

La croissance de l'instrument vocal : contrôle, modélisation, potentialités acoustiques et conséquences perceptives

Louis-Jean Boë, Univ. de Grenoble / Muséum d'Histoire Naturelle-Paris

Lucie Ménard, Université du Québec à Montréal

Jihène Serkhane, Mitra Laboratory, New-York

Peter Birkholz, University of Rostock, Bernd Kröger, Aachen Universität

Pierre Badin, Université de Grenoble, Guillaume Captier, Université de Montpellier,

Mélanie Canault, Université de Strasbourg, Nicolas Kielwasser, Cluses

Résumé : *Pour pouvoir étudier précisément l'évolution du conduit vocal et des articulateurs, et en induire des informations sur le contrôle moteur et les potentialités acoustiques durant la croissance, il est essentiel de disposer de données tout au long de l'ontogenèse, de la gestation à l'âge adulte. La morphogenèse du conduit vocal, qui s'étend sur une vingtaine d'années et dépend, pour l'essentiel, des structures osseuses cranio-faciales, est loin d'être linéaire. Les nouvelles données de la génétique rendent lisibles les données de croissance osseuse et donc de celles du conduit vocal. Pour pouvoir prédire les conséquences de la croissance sur les premières manifestations vocales pré-linguistiques (les proto-syllabes et le babillage), les stratégies de contrôle articulaire, et quantifier les potentialités acoustiques du conduit vocal, des modèles sont mis en œuvre le plus proche possible des données articulatoires. Les stimuli de synthèse articuloirement réalistes permettent de tester des hypothèses sur le processus perceptuo-moteur de l'enfant. Enfin ces modèles sont utilisables dans le cadre des recherches phylogénétiques sur l'émergence de la parole. L'étude de la croissance du conduit vocal constitue un paradigme expérimental de choix pour les recherches en parole.*

Abstract: *In order to accurately follow the growth of the vocal tract and the speech articulators and to infer their motor control and acoustic potentialities, it is essential to have anatomic and physiologic data throughout ontogenesis, from gestation to adulthood. The morphogenesis of the vocal tract, which involves the bony structures, their development, and cranio-facial growth during ontogenesis, is far from linear. The new data provided by genetics help interpreting the process of bony growth and thus of vocal tract reshaping from fetus to adult. To predict the consequences of vocal tract growth on the first speech productions (proto-vocalic and proto-consonantal articulations, babbling), anthropomorphic articulatory models are developed based on articulatory data. The articulatory models can generate realistic acoustic stimuli, which enable the testing of hypotheses about adult and newborn perceptuo-motor processes. These articulatory models can provide phylogenetical cues to the debate around the emergence of speech. The study of vocal tract growth therefore constitutes a key experimental paradigm for speech research.*

1. Introduction

Dans les manuels de référence consacrés à la production de la parole, le conduit vocal est décrit comme un tout (un tube) modelé par la disposition des articulateurs et, pour l'essentiel, par celle de la langue et des lèvres. Il n'est que très rarement fait référence à l'architecture osseuse cranio-cervico-faciale dans laquelle et par rapport à laquelle ce conduit est disposé.

Du fait de leur visibilité sur les radiographies, les vertèbres cervicales et l'os hyoïde ne sont évoqués que pour situer approximativement la position du larynx et celle des cordes vocales. Seules les dents et le palais dur constituent des repères explicites dans les descriptions phonétiques. La mandibule porteuse de la langue, des lèvres et du larynx, a été longtemps absente des descriptions des modèles de production, la langue faisant, elle seule, l'objet des analyses statistiques permettant d'extraire des composantes principales interprétables en termes articulatoires. Depuis les propositions de Browman et Goldstein (1986), la phonologie articulatoire, basée sur les gestes de production porteurs de contrastes phonologiques, ne réfère qu'aux mouvements des lèvres, de la langue, du vélum et à l'aperture glottique, les gestes de la mandibule n'étant qu'implicites. Et pourtant, le rôle de cette pièce osseuse est crucial pour sa contribution dans la position du larynx (l'os hyoïde est suspendu à la mandibule), de la langue, des lèvres ; la mandibule joue un rôle clé dans l'acquisition et le développement de la parole, comme l'ont montré les travaux sur le babillage. En effet celui-ci, qui apparaît autour de 7 mois, et qui est produit par l'oscillation rythmique de la mandibule, peut être considéré comme le véritable départ de l'acquisition de la parole.

Comme le conduit vocal est inséré dans l'architecture osseuse, il est important de suivre en détail la croissance de repères anatomiques osseux de la tête et du cou, tout au long de l'ontogenèse (de la gestation à l'âge adulte) pour en analyser les évolutions et en induire les conséquences pour le contrôle des articulateurs, la configuration du conduit vocal et ses potentialités acoustiques. D'autant plus que le conduit vocal d'un nouveau-né n'est pas du tout l'homothétique de celui d'un enfant, d'un(e) adolescent(e) ou d'un(e) adulte. Sa croissance dépend essentiellement de celle du crâne, de la face et des vertèbres qui s'accompagne de la rotation basicrânienne (du sphénoïde et du basi-occipital). Elle est orchestrée par la biologie du développement génétique.

