

# A propos des degrés de liberté dans la coarticulation d'un locuteur français filmé 50 fois par seconde

VILAIN Anne, ABRY Christian, BADIN Pierre

Institut de la Communication Parlée, CNRS UPRESA 5009  
Université Stendhal, DU, 1180 avenue centrale  
BP 25, 38040 Grenoble cedex 9, France  
Tél. : 04 76 82 41 19 - Fax : 04 76 82 43 35  
e-mail : vilain,abry,badin@icp.inpg.fr - <http://www.icp.inpg.fr>

## ABSTRACT

Questions on speech production are here addressed on the basis of a 30-second X-ray film built on a corpus of VCV sequences uttered by a skilled French speaker. This high-quality observation, in addition to other previously recorded speakers, allows us to establish a coarticulation typology in relation to the degrees of freedom of the vocal tract. These data stand as very meaningful cues for the study of the economy and evolution of phonetic systems, starting from the emergence of coarticulation in babbling.

## 1. INTRODUCTION

Nous présentons ici un corpus de vues de profil du conduit vocal issues d'un film cinéradiographique, qui représente une base de données de grande valeur pour l'étude de la production de la parole, ne serait-ce que par sa rareté et sa qualité. Ce film a donné lieu à une modélisation articulatoire, le modèle Gentiane, à partir duquel il est désormais possible d'acquérir une quantité illimitée de données sans plus exposer le sujet aux rayons X.

Ces données et simulations nous permettront d'établir les contraintes de production des segments en séquences et ceci de façon à pouvoir les relier à l'organisation et à l'évolution des systèmes sonores.

Dans ce sens, nous avons appliqué ici-même la notion — venue de la robotique — de degrés de liberté en excès à un *plant* sous *contrôle*, afin d'en retirer plusieurs bénéfices : dégager les degrés de liberté de notre locuteur pour élaborer le *plant* Gentiane (section 2); observer l'exploitation de différents degrés de liberté dans la coarticulation de types articulatoire-acoustiques majeurs (section 3); mettre en évidence l'émergence de nouveaux degrés de liberté au cours des processus évolutifs de la coarticulation (section 4); tester enfin une théorie de l'ontogénèse de la parole, en vue de modéliser l'émergence pas à pas des degrés de liberté dans la coarticulation du babillage canonique (section 5).

Le modèle Gentiane s'inscrit dans une modélisation "orientée sujet". Cette approche *non normalisatrice* nous permet de modéliser la génération des premiers pas de la parole, soit la morphogénèse du babillage, sur plusieurs "robots parlants". Il devient alors possible de mettre en évidence, à la fois : (i) une invariance de commande centrale cyclique implémentée périphériquement sur l'articulateur porteur mandibulaire; (ii) une variabilité

radicale dans les produits résultants, s'expliquant par la conformation du sujet et les degrés de liberté de ses articulateurs, en fonction de sa stratégie individuelle. Se trouvent ainsi modélisés les premiers pas de la coarticulation, pour ainsi dire *ab ovo*.

## 2. MÉTHODE, DONNÉES ET MODÈLE : GENTIANE

### 2.1 Dispositif expérimental

Ce corpus cinéradiographique a été réalisé pour nous à l'hôpital Schiltigheim de Strasbourg, grâce à la collaboration de l'Institut de Phonétique de Strasbourg, avec la participation de Gilbert Brock, Agnès Hennel, Tahar Lallouache, Péla Simon et Jean-Pierre Zerling.

**Procédure** Le corpus a été enregistré simultanément en cinéradiographie par une caméra 35 mm avec un émetteur de rayons X et un amplificateur de radiances, en vidéo par une caméra et en audio par un DAT, avec des repères permettant une resynchronisation ultérieure de ces 3 enregistrements. Quatre pastilles de plomb ont été utilisées comme points de repère supplémentaires pour les images : l'une était collée sur le menton du sujet et les trois autres sur le dos, la lame et la pointe de la langue. Ces plombs apportent des informations précises et précieuses, en particulier sur les coordonnées de l'apex et la cavité sub-linguale.

**Corpus** Le corpus enregistré est constitué de combinaisons de séquences Voyelle-Consonne-Voyelle avec pour Consonnes [b, d, g, ʒ, v] et pour Voyelles [i, y, u, a]. Il a été établi de façon à former la "partition phonétique" qui permette, en un minimum de temps, d'explorer au maximum l'ensemble des degrés de liberté du conduit vocal.

