Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

Synthèse de syllabes avec un modèle de Maeda piloté par une représentation complexe

Frédéric Berthommier
Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
frederic.berthommier@gipsa-lab.grenoble-inp.fr

RÉSUMÉ

Un modèle mathématique est construit sur une notion de coordination des articulateurs à partir d'une représentation bidimensionnelle complexe. Les voyelles sont représentées par des positions en bordure du cercle unité, et pour le modèle de Maeda, les paramètres articulatoires sont générés avec une fonction de coordination facile à configurer. Les consonnes plosives /bdg/ sont encodées de la même manière, mais pour produire des syllabes, le graphe reliant les positions phonétiques distingue les arcs vocaliques et les arcs consonantiques. Un flux de paramètres articulatoires est dérivé par application sélective de la fonction de coordination. Les contributions de deux groupes d'articulateurs sont ainsi superposées et synchronisées pour piloter le modèle de Maeda et obtenir la synthèse de trajectoires formantiques. Ce modèle possède un schéma déterministe similaire à celui de la phonologie articulatoire, mais de nombreuses simplifications sont opérées.

ABSTRACT

Syllable synthesis with a Maeda model driven by a complex representation

A mathematical model is built on a notion of articulator coordination based on a complex two-dimensional representation. Vowels are represented by positions at the edge of the unit circle, and for the Maeda model, articulatory parameters are generated with an easy-to-configure coordination function. The plosive consonants /bdg/ are encoded in the same way, but to produce syllables, the graph linking phonetic positions distinguishes between vowel arcs and consonant arcs. A stream of articulatory parameters is derived by selective application of the coordination function. The contributions of two groups of articulators are thus superimposed and synchronised to drive a Maeda model and obtain the synthesis of formantic trajectories. This model has a deterministic scheme similar to that of articulatory phonology, but many simplifications are made.

MOTS-CLÉS : synthèse articulatoire, relation articulatori-acoustique, voyelles, consonnes, syllabes, coordination des articulateurs, phonologie articulatoire.

KEYWORDS: articulatory synthesis, articulatory-acoustic relationship, vowels, consonants, syllables, coordination of articulators, articulatory phonology.

1 Introduction

Lors de la production de la parole, les mouvements de la langue, de la mâchoire et des lèvres sont pseudo-périodiques. Prosaïquement, le geste associé à /aiua/ est une rotation de la langue, d'abord vers le haut et l'avant pour /ai/, puis vers l'arrière avec /iu/ et enfin vers l'arrière et vers le bas pour clore le cycle avec /ua/. Ici, la cyclicité de la position des formants est observable avec un simple

spectrogramme et on en déduit une dépendance entre gestes de la langue et position des formants. Des *nomogrammes* synthétisant la relation articulatori-acoustique ont été construits avec le tube de Fant (Badin *et al.*, 1990) en faisant varier la position et l'ouverture d'une constriction représentant la langue ainsi que l'ouverture à une extrémité. Malheureusement, cette relation est très complexe et elle ne peut pas être modélisée mathématiquement. Dans un spectrogramme ordinaire, les modulations formantiques sont bien présentes, mais on ne peut pas en déduire facilement les mouvements du tractus vocal à cause de la non-linéarité de la relation et surtout de la multiplicité des configurations articulatoires produisant un ensemble de formants donné. De fait, ceci limite considérablement l'étude de la relation entre perception et production de la parole.

Les consonnes sont engendrées par des constrictions du conduit vocal disposées en demi-cercle depuis les lèvres pour le /b/, en position coronale pour le /d/, palatale ou vélaire pour /g/ jusqu'aux lieux d'articulation pharyngal et épiglottal (Schwartz *et al.*, 2012). La Task Dynamics (TD) avec (Nam *et al.*, 2004) encode explicitement les lieux d'articulation des consonnes comme des angles et il est suggéré que des positions angulaires discrètes émergent aussi pour les voyelles avec le modèle de Maeda (Gaines *et al.*, 2021). Pour réaliser les constrictions de concert avec la production des voyelles, la langue effectue des mouvements qui sont potentiellement planifiables dans une représentation complexe par le biais d'un codage angulaire. L'intérêt de celle-ci semble secondaire si on ne coordonne pas tous les articulateurs (la langue et les lèvres) au cours du temps. Coordination et synchronisation sont les clefs de la planification syllabique selon (Xu, 2017, 2020) et nous introduisons ces deux mécanismes en faisant appel à une représentation complexe.

C'est à rebours des approches fondées sur les réseaux neuronaux que nous proposons une méthode de synthèse des syllabes basée sur des régularités du mécanisme de production exprimables dans des espaces de très faible dimension. Le modèle de Maeda (Maeda, 1979, 1990) est construit à partir de coupes sagittales et la factorisation de ces données en termes de commandes articulatoires favorise cette approche. Les aspects cognitifs et la notion d'apprentissage sont relégués au second plan pour privilégier l'identification de ces régularités. En établissant une continuité entre l'espace vocalique et la structure syllabique de la parole avec l'appui de quelques travaux, le but n'est pas de construire des applications, mais d'offrir un éclairage sur la structure profonde de la parole, complémentaire de celui apporté par les réseaux neuronaux (Dupoux, 2018).