Les domaines pour lesquels il nous semble nécessaire de croiser des données concernent :

- la génétique du développement osseux (à partir de la fécondation et au cours de la période fœtale), la biométrie physique du crâne, de la face et du rachis tout au long de l'ontogenèse pour pouvoir disposer d'une vue d'ensemble dans la continuité ;
- la modélisation anthropométrique du conduit vocal et sa commande (éclairée par les principes de la robotique et/ou les réseaux neuronaux) ;
- la phonétique de l'acquisition de la parole et du développement adossée à une théorie de la production-perception de la parole et du contrôle ;
- et enfin la simulation acoustique dynamique permettant, pour le moins, de synthétiser le babillage et des séquences de syllabes.

2. Nouvelles données génétiques : les gènes architecturaux de la tête et du cou

Les mesures céphalométriques délivrent, tout au long de l'ontogenèse, les coordonnées 2D et 3D de points de repères anatomiques et géométriques, mais ces informations ne sont pas suffisantes pour interpréter, dans son ensemble, la morphogenèse de la tête et du cou et surtout pour en percevoir le fil directeur. Les avancées de la génétique, au cours de ces vingt dernières années, ont débouché sur une nouvelle approche permettant de rendre lisible ce qui n'était que visible de la croissance osseuse de la tête et du cou.

À partir des 15-20 premiers jours qui suivent la fécondation (Couly & al., 1993; Couly & al., 2002), les gènes du développement, ou gènes homéotiques, sont responsables de la construction embryonnaire (figure 1a) et déterminent l'organisation antéropostérieure et dorso-ventrale de l'embryon, la mise en place de la base du crâne, de la tête et du corps. La synthèse de ce développement (Benoît, 2001, 2008) nous permet de disposer des grandes lignes de l'ossification de la tête, de l'os hyoïde, et du cou. Les gènes non HOX sont

responsables de l'ossification membraneuse de la partie antérieure et supérieure de la tête et de la partie antérieure de la mâchoire ; les gènes HOX de l'ossification enchondrale de la partie postérieure du crâne (l'occipital), de la base du crâne, de l'os hyoïde, et du rachis cervical (les vertèbres C1 à C7). L'expression des gènes du développement (HOX et non HOX) *architecture* ainsi deux champs, situés de part et d'autre d'une limite passant dans le plan sagittal médian par le point lambda (à la rencontre des sutures interpariétale et occipito-pariétales), la crête synostotique antérieure du sphénoïde, l'insertion inférieure (la lingula) du ligament sphéno-mandibulaire et le corps de l'os hyoïde (figure 1b). Dans une étude récente, nous avons montré que pour une population d'enfants autour de 6 ans on retrouve bien, sur les structures osseuses de la tête et du cou, la partition entre l'influence des gènes HOX et non HOX ; avec, comme conséquence, la possibilité d'associer le déplacement ontogénétique vertical de repères du domaine HOX (par exemple celui de l'os hyoïde avec celui des vertèbres cervicales et de la partie postérieure de la mandibule) (Boë & al., 2007).

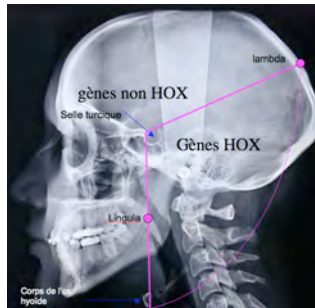
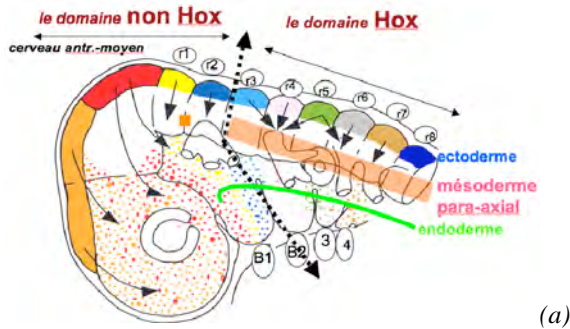


Figure 1a. L'expression des zones des gènes HOX et non HOX.

Représentation schématisée de la migration et de la destination des cellules des crêtes neurales.

Représentation schématisée d'après Charrier, Creuzet (2007).

Figure 1b. Zones d'influence des gènes HOX et non HOX sur les structures osseuses de la tête et du cou (d'après R. Benoît, 2001; 2008).