**Locuteur** Le corpus de base a été produit par un locuteur unique qui, étant à la fois phonéticien et chanteur, possède une intelligence des tâches et une maîtrise exceptionnelle de son conduit vocal, pour bien mettre en évidence l'espace de travail de la parole, ceci en un temps record (30 s) afin de minimiser son exposition aux rayons X.

### 2.2. Traitement des données

**Expertise des tracés** Les images cinéradiographiques ont été projetées sur papier en chambre noire à l'aide d'un agrandisseur photo; puis les contours du conduit vocal ont été tracés au crayon à la main en taille réelle, pour obtenir

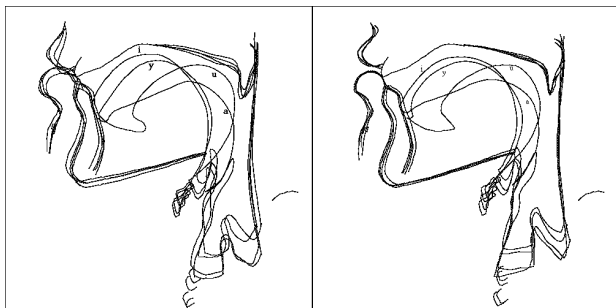
une banque de 1282 vues médio-sagittales. Il n'est actuellement pas envisageable de détecter automatiquement les contours d'une image radiographique, en raison de la superposition des ombres des différents organes. Le repérage des contours nécessite de prendre continuellement des décisions en fonction de la bonne connaissance des événements du conduit vocal. Seul un expert humain peut actuellement assumer ce rôle.

**Numérisation** Après scannérisation, chacun des sous-contours qui constituent le conduit vocal a été numérisé séparément de façon à permettre des mesures automatiques sur chaque articulateur. A chaque image du film cinéradiographique a ensuite été appliquée une grille de mesure bipolaire composée de 27 lignes. Les intersections des contours du conduit vocal avec chaque ligne de la grille correspondent à une série de 27 mesures qui caractérisent chaque configuration articuloire, et fournissent les données destinées à l'élaboration d'un modèle articuloire statistique.

### 2.3 Modélisation anthropomorphique individuelle

Selon nous, l'étude du polymorphisme des productions phonétiques sera rendue possible par la diversification de modèles articuloires "orientés sujet" — soient des modèles anthropomorphiques individualisés — établis selon des méthodologies comparables. C'est pourquoi nous avons développé le modèle Gentiane [Vil97], à partir des données cinéradiographiques dont nous disposons. C'est un modèle à 8 degrés de liberté (1 pour la mandibule, 2 aux lèvres, 4 sur la langue, 1 pour la hauteur du larynx), extraits à partir d'une analyse statistique inspirée de Maeda [Mae89]. Outre le modèle de ce dernier, cette analyse a déjà éprouvée par nous pour le modèle Bergame [Bad95], élaboré à partir d'un autre locuteur. Grâce à Gentiane, nos données pourront être complétées à volonté, sans autre enregistrement cinéradiographique, ceci par une méthode de reconstruction du contour de la langue à partir de quelques points acquis par articulographie électro-magnétique sur le même sujet [Bad97].

## 3 DEGRÉS DE LIBERTÉ DANS LA COARTICULATION DES CONSONNES



**Figure1** : Réalisations de [b] et [v] en contextes [i,y,u,a]

Les réalisations de [b,d,g,ʒ,v] en divers contextes vocaliques font apparaître un certain nombre de contraintes et de degrés de liberté caractéristiques de chaque consonne. Le lieu et le mode de constriction impliquent certaines

contraintes, tandis que les articulateurs non recrutés restent mobilisés par la coarticulation avec le contexte vocalique. De fait les articulations vocaliques mobilisent essentiellement la langue. Les articulations consonantiques nécessitant une constriction autre que linguale permettent donc une plus grande préservation de la position vocalique de la langue : c'est le cas du [b] et [v]. Cependant ces deux articulations labiales amènent deux types de comportement distincts.

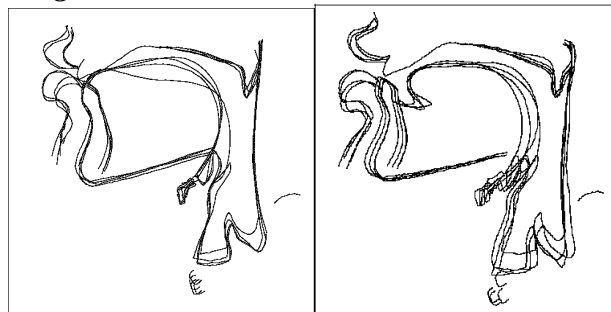
### 3.1 Occlusion labiale : une constriction simple et "perméable"

Le [b] présente une constriction dont la seule contrainte semble être qu'elle soit bilabiale, les autres articulateurs restant parfaitement libres. L'espace de la rencontre des deux lèvres possède donc une certaine variabilité. Les lèvres sont même suffisamment libres pour accompagner la protrusion de [y, u], ce qui modifie le lieu de l'occlusion (figure 1).