2 La construction de l'espace vocalique

Préalablement, nous avons montré que le DRM, constitué de huit tubes (Mrayati *et al.*, 1988), est très approprié pour construire une *bijection* entre le cercle unité et l'espace vocalique F1-F2 en exploitant les relations décrites entre variations formantiques et variations du diamètre de chacun des tubes (Berthommier, 2021). Ce modèle n'est pas anatomique et les diamètres des tubes représentent grossièrement, par régions distinctives, la fonction d'aire obtenue à partir d'une coupe sagittale du conduit vocal. Nous avons remplacé la méthode itérative de couverture de l'espace vocalique partant du tube neutre par une détermination directe de la périphérie, où sont situées les voyelles cardinales. Une fonction de coordination du diamètre des tubes est définie à partir des diamètres fixés pour les trois voyelles cardinales /aiu/ (Berthommier, 2021). Cette fonction relie une position dans le cercle unité avec les diamètres des tubes en les corrélant. Pour aller vers les modèles articulatoires, une comparaison a été réalisée entre le DRM et le modèle de Fant pour la construction du triangle vocalique. Dans les deux cas, la détermination de la fonction de coordination n'est pas triviale, car

on associe une représentation trigonométrique à la géométrie linéaire d'un système de tubes. En revanche, avec un modèle de Maeda, ici représenté par VLAM (Boë & Maeda, 1998), nous constatons que cette fonction est déductible sans calcul. Ceci atteste la compatibilité entre un modèle dérivé d'une statistique de configurations articulatoires réelles projetées sur une grille semi-polaire et la fonction de coordination qui *génère* des configurations à partir d'une représentation complexe.

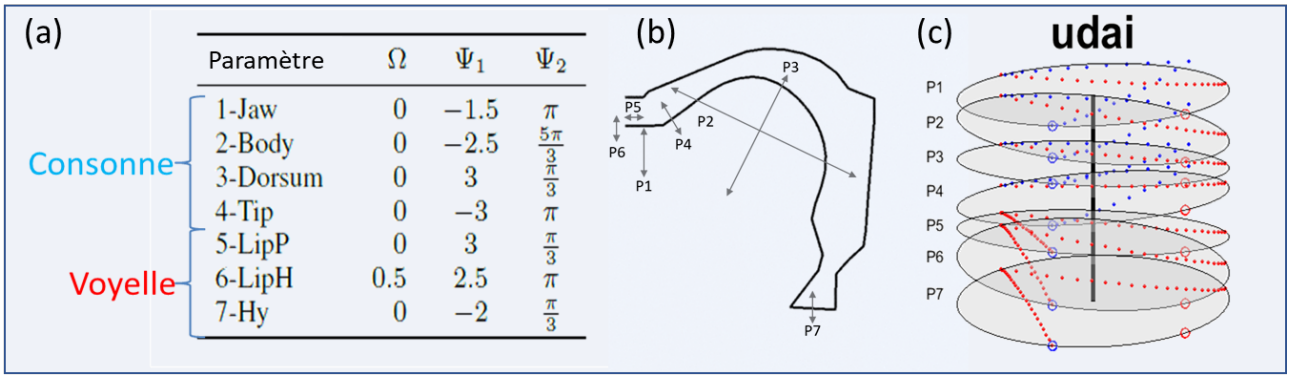

FIGURE 1 – (a) Configuration du modèle de Maeda. (b) Action des 7 paramètres articulatoires. (c) Représentation des trajectoires paramétriques en pile pour la syllabe /udai/ où les valeurs apparaissent verticalement. Les trajectoires vocaliques /ua/ et /ai/ sont en rouge et en bleu pour la consonne /d/.

La configuration de la fonction de coordination Fig 1a est basée sur la moyenne Ω et l'étendue Ψ_1 de chaque paramètre (données a priori de VLAM) plus un angle à déterminer Ψ_2. La valeur de chaque paramètre articulatoire $P_i, i = 1..7$ est calculée indépendamment pour un point donné (ρ_V, θ_V) du domaine complexe. La coordination entre P_i est assurée par le produit du même complexe conjugué $\rho_V e^{-j\theta_V}$ avec chaque valeur complexe Ψ_i du modèle :

$$P_i - \Omega_i = Re\left[\Psi_i \rho_V e^{-j\theta_V}\right]$$
$$\boldsymbol{P - \Omega} = Re\left[\boldsymbol{\Psi}\rho_V e^{-j\theta_V}\right] = \rho_V \boldsymbol{\Psi_1}\, cos(\boldsymbol{\Psi_2} - \theta_V) \quad (1)$$