3. Les repères de la biométrie physique : la céphalométrie crânienne

Les données de la génétique rendent ainsi nettement plus lisibles et compréhensibles l'évolution de la structure osseuse du crâne, de la face et du rachis et, par cela

4. Les données anatomiques du conduit vocal : les parties molles du larynx aux lèvres

L'anatomie des structures osseuses et des parties molles qui délimitent et constituent le conduit vocal est bien connue pour l'adulte, l'adolescent et l'enfant, mais nettement moins décrite pour le nouveau-né et le fœtus. Les principales différences entre un fœtus de 5 mois et un adulte (figure 3) se caractérisent par :

- la rotation basicrânienne, due à la poussée du cerveau, qui peut être spécifiée par une diminution de l'angle sphénoïdal qui passe de 135° à 110° (sur la figure 3 : nasion - selle turcique - basion) ;
- l'abaissement du palais dur par rapport à la base du crâne (le basion, sur la figure 3), dû à l'augmentation de la dimension verticale du maxillaire ;
- le déplacement antéro-postérieur du corps de la langue qui est associé à un rapprochement de la paroi pharyngale vers le rachis cervical ;
- un remodelage très important de la mandibule avec une diminution de l'angle goniale, l'apparition du menton, la verticalisation du ramus et son rapprochement vers le rachis cervical ;

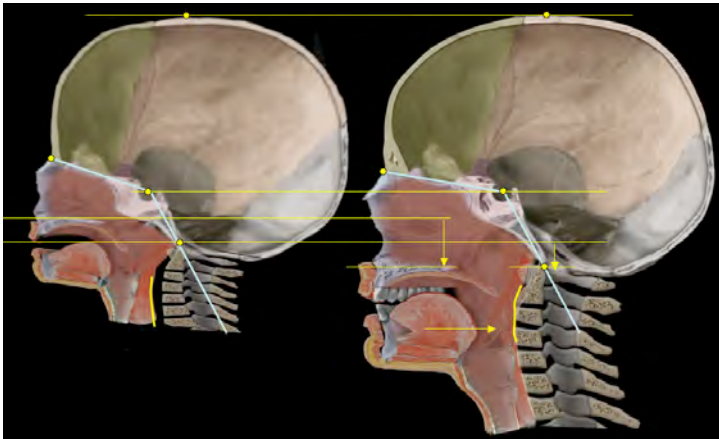


Figure 3. Les principales modifications du crâne qui déterminent celles du conduit vocal, du fœtus à l'âge adulte (homme) : diminution de l'angle sphénoïdal, déplacement du palais vers le bas, à hauteur du basion, rapprochement de la paroi pharyngale vers le rachis.

- le déplacement vertical de l'os hyoïde et du larynx, dû pour l'essentiel à la morphogenèse de la mandibule : à 5 mois les cordes vocales sont au niveau de la troisième vertèbre cervicale C3, pour une femme au niveau de C5 et pour un homme entre C5 et C6.

Pour pouvoir appréhender dans leur globalité ces remodelages, un effort important de visualisation a été fait par l'un d'entre nous (N.K.) grâce à l'utilisation de la

synthèse numérique dynamique appliquée à la croissance du cerveau, de l'architecture osseuse, et du conduit vocal.

Sur le plan musculaire, le remodelage de toute la structure osseuse du crâne, de la face et du rachis provoque une relocalisation de l'ensemble des muscles de la langue avec des différences de volume importantes et des variations très nettes pour la longueur et l'angulation du styloglosse (figures 4a/b), ce qui ne sera pas sans conséquence sur son recrutement (pour la voyelle [u] ou la consonne [k] par exemple).



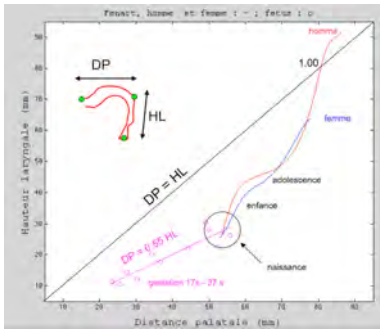
Figure 4a. Modifications de la face et de la mandibule qui entraînent un remodelage de la morphologie musculaire de la langue.

Figure 4b. Détail de l'implantation et de la direction des fibres musculaires du styloglosse pour un adulte.

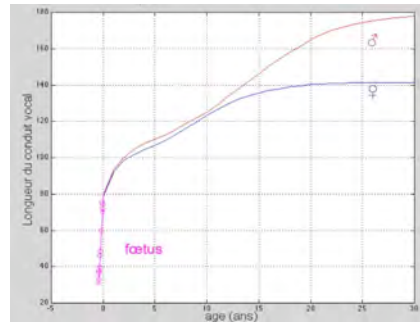
Il existe de nombreuses données articulatoires concernant l'anatomie du conduit vocal pour les adultes, beaucoup moins pour les enfants et très peu pour les bébés. Celles qui couvrent de manière continue la croissance de la naissance jusqu'à l'âge adulte sont exceptionnelles : à notre connaissance, il n'existe que celles de Goldstein (1980). Elles constituent, encore aujourd'hui, une source précieuse, d'autant plus qu'elles sont référencées par rapport à l'architecture craniofaciale. Pour les données osseuses, mis à part quelques légères différences, elles sont tout à fait compatibles avec celles de Fenart (2003). Pour les données fœtales du conduit vocal nous avons utilisé les données de Captier qui sont pratiquement les seules disponibles (Boë & al., 2008).

