 1977).

[3] Pour des raisons de clarté de la représentation, seuls 100 échantillons tirés aléatoirement parmi l'ensemble des commandes obtenues par l'échantillonnage MCMC ont été représentées.

[4] Les résultats auditifs obtenus pour le phonème /ɛ/ en planification proprioceptive présentent une variabilité bien plus importante où on retrouve une partie importante des résultats auditifs qui s'écartent de la cible correspondante. Ceci met en évidence des discontinuités ou non convexités de la projection proprioceptive des cibles auditives que le noyau gaussien utilisé retient à tort. Ce phénomène a vraisemblablement pour origine l'imprécision des modèles internes utilisés.

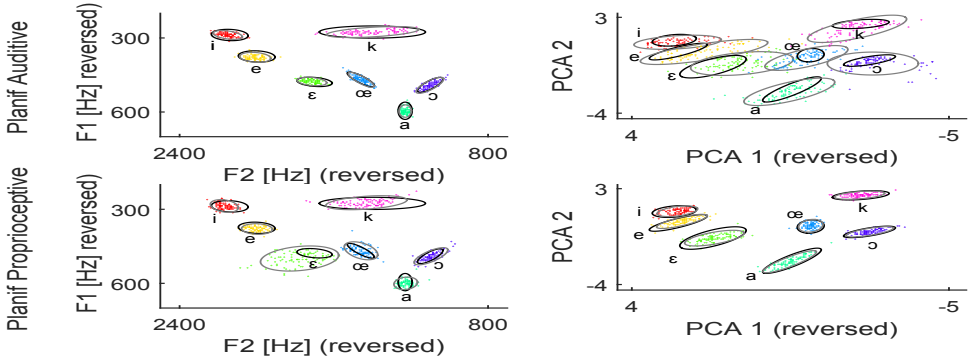


FIGURE 3 : Réalisations auditives (gauche) et proprioceptives (droite) obtenues par le modèle biomécanique de langue à partir de 100 commandes motrices obtenues par planification auditive (haut) et planification proprioceptive (bas) pour les sept phonèmes considérés. Les ellipses grises représentent les ellipses de dispersion des données générées. Les ellipses noires représentent les noyaux gaussiens caractérisant les cibles auditive (images de gauche) et proprioceptive (images de droite) pour chaque phonème

aussi gaussiennes, ne permet pas d'extraire toute la complexité de la distribution des configurations articulatoires donnant lieu à un même phonème. La planification dans l'espace proprioceptif donne donc naturellement des réalisations d'un même phonème moins variables, plus prototypiques. Cette planification ne permet pas l'exploitation de toutes les possibilités d'équivalence motrice.

On peut alors s'attendre à ce que la planification de séquences de phonèmes dans le domaine proprioceptif révèle une coarticulation de moins grande amplitude qu'une planification dans l'espace auditif. Ensuite, malgré la restriction des configurations articulatoires imposée par notre hypothèse de planification proprioceptive, la variabilité auditive observée est qualitativement similaire à celle générée par la planification auditive, bien que cette dernière exploite davantage de configurations articulatoires. Ainsi, d'un point de vue purement acoustique nos résultats suggèrent que, pour ce qui est de la production de phonèmes isolés du moins, les deux stratégies de planification seraient difficiles à distinguer. Dans la production de séquences de phonèmes cependant, on s'attend à ce qu'une planification reposant sur des cibles proprioceptives prototypiques induise des patrons de coarticulation différents.

Globalement, la modélisation que nous proposons des cibles auditives ou proprioceptives, ainsi que de la planification auditive ou proprioceptive, offre un cadre précis qui permet de prédire les propriétés des phonèmes planifiés selon l'une ou l'autre de ces modalités. Il convient maintenant d'évaluer la validité de ces prédictions par des expériences où des sujets seront placés dans des conditions favorisant l'une ou l'autre de ces planifications. C'est ce à quoi nous nous attachons désormais.

Remerciements

Ces recherches ont bénéficié du soutien financier du Conseil Européen de la Recherche sous le septième programme-cadre de l'Union Européenne (FP7/2007-2013 Grant Agreement no. 339152, "Speech Unit(e)s", PI : Jean-Luc-Schwartz).

Références

BADIN P. & FANT G. (1984). Notes on vocal tract computation. *Quarterly Progress and Status Report, Dept for Speech, Music and Hearing, KTH, Stockholm*, p. 53–108.

BESSIÈRE P., MAZER E., AHUACTZIN J. M. & MEKHNACHA K. (2013). *Bayesian Programming*. Boca Raton, Florida : CRC Press.

BLAKEMORE S.-J., WOLPERT D. & FRITH C. (2000). Why can't you tickle yourself ? *Neuroreport*, **11**(11), R11–R16.

BROWMAN C. P. & GOLDSTEIN L. (1992). Articulatory phonology : An overview. *Phonetica*, **49**(3-4), 155–180.

CALLAN D. E., JONES J. A., CALLAN A. M. & AKAHANE-YAMADA R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. *NeuroImage*, **22**(3), 1182–1194.

FELDMAN A. G. (1986). Once more on the equilibrium-point hypothesis (λ model) for motor control. *Journal of motor behavior*, **18**(1), 17–54.

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.

HARSHMAN R., LADEFOGED P. & GOLDSTEIN L. (1977). Factor analysis of tongue shapes. *The Journal of the Acoustical Society of America*, **62**(3), 693–707.

HICKOK G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, **29**(1), 2–20.

LAMETTI D. R., NASIR S. M. & OSTRY D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of neuroscience*, **32**(27), 9351–9358.

NOIRAY A., CATHIARD M.-A., MÉNARD L. & ABRY C. (2011). Test of the movement expansion model : Anticipatory vowel lip protrusion and constriction in french and english speakers. *The Journal of the Acoustical Society of America*, **129**(1), 340–349.

PALIWAL K. K., AINSWORTH W. A. & LINDSAY D. (1983). A study of two-formant models for vowel identification. *Speech Communication*, **2**(4), 295–303.

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.

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. L. & MUNHALL K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, **1**(4), 333–382.

STEVENS K. N. (1993). Models for the production and acoustics of stop consonants. *Speech communication*, **13**(3), 367–375.