Remarquons que c'est une simple cosinusoïde. Les angles inconnus $\boldsymbol{\Psi_2}$ sont fixés afin que les angles des voyelles /iau/ sur le cercle unité ($\rho_V = 1$) soient $\theta_V = \{\frac{5\pi}{3}, \pi, \frac{\pi}{3}\}$. Pour cela, on assigne l'une des voyelles /iau/ à chaque articulateur de telle sorte que $P_i - \Omega_i = \Psi_{1i}$ pour cet articulateur lorsque $\theta = \Psi_{2i}$. Le paramètre corps de la langue (Body) est aligné sur la réalisation du /i/, l'ouverture des lèvres (LipH), l'abaissement de la pointe (Tip) et de la mâchoire (Jaw) sur /a/ et l'allongement du conduit vocal (LipP, Hy) ainsi que la flexion de la langue (Dorsum) avec /u/ (Fig. 1b pour l'action des paramètres). Ces spécifications établissent une bijection entre le domaine du cercle unité Fig. 2a et une surface dans l'espace des formants F1-F2-F3 (dite surface vocalique Fig. 2c). La répartition angulaire d'autres voyelles cardinales est déductible par antisymétrie centrale. Par exemple, l'angle de /ɛ/ est égal à $\frac{\pi}{3} + \pi = \frac{4\pi}{3}$ avec comme conséquence un renversement de la valeur des paramètres articulatoires du /u/ par rapport à la valeur du neutre car $cos(x + \pi) = -cos(x)$. Les angles $\theta_V \in \{0, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \pi, \frac{4\pi}{3}, \frac{3\pi}{4}, \frac{5\pi}{3}\}$ sont associés respectivement aux voyelles /ɨ, u, o, ɔ, a, ɛ, e, i/. Celles-ci résultent de l'application de la fonction de coordination et elles ne nécessitent pas d'ajustement particulier. En revanche, nous avons retrouvé à l'écoute l'angle $\frac{11\pi}{6}$ pour la voyelle /y/ moins fréquente (Moran & McCloy, 2019). Les paramètres articulatoires du modèle de Maeda reflètent les commandes musculaires du tractus vocal essentiellement basées sur des relations *agonistes-antagonistes* (Kröger & Bekolay, 2022). L'antisymétrie centrale est une conséquence directe de la

structure anatomique associant forme semi-circulaire du tractus vocal et commandes musculaires. Cet ensemble est physiquement relié à la structure de l'espace vocalique (Schroeder, 1967; Mrayati *et al.*, 1988; Berthommier, 2021). Ces régularités sont prises en compte par la fonction de coordination, expression mathématique qui ancre l'espace vocalique dans l'anatomie du tractus vocal en établissant une continuité structurelle forte (Fig. 2).

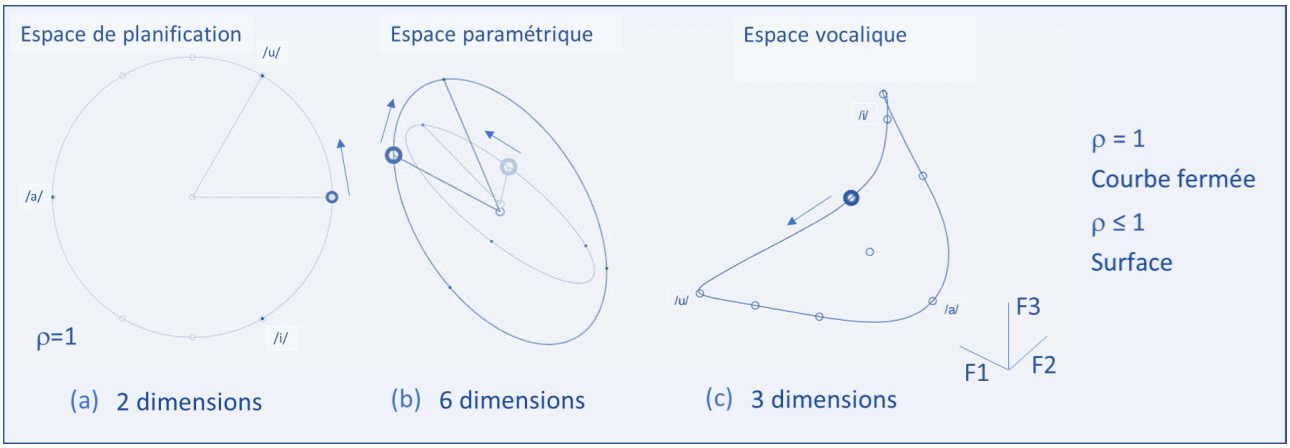

FIGURE 2 – Illustration de l'effet de la fonction de coordination. Un parcours entre 0 et 2π dans le sens trigonométrique dans (a) l'espace de planification où les voyelles cardinales sont codées par des angles se traduit par des parcours *elliptiques* dans (b) l'espace paramétrique. Le point de départ $\odot$ et le secteur $\{0, \frac{\pi}{3}\}$ sont figurés. Ici, les paramètres articulatoires sont séparés en 2 groupes visibles {Jaw, LipP, LipH} et non visibles {Body, Dorsum, Apex} pour figurer les corrélations induites par la fonction de coordination. La représentation du paramètre Hy est omise bien qu'il soit pris en compte. (c) Ces mêmes paramètres entraînent un parcours de la périphérie d'une surface vocalique exprimée par la position des formants F1-F2-F3. On retrouve dans l'ordre les voyelles cardinales /u, o, ɔ, a, ɛ, e, i/ notées ∘ en périphérie de cette surface. La fonction de coordination établit une bijection qui associe tout point du domaine complexe situé à l'intérieur du cercle unité avec un vecteur de paramètres inclus dans les deux ellipses et un point situé sur la surface vocalique, calculé avec (Badin & Fant, 1984) à partir de ce même vecteur.