### 3.2 Constriction labio-dentale : une précision "pointilleuse" et "gratuite" de la lèvre inférieure

Le [v] quant à lui présente chez notre locuteur un comportement extrêmement stable, puisque la position de la lèvre inférieure, soit la constriction labio-dentale ne varie absolument pas quel que soit le contexte (figure 1). Cette précision n'est pourtant pas imposée par des facteurs articulatoire-acoustiques, puisqu'une variation du point de constriction n'apporterait aucune modification notable de la qualité acoustique de la fricative, comme c'est le cas de la légère variation de protrusion observée sur la lèvre supérieure. Il semble qu'il s'agisse ici d'une stratégie individuelle, puisque le locuteur PB, à partir duquel fut établi le modèle Bergame [Bad95], ne présentait pas ce type de précision. Ce dernier locuteur avait par contre la particularité de relever la lèvre supérieure au moment où la lèvre inférieure rejoignait l'incisive supérieure, sans doute pour l'acoustique de la friction.

### 3.3 Occlusion lamino-apicale : un "point" de contrainte acoustique sur la langue



**Figure2** : Réalisations de [d] et [ʒ] en contextes [i,y,u,a]

Le [d] nécessite un contact de la pointe de la langue contre les alvéoles pour obtenir une occlusion franche, c'est-à-dire qui utilise la précision de l'apex à l'occlusion et à la détente [Lab95]. Du fait du découplage de l'apex par

rapport à la masse de la langue (degré de liberté apical ou laminal), le corps de celle-ci garde une certaine flexibilité : la configuration propre à la voyelle est ainsi modifiée mais reste cependant visible (figure 2).

### 3.4 Constriction lamino-dorsale : une zone de contrainte acoustique encore plus forte

Le [ʒ] montre le même type de comportement, mais avec des contraintes plus fortes, puisque la constriction nécessaire à la turbulence voulue demande une grande précision, un chenal assez long sur la lame de la langue, et une tenue de cette position pendant la friction. De plus le [ʒ] optimise son acoustique par une protrusion des lèvres qui atténue, certes, la distinction entre voyelles arrondies et non-arrondies (plus visible avec le [d], figure 2), sans pourtant la mettre en danger perceptif [Abr80].

### 3.5 Occlusion vélaire : une coarticulation massive du corps de la langue

Le [g] est une consonne très plastique: elle présente une articulation dorsale, mobilisant la masse de la langue, qui est déjà recrutée pour la voyelle. La masse de la langue est peu mobile et possède seulement deux degrés de liberté (corps et dos), ce qui a pour conséquence une forte influence de la position vocalique sur le lieu de constriction consonantique. Les lieux d'occlusion du [g] des contextes [u] à [i] sont ainsi distants pour ce locuteur de près de 2 cm, allant de la région post-palatale à la zone vélaire (figure 3).

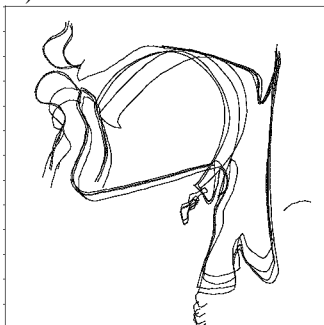


Figure3 : Réalisations de [g] en contextes [i,y,u,a]

## 4 DEGRÉS DE LIBERTÉ ET VIABILITÉ ARTICULATORI-ACOUSTIQUE

Une séquence dont la réalisation articulatoire est rendue difficile par le manque de degrés de liberté des articulateurs risque de se transformer selon les variations de la coarticulation. Ainsi, dans une séquence [gi], la consonne peut être modifiée par le contexte de voyelle antérieure haute au point de se palataliser (figure 4 : par comparaison avec [gu]). Puis, en raison du faible découplage des articulateurs, avoir une détente sous forme d'occlusive [ji] ou constrictive [ji] palatales. Mais alors une telle séquence — soit une semi-voyelle coarticulée avec une voyelle de même lieu — présente un piètre contraste acoustico-perceptif. De ce fait, ce genre de séquence est peu viable, et effectivement, elle est majoritairement évitée dans les systèmes des langues du

monde [Mad93] : "[ji] and [wu] are quite consistently avoided entirely or sharply lower in occurrence than [expected]".