Il est important de noter une différenciation très nette entre les croissances antéropostérieure et verticale du conduit vocal (la distance palatale et la hauteur pharyngale) : le rapport de ces valeurs constitue l'Indice de Hauteur Laryngale (Honda & Tiede, 1998). Durant la croissance du fœtus (nos données commencent à 17 semaines) cet indice reste constant jusqu'à la naissance : le conduit vocal croît de façon homothétique, avec une dimension palatale à peu près deux fois plus grande que la hauteur laryngale. Puis, dès la naissance, cette hauteur va croître beaucoup plus vite que la dimension palatale, plus tôt pour les femmes, plus tard et plus longtemps pour les hommes ; pour ceux-ci, le conduit vocal atteint, à maturité, les mêmes dimensions horizontale et verticale (figure 5a). Globalement la dimension du conduit vocal varie du simple au double : 6 cm pour un fœtus de 5 mois, 8 cm à la

naissance, 15 cm pour une femme, 17,5 cm pour un homme. Cet allongement ne varie pas linéairement avec l'âge et on observe des différences de vitesse de croissance tout au long de la gestation jusqu'à la fin de l'adolescence. Contrairement aux données crâniennes, la longueur du conduit vocal, et plus précisément la hauteur pharyngale, est liée au sexe (figure 5b).



(a)



(b)

Du fœtus jusqu'à l'âge adulte.

Figure 5a. Variations de l'indice de hauteur laryngale (rapport de la hauteur laryngale HL à la distance palatale DP).

La droite DP = HL correspondrait à une simple homothétie du conduit vocal.

Figure 5b. Variations de la longueur du conduit vocal (mm) pour les femmes et les hommes. Données de Boë & al. (2008) pour les fœtus et de Goldstein (1980) de la naissance à l'âge adulte.

5. Le contrôle du conduit vocal lors du développement

Pour pouvoir contrôler un modèle articulatoire de synthèse pour l'acquisition de la parole au cours des deux premières années, on dispose actuellement de la théorie *Frame then Content* de MacNeilage et Davis et de plus en plus de données acoustiques et articulatoires qui permettent de comprendre les stratégies de contrôle et de couplage des articulateurs qui se mettent en place au cours de l'ontogenèse.

Lèvre-mandibule

Le mouvement des lèvres au cours du babillage serait initialement une conséquence passive de l'oscillation mandibulaire, les productions précoces étant marquées par un déplacement prononcé de la mandibule associé à une très nette compression des lèvres. Puis, l'habileté labiale va se développer. À 6 ans, les patrons de l'enfant se rapprochent de ceux de l'adulte : les mouvements, bien qu'encore variables, vont se dissocier et devenir indépendants avec une priorité de la lèvre inférieure (Green & al., 2002), pour se stabiliser peu à peu. Toutefois, les mouvements des lèvres et de la mandibule maintiendront un couplage étroit au niveau de leur *timing* (Gracco & Lofqvist, 1993).

Lèvre-langue

De récents travaux (Ménard & al., 2008) ont été menés sur les possibilités compensatoires des enfants français, âgés de 4 ans, produisant un [u] avec un *lip-tube*, c'est-à-dire contraints de réaliser cette voyelle avec une aire aux lèvres nettement plus importante (1,76 cm²) que dans les conditions naturelles (nettement inférieure à 1 cm²). Pour arriver à produire la même cible acoustique dans une telle situation, il faut opérer une compensation en déplaçant le corps de la langue vers l'arrière. Les résultats montrent que, comme chez les adultes (Savariaux & al., 1995, 1999), tous les sujets sont capables d'opérer cette compensation, mais qu'à 4 ans la maîtrise du contrôle moteur n'est pas assez mature pour qu'elle soit apprise par intégration successive des différentes stratégies à l'intérieur d'un modèle interne, comme pour l'adulte. Il semble, au contraire, que l'enfant de cet âge procède plutôt par essais et erreurs.

Mandibule-langue

Autour de 12 mois, l'enfant va commencer à contrôler séparément la langue et la mandibule. À cet âge, les mouvements linguaux sont moins corrélés à ceux de la mandibule qu'à 8 mois. Le bébé accroît l'amplitude de ses mouvements linguaux dans la dimension antérieure-postérieure de la cavité buccale entre le stade du babillage redupliqué et celui du babillage varié. Peu à peu, il apprend à déplacer simultanément sa mandibule sur le plan vertical et la langue dans un plan antéro-postérieur (Canault & al., 2006; Canault, 2007a/b). Le manque de contrôle différencié de la mandibule et de la langue, au stade du babillage canonique, est également décelable dans les patrons de cooccurrences (Davis & MacNeilage, 1990; MacNeilage, 1998; MacNeilage & Davis, 2000a/b, 2001). Ces schémas associatifs s'établiraient tous sur la base du cadre biphasique de l'oscillation mandibulaire et leur organisation ferait apparaître une relative inertie de la langue lors de la transition *consonne voyelle* CV qui accentuerait l'hypothèse de la dominance du cadre dans les réalisations précoces. Trois types d'association CV semblent alors privilégiés. La première configuration décrite par la théorie *Frame then Content* (MacNeilage, 1998) résulterait des propriétés mécaniques de la mâchoire inférieure. La simple oscillation mandibulaire serait responsable des associations préférentielles du type Consonne labiale + Voyelle neutre, alors qualifiées de *cadre pur* (*pure frame*). De cette manière, l'élévation de la mandibule suffirait à réaliser l'occlusion du conduit vocal, alors que le maintien de la langue en position de repos au cours de son abaissement impliquerait le patron de résonance d'une voyelle centrale ou neutre. Deux autres patrons de cooccurrence, les cadres antérieur et postérieur, sont observables : la langue peut se positionner plus en avant dans la cavité buccale et, dans ce cas, les consonnes coronales seront de préférence combinées aux voyelles antérieures (ex. : [te], [de]). Le cadre assigné à ce mouvement frontal est dénommé *cadre antérieur* (*fronted frame*). À l'inverse, lorsque l'association *consonne vélaire/voyelle postérieure* est privilégiée, le positionnement lingual est plutôt postérieur (ex. : [ku], [gu]) et le cadre résultant est, par conséquent, désigné comme *cadre postérieur* (*backed frame*). La robustesse de tels patrons a pu être vérifiée par Davis et MacNeilage (1994) entre 7 et 12 mois. De la même manière,