La relation qu'entretiennent les deux surfaces Fig. 2a-c avec l'espace paramétrique de dimension 6 est visualisée Fig. 2b à l'aide de deux ellipses placées dans un espace 3D. Chacune d'elles représente 3 paramètres visibles et non visibles avec deux points qui sont en rotation lorsque l'on parcourt le cercle unité Fig. 2a. À droite Fig. 2c, la périphérie de la surface vocalique est décrite de façon concomitante. Si l'on se dirige vers le centre du cercle unité, les paramètres se rapprocheront de ceux de la voyelle neutre, et les formants des valeurs (f1n, f2n, f3n) en suivant cette surface. Cela réduit considérablement le calcul de la trajectoire d'une diphtongue dans l'espace paramétrique à sept dimensions. Sans une telle simplification et dans un espace plus grand, la synthèse des diphtongues avec VocalTractLab se montre laborieuse (Xu *et al.*, 2023). (Story *et al.*, 2018) construisent une bijection entre une paramétrisation 2D des formes du conduit vocal et F1-F2. Il s'agit du résultat le plus proche du nôtre, mais il est obtenu avec un modèle à tubes et l'espace vocalique résultant présente des défauts notables. Ici, le *pointage* exercé depuis le domaine complexe vers la surface F1-F2-F3 est très régulier et les trajectoires dessinées dans l'un apparaissent peu déformées sur l'autre.

3 La représentation des consonnes et leur coproduction

Nous proposons d'étendre l'usage de la fonction de coordination aux consonnes en les plaçant dans le même référentiel. Selon (Öhman, 1966) avec des syllabes V1CV2, les consonnes plosives perturbent la trajectoire de F2 marquant une transition entre les voyelles V1 et V2. La modélisation la coproduction voyelles/consonne (Öhman, 1967) est basée sur une pondération qui ne *sépare pas* les parties du conduit vocal pour les affecter à la constriction ou à cette transition vocalique. Par contre, la structure du DRM est très appropriée pour effectuer une telle séparation. L'effet de perturbation de F2 est obtenu en sélectionnant le tube du lieu d'articulation pour effectuer la constriction et en laissant les autres tubes contribuer à la transition vocalique (Carré & Chennoukh, 1995). Fondé sur un principe équivalent, TubeTalker (Story, 2009, 2013) utilise des formes anatomiques issues d'IRMs, mais le pincement des tubes dans la région orale n'est pas anatomique. On obtient les plosives /bdg/ sur le plan acoustique avec une trace correcte de la coarticulation (Story & Bunton, 2021) mais cela ne renseigne pas bien sur la relation articulatori-acoustique et son contrôle. Cependant, la propriété de séparabilité des paramètres mise en évidence avec ces modèles est intéressante car elle simplifie la coproduction envisagée par (Öhman, 1967). Comme amélioration aboutissant à un résultat audible, (Birkholz, 2013) dispose pour chaque consonne de 3 cibles articulatoires pour /aiu/ et il réalise une moyenne pondérée pour les autres voyelles. Pour planifier la synthèse de syllabes, VocalTractLab possède un niveau vocalique spécifique dans son panel de planification gestuelle. Mais n'y a pas de séparabilité comme avec les modèles à tubes. La TD adopte une solution flexible en coordonnant et en pondérant la participation des articulateurs à une tâche de constriction. La synthèse d'un sous-ensemble de VCVs avec un modèle composite suivant ses principes et à l'aide du modèle de Maeda a été décrite récemment (Alexander *et al.*, 2019). La propriété de séparabilité des articulateurs n'est pas ou peu appliquée en dehors des modèles à tubes.

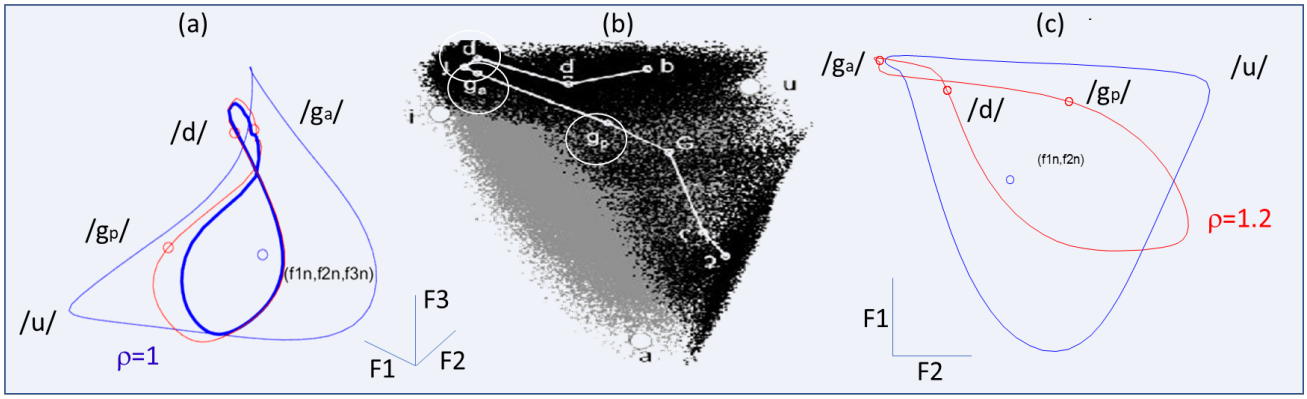

FIGURE 3 – Répartition des consonnes plosives relative à la surface vocalique. La marge périphérique du cercle unité est parcourue tout en bloquant les 3 paramètres {LipH, LipP, Hy} en position neutre. (a) Dans l'espace F1-F2-F3 (b) Référence F1-F2 (Schwartz *et al.*, 2012) (c) Dans l'espace F1-F2.