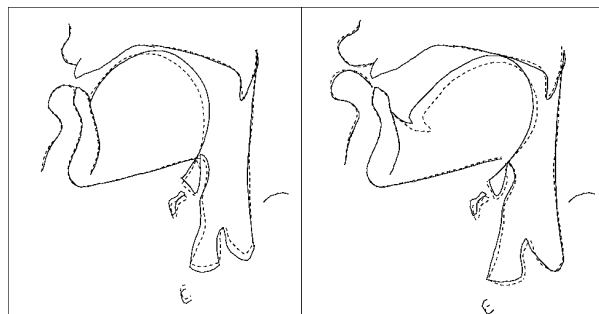


Figure4 : Transition Consonne-Voyelle dans les séquences [gi] et [gu] (La consonne est en trait plein, la voyelle en pointillés)

A ce moment de l'évolution phonétique, la viabilité contrastive de la séquence peut passer par un changement de régime articulatoire-acoustique. En effet si l'on modifie légèrement le degré de la constriction, un phénomène de turbulence peut apparaître, et ainsi une séquence du type [ʒi] puis [ʒi]. L'exemple du français est sur ce point instructif. On peut penser que les mots commençant par [ʒi] dans cette langue proviennent du stade [ji], affriqué en [ʒi], avant de donner [ʒi] (comme *girolle* de *girare*). Mais l'évolution de [j] latin, devant [i] (indirectement pour *gîte* de *jacere*) et devant les autres voyelles antérieures (cf. *jam* dans *déjà* ou *jamais*), et surtout devant les postérieures (cf. *joug* de *jugum*), cette évolution montre bien qu'il y a eu optimisation du contraste par passage à la friction. Celle-ci est obtenue en exploitant la non-linéarité des relations entre les niveaux articulatoire et acoustique: on passe en régime fricatif avec de très petites variations d'aire à la constriction.

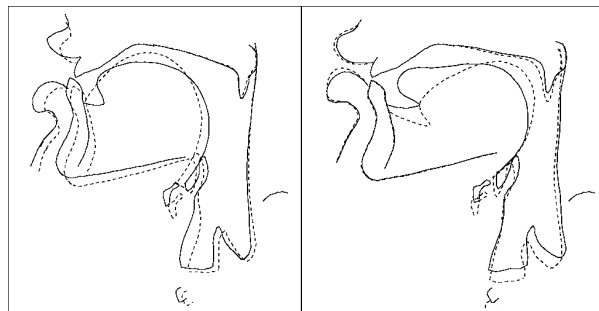


Figure5 : Transition Consonne-Voyelle dans les séquences [ʒi] et [ʒu]

La production de cette friction ne sera reprise sous contrôle véritablement actif, qu'en recrutant les degrés de liberté de l'apex ou de la lame de la langue (visible pour [ʒi] et surtout pour [ʒu], figure 5). La séquence obtenue présente un découplage des articulateurs lame/masse de la langue, bien meilleur, même pour [dʒu] par comparaison avec [ju], lequel doit avancer au maximum la position de [u] vers [j], une assimilation de lieu qui se produit en quelque sorte "dans le dos du son", en profitant : (i) de l'anticipation dans le [j] de la fermeture des lèvres (à laquelle cette constrictive n'est pas sensible, cf. [juju] arrondi de part en part), (ii) et de l'inaudibilité de l'avancée du lieu de [u] une fois les lèvres closes [Bai91]. Tout en

respectant les viabilités du système phonologique de la langue, un tel découplage assure la stabilité articulatoire-acoustique du son produit et améliore notablement le contraste acoustico-perceptif de la séquence.

Comme nous l'avons déjà constaté, la qualité acoustique de ce [ɕ] ou [ʒ] sera par la suite optimisée par une articulation labialisée, qui a pour effet d'augmenter la cavité de résonance de la friction.

## 5 UN SEUL DEGRÉ DE LIBERTÉ OU LES PREMIERS PAS DE LA COARTICULATION

Que se passe-t-il lorsque ces différents degrés de liberté — que nous avons vu se maintenir ou disparaître à différents degrés dans la coarticulation (section 3), voire émerger pour des raisons de stabilité articulatoire-acoustique (section 4) — se réduisent au seul contrôle de l'organe porteur de la langue et des lèvres : la mandibule ? Autrement dit lorsque le conduit vocal — respectivement le modèle articulatoire — ne fonctionne qu'avec un seul degré de liberté, tous les autres articulateurs habituellement actifs étant "gelés" ?