Vihman (1992) a pu évaluer leur viabilité à travers différentes sources de données (enregistrements audio et vidéo hebdomadaires de 10 bébés de 9 à 16 mois, enregistrements vidéo mensuels de 10 enfants anglais âgés de 8 à 24 mois et de données interlinguistiques concernant les productions de 5 enfants français, 5 suédois et 5 japonais). La prévalence des associations prédites a été confirmée avec une majorité des schémas *Consonne labiale + Voyelle centrale* et *Consonne vélaire + Voyelle postérieure*. Plus tard, en focalisant sur la production des premiers mots de 10 enfants monolingues anglais, Davis et al. (2002) ont attesté que ces tendances prédominaient au-delà de la période du babillage. L'analyse des patrons intracycliques, CV et VC, a révélé que les associations *coronale-antérieure et labiale-centrale* étaient présentes dans tous les environnements et ressortaient comme les patrons les plus fréquents. L'association *dorsale-postérieure* apparaissait, quant à elle, uniquement dans les séquences CV au sein d'énoncés multisyllabiques.

Dans une situation de contrainte, avec la mâchoire calée (*biteblock experiment*), les expériences de Oller et MacNeilage (1983) ont permis de conclure que des enfants de 4 et 8 ans n'arrivaient pas à compenser pour atteindre les mêmes cibles acoustiques. Il leur était demandé de répéter des [i] et [æ] avec leur mandibule libre et avec un *biteblock*. Les analyses spectrales ont montré des différences acoustiques entre les deux conditions. Pourtant les auditeurs ont jugé que les productions étaient assez bonnes, amenant à conclure que les enfants préservaient d'autres paramètres acoustiques (comme la durée).

Dans un récent travail combinant modélisation et données perceptives, Ménard et al. (à paraître) ont montré que pour une même cible vocalique perçue en anglais, le couplage entre la position haut-bas de la mâchoire et la position antéro-postérieure du corps de la langue était moins important que chez l'adulte. Ce résultat, basé sur des analyses de régressions linéaires, révèle que la morphologie du conduit vocal contraint le contrôle moteur des articulateurs de la parole.

Vélu

Les mouvements du vélu ne seraient pas contrôlés à 7 mois (Lalevée & Vilain, 2003). Pourtant, de nombreux auteurs relatent la présence de séquences du type *mama*, au sein desquelles les articulations nasales et orales sont alternées. À première vue, le changement de mode articuloire pourrait être considéré comme une première forme de contrôle vélaire dans la mesure où la position du vélu changerait au cours d'un cycle mandibulaire en adoptant successivement une position élevée puis abaissée. Cependant, Rossato et al. (2003) ont émis l'hypothèse qu'un contrôle très fin du vélu ne serait pas nécessaire pour produire des séquences du type *mama*. Il existe une gamme de hauteurs du vélu où les voyelles orales et les consonnes nasales peuvent être produites. Une séquence *mama* pourrait donc être produite avec une position constante du vélu, qui serait sa configuration *par défaut* relâchée et non contrôlée (Lalevée & Vilain, 2003).

Toutes ces données seront donc à implémenter dans les modèles de croissance pour générer des stimuli conformes à ceux qui sont observés au cours de l'acquisition de la parole.

6. La modélisation anthropomorphique de la croissance du conduit vocal

La modélisation en deux et trois dimensions

Les deux modèles que nous avons choisis de présenter intègrent les données de Goldstein (1980), jusqu'aux plus récentes (Fenart, 2003 ; Vorperian & al., 2005 ; Boë & al., 2008). Ils permettent de simuler en 2D la croissance du conduit vocal du fœtus jusqu'à l'âge adulte et, en 3D, depuis 1 an jusqu'à la maturité ; ils tiennent compte du dimorphisme sexuel. Pour pouvoir assurer un anthropomorphisme du conduit vocal la modélisation est assurée par rapport à l'architecture osseuse de la tête et du cou, pour l'essentiel la base du crâne, le palais dur, la mandibule, l'os hyoïde et les vertèbres cervicales (figure 6).