Les consonnes plosives /bdg/ ont des caractéristiques dynamiques complexes et elles sont présentes en grande proportion dans les langues du Monde, ce qui motive leur choix. En phonologie articulatoire (Saltzman & Munhall, 1989), la notion de tâche pour la réalisation de constrictions *privilégie* implicitement les consonnes par rapport aux voyelles associées à des formes globales du tractus vocal. Cependant, une analogie de production entre consonnes et voyelles est inhérente au lieu de

constriction (Gaines *et al.*, 2021). Une statistique exhaustive des lieux de constriction a été réalisée avec VLAM par (Schwartz *et al.*, 2012). Par le biais d'un nomogramme, elle met en relation les lieux d'articulation des consonnes avec les formants F1-F2-F3 produits au relâchement de la constriction.

Avec la fonction de coordination, les résultats de (Schwartz *et al.*, 2012) sont reproduits sans recourir à des simulations de Monte Carlo Fig. 3b. Avec le même principe que pour les voyelles et en délimitant $1 \leq \rho_C \leq 1.2$ tout en parcourant le cercle unité, on ne sélectionne que les 4 paramètres contrôlant la langue et la mâchoire Fig. 1a-b, en fixant en position neutre les trois paramètres restants Fig. 3a-c. D'une part, cette sélection laisse libres trois paramètres qui sont séparables pour effectuer une tâche de coproduction des voyelles. D'autre part, en augmentant les valeurs de ρ, une fermeture du conduit vocal, limitée par une rectification douce, est engendrée sur une partie du cercle unité. Consonnes et voyelles se retrouvent plongées dans une représentation commune. La procédure de découverte proposée comme analogue au babillage (Schwartz *et al.*, 2012) est également simplifiée puisque les angles attribuables aux consonnes /dg/ (respectivement $\theta_C \cong \{\frac{3\pi}{2}, -\frac{\pi}{12}, \frac{\pi}{3}\}$ avec deux positions, antérieure et postérieure, pour /g/) sont très proches de ceux des voyelles cardinales /e, i, u/. Avec un tel a priori, leur découverte ne demande que peu de tâtonnements et nous restons dans le cadre de l'émergence structurale mise en évidence pour les voyelles cardinales. Les phonèmes sont replacés dans un cadre intrinsèquement structuré et le conflit inhérent au choix de contrôle des lieux de constriction plutôt que celui de la forme globale du tractus vocal est résolu naturellement.

4 La synthèse des syllabes

Pour définir une coproduction fondée sur la séparabilité des articulateurs, nous proposons que la structure syllabique soit représentée par un graphe constitué de noeuds pour les phonèmes et d'arcs représentant des trajectoires entre chaque noeuds, planifiées dans le plan complexe. On distingue les trajectoires vocaliques et consonantiques entre deux points de rendez-vous avec une *synchronisation* (Xu, 2017, 2020). Les articulateurs associés à chacune des branches sont définis par un processus de sélection. La synchronisation est quant à elle obtenue en associant les arcs à des multiples entiers d'une période de référence T. La structure temporelle de la production est liée à l'enchaînement des gestes articulatoires qui forment des unités syllabiques de taille intermédiaire puis des mots par concaténation. Sur le plan temporel, nous faisons l'hypothèse que les segments ont une durée relativement constante à court terme mais que la structure superficielle reste pseudo-périodique.

La synthèse des syllabes CV découle de cette représentation des consonnes. En effet, les articulateurs qui ne sont pas affectés sont *libres* pour les transitions vocaliques. S'il est couramment admis que C et V ont un début synchrone, en phase selon la phonologie articulatoire (Browman & Goldstein, 1992), il faut spécifier l'anticipation de la forme des lèvres. Exemple classique, l'arrondissement des lèvres précède l'onset consonantique pour /Cu/ (Daniloff & Moll, 1968). Pour réaliser la coproduction CV et assurer la synchronisation des articulateurs, nous introduisons comme point d'ancrage une voyelle réduite (centralisée). Ici, on assigne à la trajectoire /u_r/ (réduit) vers /u/ les paramètres qui ne sont pas nécessaires à l'articulation de la consonne. Pour /Cu/, l'anticipation de la protrusion des lèvres apparaît avec le paramètre LipP.

Les trajectoires des voyelles et des consonnes sont planifiées dans le plan complexe Fig. 4a en formant des arcs entre 2 points (ρ_1, θ_1) et (ρ_2, θ_2) :

$$z(t) = (1 - \rho(t))\, \rho_1\, e^{j\theta_1} + \rho(t)\, \rho_2\, e^{j(\theta_2 + \frac{\nu}{K}\theta(t))} \tag{2}$$

où t varie entre 0 et nT, $\rho(t) = cos(\frac{\theta(t)}{2})$ détermine le profil de vitesse et $\nu = \pm 1$ et K sont des paramètres de forme de la trajectoire (voir aussi (Berthommier, 2023)). Lorsque K est grand ($K = 30$ pour les arcs de voyelles et $K = 10$ pour les arcs de consonnes), les trajectoires deviennent plus rectilignes dans le plan complexe (voir Fig. 4a).