Dans cette perspective de recherche, nous avons pu commencer à tester certaines hypothèses concernant l'émergence du langage dans l'ontogenèse selon la théorie "Frame, then Content" de MacNeilage [Mac98, Abr98], basée sur l'observation du babillage canonique. Cette expérimentation a mis en évidence, à partir d'une

commande centrale identique, un phénomène de variation radicale des produits périphériques, aussi différents que [bəbə] vs. [dædæ], selon les modèles individuels utilisés [Abr97]. Ce phénomène nous semble être un apport majeur dans l'émergence de la coarticulation, puisque la forme finale peut être considérée simplement comme le "sous-produit" de trois types de contraintes : (i) la différence de conformation des sujets ; (ii) leurs stratégies de synergie individuelles ; (iii) et finalement le "gel" de tous les autres degrés de liberté, excepté la mandibule, ce qui simule l'absence d'indépendance des articulateurs portés, à ce stade du développement (7 mois).

## 6 CONCLUSION : DE DEGRÉ EN DEGRÉS

C'est par la suite qu'émergeront les autres degrés de liberté, dégagés pour Gentiane (section 2), permettant à la fois le contrôle de produits acoustiquement différenciés (après 7 mois, ce qui nécessitera de développer d'au moins autant la modélisation en section 5), une coarticulation variée, dans l'ensemble relativement souple, donc fluide (grâce aux degrés de liberté observés en section 3), et une stabilité articulatoire-acoustique de haute volée (par une optimisation évolutive rencontrée en section 4), soit au bout du compte une dextérité permettant de produire sur une porteuse à 6 cycles par seconde (la mandibule fonctionnant comme "pace-maker" de la syllabe), une cadence allant jusqu'à 15 modifications quantales du conduit vocal, dans le même temps: celui de la parole.

## RÉFÉRENCES

- [Abr80] Abry C., Boë L.-J. (1997), "À la recherche des corrélats géométriques discriminants pour l'opposition d'arrondissement vocalique en français", *Labialité et Phonétique...* (Abry Ed), Ellug, Grenoble, pp. 217-237.
- [Abr97] Abry C., Badin P., Vilain A., Stefanuto M., Boë L.-J. (1997), "Une théorie de l'ontogenèse à l'épreuve de trois modèles articulatoires anthropomorphiques", Programme GIS Cognisciences, Rapport mi-parcours.
- [Abr98] Abry C., Boë L.-J., Laboissière R., Schwartz J.-L. (1998), "Emergence of a new puzzle for the evolution of speech and language?", commentaire à [Mac98], à paraître.
- [Bad95] Badin P., Gabioud B., Beautemps D., Lallouache T.M., Bailly G., Maeda S., Zerling J.P., Brock G. (1995), "Cineradiography of VCV sequences : articulatory-acoustic data for a speech production model", 15th ICA, Trondheim, Norway, pp. 349-352.
- [Bad97] Badin P., Baricchi E., Vilain A. (1997), "Determining tongue articulation : from discrete fleshpoints to continuous shadow", Eurospeech, Rhodes, Greece, 1, pp. 47-50.
- [Bai91] Bailly G., Laboissière R., Schwartz J.L. (1991), "Formant trajectories as audible gestures : an alternative for speech synthesis. *Journal of Phonetics*", 19, pp. 9-23.
- [Lab95] Laboissière R., Galvan A. (1995), "Inferring the commands of an articulatory model from acoustical specifications of stop/vowel sequences", *ICPhS 95*, Stockholm, Suède, vol. 1, pp. 358-361.
- [Mac98] MacNeilage P.F. (1998), "The Frame/Content theory of evolution of speech production", *Behavioral and Brain Sciences*, à paraître.
- [Mad93] Maddieson I. (1993), "The structure of segment sequences", *UCLA Working Papers in Phonetics*, 83, pp. 1-8.
- [Mae89] Maeda S. (1989), "Articulation compensatoire des voyelles : analyse de données cinéradiographiques avec un modèle linéaire", *Mélanges de phonétique générale et expérimentale à Péla Simon*, Institut de Phonétique de Strasbourg, pp. 545-562.
- [Vil97] Vilain A. (1997), Un nouveau modèle articulatoire pour la synthèse et le contrôle robotique de la parole : Gentiane. DEA Sciences du Langage, ICP, Grenoble.