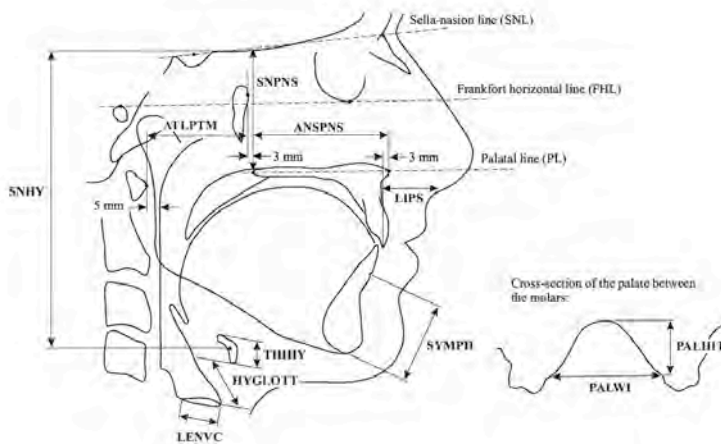


Figure 6. Les principaux points de références et distances qui permettent de repérer le conduit vocal dans l'architecture cranio-cervico-faciale (d'après Goldstein, 1980).

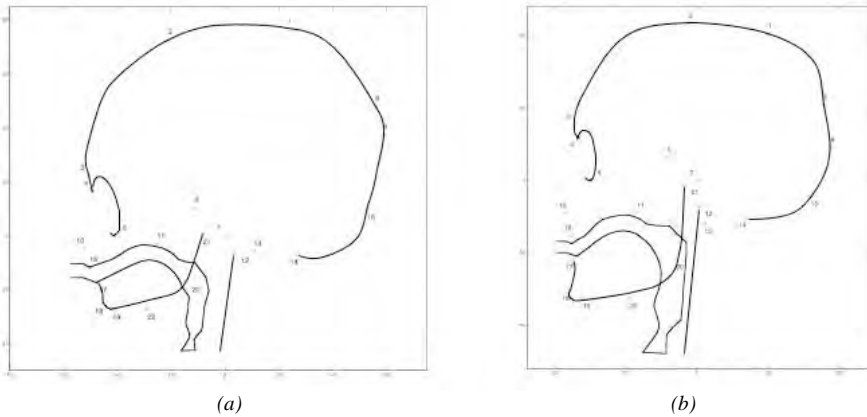
Le modèle 2D

C'est une amélioration de VLAM (*Variable Linear Articulatory Model*) développé par Maeda (Boë & Maeda, 1998). Le conduit vocal est divisé en trois zones : antérieure, pharyngale et intermédiaire : tout au long de l'ontogenèse, mois par mois, leurs dimensions sont ajustables séparément et linéairement, ce qui permet de réaliser pour l'ensemble du conduit vocal une anamorphose non linéaire. Le modèle initial a été remanié : l'ensemble des paramètres de commande permet de modifier la profondeur du palais en fonction de l'âge, l'aplatissement correspondant de la langue, et les conséquences de la variation de l'angle d'inclinaison de la tête. Il a été étalonné pour refléter les dimensions mesurées au cours de la croissance et systématiquement testé au niveau acoustique (Ménard, 2002). Le modèle actuel (Boë & al., 2006) est ainsi piloté par :

- des paramètres anatomiques : profondeur du palais, rapport entre la dimension antéro-postérieure de la cavité buccale et de la hauteur pharyngo-laryngale, l'aplatissement de la langue ;
- un paramètre positionnel : l'inclinaison de la tête par rapport au rachis cervical ;

- des paramètres articulatoires : l'ouverture et la protrusion des lèvres, les positions du corps du dos et de la pointe de la langue, l'ouverture de la mâchoire et la position du larynx ;
- l'âge et le sexe : ils déterminent le rapport entre la dimension antéro-postérieure de la cavité buccale et la hauteur pharyngo-laryngale et fixent la fréquence de vibration des cordes vocales.

Le modèle délivre la coupe sagittale (figures 7) à partir de laquelle est calculée la fonction d'aire et la fonction de transfert qui permet de générer un son.



Figures 7. Le modèle de croissance 2D :

génération de la coupe sagittale pour un nouveau-né (a) et un homme adulte (b).

Le conduit vocal est localisé par rapport aux principaux repères anatomiques crano-cervico-faciaux.

On peut noter les différences respectives entre les dimensions antéro-postérieures et pharyngales du conduit vocal, la taille de la mandibule et la position de la paroi pharyngale par rapport au rachis (ici sa limite antérieure).

Le modèle peut être ajusté à l'architecture osseuse, à partir d'une photographie du crâne, une coupe anatomique, une radiographie ou une image IRM (logiciel *LabSkull*) (figures 8 infra).

Le modèle 3D

Développé par Birkholz et Kröger (2006, 2007), ce modèle est le premier qui permet de simuler la croissance en 3D. Il est piloté par une dizaine de paramètres permettant de fixer : la position de l'os hyoïde (HX, HY), l'ouverture du voile du palais (VA), la protrusion et l'ouverture des lèvres (LP, LH), la position et la rotation de la mandibule (JX, JY, JA), le contour sagittal médian de la langue (de la partie pharyngale à l'apex : TRE, TCX, TCY, TBX, TBY, TTX, TTY) ; quatre paramètres additionnels déterminent la hauteur des bords de la langue par rapport à la partie sagittale médiane (pour les détails, voir Birkholz, 2005 ; Birkholz & al., 2006). Ces paramètres sont ajustés en fonction de l'âge à partir des données de Goldstein (1983).

de fonction d'aires pour la séquence temporelle correspondante (Birkholz, 2007) permettant notamment de générer le babillage de type *bababa* ou *dadada* (Boë & al., 2008).