À chaque instant t, l'ensemble des paramètres est coordonné avec l'équation 1 et la trajectoire paramétrique est obtenue par enchaînement de périodes de temps de durée nT. Au cours de chaque période, les trajectoires des paramètres sont la partie réelle du produit du vecteur colonne complexe $\mathbf{\Psi}$ qui représente le modèle articulatoire, et du vecteur ligne complexe $\bar{z}(t)$ issu de la planification. Il en résulte des matrices de dimension $7 * nT$ qui sont concaténées :

$$\begin{aligned} \boldsymbol{P}(t) - \mathbf{\Omega} = Re\left[\mathbf{\Psi}\bar{z}(t)\right] = (1-\rho(t))\,\rho_1\,\mathbf{\Psi_1}\,cos(\mathbf{\Psi_2} - \theta_1) \\ + \rho(t)\,\rho_2\,\mathbf{\Psi_1}\,cos(\mathbf{\Psi_2} - \theta_2 - \frac{\nu}{K}\theta(t)) \end{aligned} \tag{3}$$

Il reste à instancier le processus de sélection reposant sur la séparabilité des articulateurs. Les trajectoires des segments vocaliques et des pauses sont définies par l'équation précédente seule, tandis que la *superposition* des trajectoires vocaliques et consonantiques nécessite une coordination séparée. Notons que la fonction de coordination s'applique alors en chaque t sur les deux trajectoires $\bar{z}_v$ et $\bar{z}_c$. La coarticulation entre voyelles et consonnes est produite par la superposition de ces 2 branches ayant la même durée nT ainsi que les mêmes points de départ et d'arrivée qui sont des voyelles éventuellement réduites (i.e. VCVr pour VC ou VrCV pour CV) :

$$\boldsymbol{P}(t) - \mathbf{\Omega} = Re\left[S_v.\mathbf{\Psi}\bar{z}_v(t) + S_c.\mathbf{\Psi}\bar{z}_c(t)\right] \tag{4}$$

où S_v et S_c sont les deux vecteurs de sélection exclusifs composés de zéros et de uns pour les articulateurs sélectionnés avec $S_v + S_c = \mathbf{1}$ (un vecteur colonne de 7*1). Avec un produit de Hadamard, les composantes non sélectionnées du vecteur complexe $\mathbf{\Psi}$ sont annulées. La composition de ces vecteurs dépend de la (ou des) consonnes. Nous avons vu que pour /dg/, la mâchoire et les 3 paramètres de la langue sont nécessaires tandis que pour /b/ la sélection de l'ouverture des lèvres est complémentée par la mâchoire et le corps de la langue. D'autres détails concernant la sélection des articulateurs et la planification des clusters consonantiques sont disponibles (Berthommier, 2023).

Dans l'exemple /udai/ Fig. 4a, la trajectoire /ua/ est perturbée par /d/. Durant cette coproduction, deux ensembles d'articulateurs notés Fig. 1a sont coordonnés séparément. Tandis que le premier ensemble (consonne) suit la trajectoire de /u/ vers /d/ puis vers /a/ en deux périodes T, le second (voyelle) va de /u/ à /a/ de façon synchronisée. Les trajectoires visibles dans la pile paramétrique Fig. 1c sont différentes et leurs effets sur l'unique trajectoire formantique se superposent. Mathématiquement, la conséquence de la superposition est d'augmenter la dimension de la planification de 2 (1 nombre complexe) à 4 (2 nombres complexes). Cette augmentation est observable dans l'espace des formants Fig. 4c : alors que la trajectoire de la diphtongue /ai/ reste sur la surface vocalique, la trajectoire des segments superposés /uda/ Fig. 4c en sort. La coarticulation du /d/ est réalisée par une attraction du locus situé au point de rebroussement de la trajectoire, vers la branche /ua/ portée par les lèvres et le larynx. Notons enfin que si la méthode est adaptée pour planifier Fig. 4a les trajectoires des formants Fig. 4c-d et offrir des animations du modèle de Maeda, ce qui concerne le contrôle effectif des articulateurs reste très hypothétique. Comme nous l'avons indiqué, le champ d'application est celui de la phonologie articulatoire où il peut apporter une série de simplifications, en particulier sur la planification des syllabes. Pour la reproduction de gestes articulatoires, la TD s'appuie sur des enregistrements de données EMA correspondant aux variables articulatoires du modèle, tandis qu'ici une telle référence n'existe pas. En revanche, les paramètres de Maeda issus d'une factorisation par