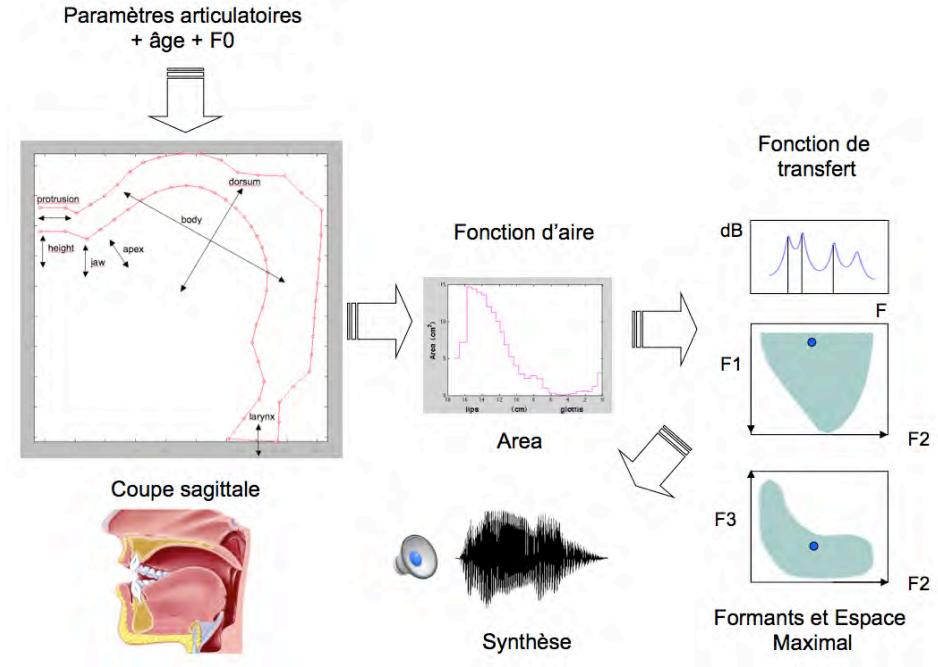


Figure 9. Les étapes du passage de l'articulatoire à l'acoustique, de la coupe sagittale au signal sonore. Pour une synthèse dynamique, il faut disposer au cours du temps (par exemple toutes les 20 ms) de toutes les séquences des paramètres de commande du modèle pour calculer les coupes sagittales, les fonctions d'aire, les fonctions de transfert correspondantes et générer le signal acoustique.

7. Les données acoustiques

La source vocale

Au cours de la croissance, l'accroissement de la longueur des cordes vocales et leur épaisseur provoquent un abaissement de la fréquence laryngienne avec une différence homme/femme de l'ordre de l'octave qui est très significative. La modélisation des variations de F0 pour les femmes et les hommes que nous proposons est dérivée des données de Beck (1997) et Lee *et al.* (1999). F0 (en Hz) est ajustée pour la croissance par une double logistique (figure 10a) avec des paramètres différents pour les femmes et les hommes :

$$F0 = F0_{\max} - (\alpha / (1 + \exp(\beta - \gamma^2 \text{age})) + \epsilon / (1 + \exp(\epsilon - \varphi^2 \text{age})))$$

Ce qui permet de disposer de valeurs moyennes de la hauteur de la voix tout au long de la croissance et donc, pour la synthèse, d'exciter un modèle articulatoire de manière réaliste.

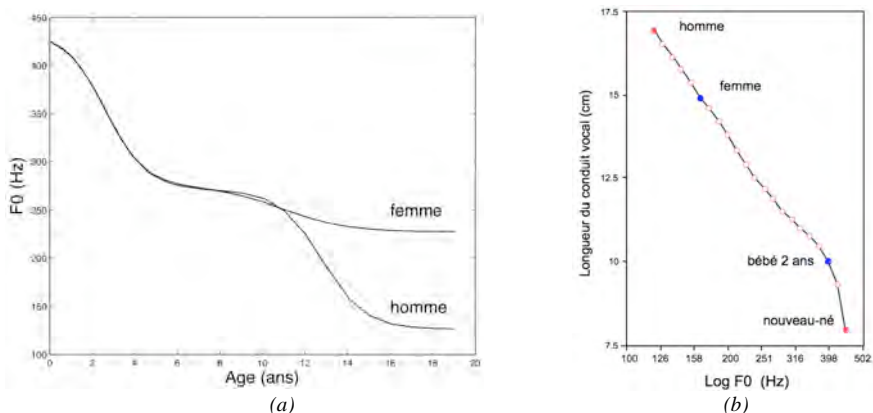


Figure 10a. Évolutions de la fréquence fondamentale moyenne au cours de la croissance pour les femmes et les hommes optimisées par deux doubles logistiques.

Figure 10b. Relation, année par année, entre la longueur moyenne du conduit vocal et le logarithme de la fréquence moyenne de vibration des cordes vocales.