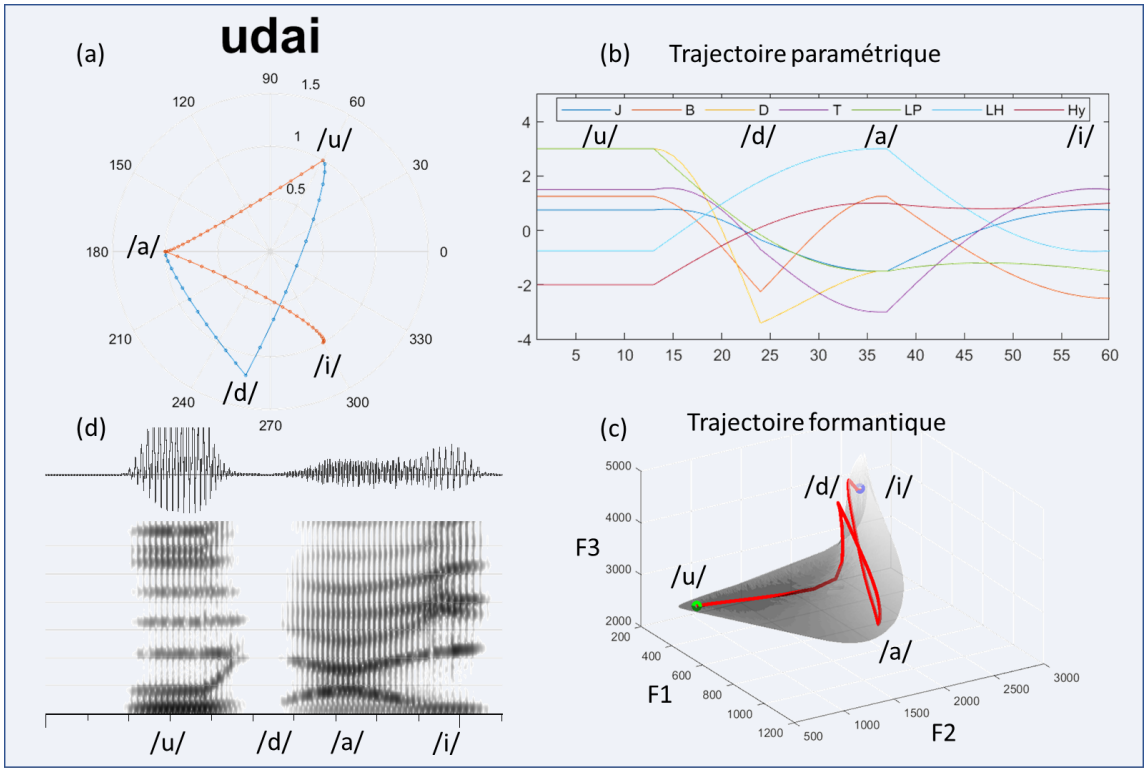

FIGURE 4 – Synthèse de la syllabe /udai/. (a) Graphe de planification syllabique avec en rouge les trajectoires vocaliques et en bleu les consonantiques. Elles relient les phonèmes situés en position périphérique du cercle unité. (b) Flux de 7 paramètres articulatoires évalués par périodes T=120 ms concaténées sans lissage (ici 5T). (c) Trajectoire des 3 premiers formants mise en rapport avec la surface vocalique. (d) Enveloppe et spectrogramme de sortie. Une modulation construite à partir de la structure syllabique et pour les consonnes plosives /bdg/ est appliquée.

PCA guidée reflètent des groupes musculaires dont le contrôle serait séparable (Maeda & Honda, 1994; Kröger & Bekolay, 2022). Très hypothétiquement, on obtiendrait ici Fig. 4b une image de leurs modulations d'activité.

5 Conclusion

Avec un dessin équivalent, combinant une planification à l'échelle syllabique et une inférence des commandes articulatoires d'un modèle, les simplifications exercées par rapport l'AP/TD (Saltzman & Munhall, 1989; Browman & Goldstein, 1992) sont nombreuses. Les relations temporelles entre gestes articulatoires étant fixées au moment de la planification syllabique, le recours à une évaluation dynamique à l'aide d'oscillateurs couplés n'est pas nécessaire (Xu, 2020). Ceci entraîne des différences de description de la structure syllabique. En ce qui concerne la TD, nous avons vu qu'il n'est pas nécessaire de réaliser une transformation pour déduire des paramètres de commande d'un modèle articulatoire à partir de variables articulatoires liées aux tâches de constriction. En effet, ces paramètres sont obtenus directement, et les trajectoires vocaliques sont pointées spatialement depuis la représentation complexe, reliant production et formes acoustiques résultantes de façon cohérente.

Références

ALEXANDER R., SORENSEN T., TOUTIOS A. & NARAYANAN S. (2019). A modular architecture for articulatory synthesis from gestural specification. *The Journal of the Acoustical Society of America*, **146**(6), 4458–4471. DOI : 10.1121/1.5139413.

BADIN P. & FANT G. (1984). *Notes on vocal tract computations*. Stl- qpsr 2-3/1984, Royal Institute of Technology, Stockholm, Sweden.

BADIN P., PERRIER P., BOË L. & ABRY C. (1990). Vocalic nomograms : Acoustic and articulatory considerations upon formant convergences. *The Journal of the Acoustical Society of America*, **87**(3), 1290–1300. DOI : 10.1121/1.398804.

BERTHOMMIER F. (2021). A mathematical model of the vowel space. DOI : 10.48550/arXiv.2111.00868.

BERTHOMMIER F. (2023). Why can big.bi be changed to bi.gbi ? a mathematical model of syllabification and articulatory synthesis. DOI : 10.48550/arXiv.2307.02299.

BIRKHOLZ P. (2013). Modeling consonant-vowel coarticulation for articulatory speech synthesis. *PLOS ONE*, **8**(4), 1–17. DOI : 10.1371/journal.pone.0060603.

BOË L.-J. & MAEDA S. (1998). Modélisation de la croissance du conduit vocal. journées d'Études linguistiques. In *La voyelle dans tous ses états*, p. 98–105.

BROWMAN C. P. & GOLDSTEIN L. M. (1992). Articulatory phonology : An overview. *Phonetica*, **49**, 155–180. DOI : 10.1159/000261913.

CARRÉ R. & CHENNOUKH S. (1995). Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures. *Journal of Phonetics*, **23**(1), 231–241. DOI : 10.1016/S0095-4470(95)80045-X.

DANILOFF R. & MOLL K. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, **11**(4), 707–721. DOI : 10.1044/jshr.1104.707.

DUPOUX E. (2018). Cognitive science in the era of artificial intelligence : A roadmap for reverse-engineering the infant language-learner. *Cognition*, **173**, 43–59. DOI : 10.1016/j.cognition.2017.11.008.

GAINES J. L., KIM K. S., PARRELL B., RAMANARAYANAN V., NAGARAJAN S. S. & HOUDE J. F. (2021). Discrete constriction locations describe a comprehensive range of vocal tract shapes in the Maeda model. *JASA Express Letters*, **1**(12), 124402. DOI : 10.1121/10.0009058.

KRÖGER B. J. & BEKOLAY T. (2022). Producing syllables : motor planning, motor programming and execution. In O. NIEBUHR, M. S. LUNDMARK & H. WESTON, Éds., *Studientexte zur Sprachkommunikation : Elektronische Sprachsignalverarbeitung 2022*, p. 1–8 : TUDpress, Dresden.

MAEDA S. (1979). Un modèle articulatoire de la langue avec des composantes linéaires. In *Actes 10 èmes Journées d'Etude sur la Parole*, p. 152–162.

MAEDA S. (1990). Compensatory articulation during speech : Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In D. J. HARDCASTLE & A. MARCHAL, Éds., *Speech Production and Speech Modelling*, NATO ASI Series, p. 131–149. Springer Netherlands, Dordrecht. DOI : 10.1007/978-94-009-2037-8_6.

MAEDA S. & HONDA K. (1994). From emg to formant patterns of vowels : The implication of vowel spaces. *Phonetica*, **51**(1-3), 17–29. DOI : 10.1159/000261955.

MORAN S. & MCCLOY D., Éds. (2019). *PHOIBLE 2.0*. Jena : Max Planck Institute for the Science of Human History.

MRAYATI M., CARRÉ R. & GUÉRIN B. (1988). Distinctive regions and modes : A new theory of speech production. *Speech Commun.*, **7**(3), 257–286. DOI : 10.1016/0167-6393(88)90073-8.

NAM H., GOLDSTEIN L., SALTZMAN E. & BYRD D. (2004). Tada : An enhanced, portable task dynamics model in matlab. *The Journal of the Acoustical Society of America*, **115**(5), 2430–2430. DOI : 10.1121/1.4781490.

SALTZMAN E. L. & MUNHALL K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, **1**(4), 333–382. DOI : 10.1207/s15326969eco0104_2.

SCHROEDER M. R. (1967). Determination of the geometry of the human vocal tract by acoustic measurements. *The Journal of the Acoustical Society of America*, **41**(4B), 1002–1010. DOI : 10.1121/1.1910429.

SCHWARTZ J.-L., BOË L.-J., BADIN P. & SAWALLIS T. R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech : On the universality of the labial–coronal–velar stop series. *Journal of Phonetics*, **40**(1), 20–36. DOI : 10.1016/j.wocn.2011.10.004.

STORY B. H. (2009). Vowel and consonant contributions to vocal tract shape. *The Journal of the Acoustical Society of America*, **126**(2), 825–836. DOI : 10.1121/1.3158816.

STORY B. H. (2013). Phrase-level speech simulation with an airway modulation model of speech production. *Computer Speech and Language*, **27**(4), 989–1010. DOI : 10.1016/j.csl.2012.10.005.

STORY B. H. & BUNTON K. (2021). Identification of voiced stop consonants produced by acoustically driven vocal tract modulations. *JASA Express Letters*, **1**(8), 085203. DOI : 10.1121/10.0005917.

STORY B. H., VORPERIAN H. K., BUNTON K. & DURTSCHI R. B. (2018). An age-dependent vocal tract model for males and females based on anatomic measurements. *The Journal of the Acoustical Society of America*, **143**(5), 3079–3102. DOI : 10.1121/1.5038264.

XU A., VAN NIEKERK D., KRUG P., PROM-ON S., BIRKHOLZ P. & XU Y. (2023). Computational models for articulatory learning of english diphthongs : One dynamic target vs. two static targets. In *Proc. of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, p. 4140–4144.

XU Y. (2017). Syllable as a synchronization mechanism. In *Proceedings of 8th Tutorial and Research Workshop on Experimental Linguistics*, p. 9–12. DOI : 10.36505/ExLing-2017/08/0003/000305.

XU Y. (2020). Syllable is a synchronization mechanism that makes human speech possible. DOI : 10.31234/osf.io/9v4hr.

ÖHMAN S. E. G. (1966). Coarticulation in vcv utterances : Spectrographic measurements. *The Journal of the Acoustical Society of America*, **39**(1), 151–168. DOI : 10.1121/1.1909864.

ÖHMAN S. E. G. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, **41**(2), 310–320. DOI : 10.1121/1.1910340.