La corrélation entre ces deux variables est très significative.

Il est important de noter que la fréquence fondamentale est donc un indicateur assez fiable du sexe et de l'âge au cours de la croissance, mais il existe aussi une synchronie, de la naissance à l'âge adulte, entre la fréquence fondamentale moyenne et la longueur moyenne du conduit vocal (figure 10b). Cette corrélation va pouvoir être utilisée par les mécanismes de normalisation perceptive entre un enfant et un adulte.

Le conduit vocal

La position du rétrécissement à l'intérieur du conduit vocal (qui caractérise le lieu d'articulation), l'aire à ce niveau, l'aire aux lèvres et l'amplitude de leur projection sont les caractéristiques géométriques du conduit vocal qui ont les conséquences les plus importantes sur les valeurs des formants. À l'intérieur du conduit vocal, le rétrécissement, contrôlé par le déplacement de la lame ou du dos de la langue, délimite ainsi le conduit vocal en deux parties (antérieure et postérieure). Celles-ci ne correspondent pas nécessairement au découpage anatomique en cavités buccale et pharyngienne. Les voyelles cardinales extrêmes [i a u] correspondent à un rétrécissement nettement marqué (une faible aire à la constriction). Si l'on veut produire ces trois voyelles universelles, la connaissance élémentaire de l'acoustique (Fant, 1960) démontre que :

- il n'y a qu'une seule solution pour produire une voyelle possédant une troisième résonance élevée comme celle du /i/ : il faut contrôler une constriction avec le front de la langue mise en gouttière contre le palais dur ; par simple conséquence, il s'ensuit que la longueur de la cavité arrière restante sera toujours suffisante pour produire une deuxième résonance, elle aussi élevée ; quant à la première résonance, elle sera toujours assez basse pour être perçue comme celle d'un /i/ ;
- produire les deux premières résonances caractéristiques des deux autres voyelles extrêmes ne pose aucun problème : grâce à deux résonateurs de Helmholtz pour le /u/ et deux tubes en forme de pavillon, pour le /a/.

Et il est possible de vérifier ceci quel que soit le modèle utilisé, qu'il soit ultra-simplifié ou hyper-sophistiqué (figure 11a).

Les trois voyelles cardinales sont positionnées dans l'espace formantique (F1 F2 F3) de manière à présenter les unes par rapport aux autres une distance maximale offrant ainsi une différenciation acoustique optimale. On peut vérifier la validité de ces prédictions en utilisant un modèle très simplifié du conduit vocal, le modèle à 4 tubes de Fant. On génère un grand nombre d'items vocaliques (10 000 par exemple) pour cartographier ce que l'on appelle l'Espece Vocalique Maximal (Boë & al., 1989). L'EVM dessine alors dans le plan F1 F2 le triangle vocalique bien connu : à ses extrémités, se retrouvent les configurations [i a u] prédites par la théorie de la dispersion (Liljencrants, Lindblom, 1972; Schwartz & al., 1997) (figure 11b).

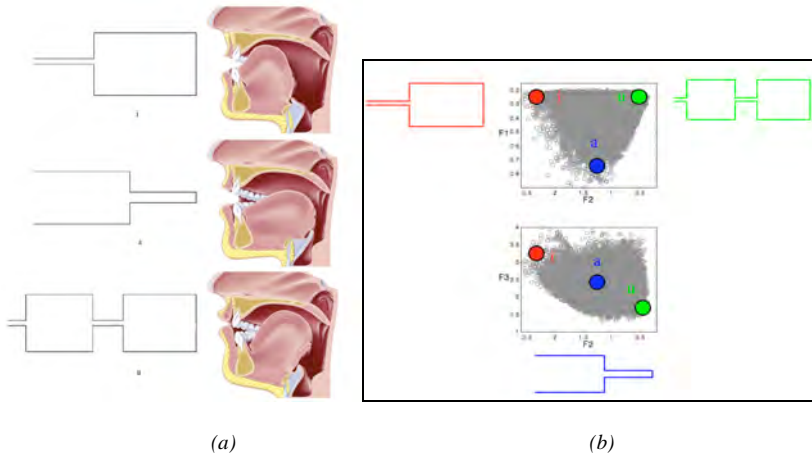


Figure 11a.. De haut en bas : les configurations vocaliques pour /i a u/ schématisées et les coupes sagittales correspondantes.

Figure 11b. L'espace vocalique maximal cartographié dans les plans F1 F2 et F2 F3 à partir de 10 000 items générés par un modèle de conduit vocal simplifié (4 tubes, Fant, 1960). Aux extrémités du triangle F1 F2 on retrouve bien les formes prototypiques des trois voyelles cardinales. Dans l'espace F1 F2 F3, ces voyelles présentent bien un contraste maximal prédit par la théorie de la dispersion.

Les conséquences de la croissance du conduit vocal sur l'espace maximal vocalique sont bien connues depuis les mesures de Peterson et Barney (1950) et prévisibles (Fant, 1960). Des conduits vocaux de longueurs différentes ont exactement les mêmes potentialités de différenciation acoustiques, à ceci près que les EVM sont décalés dans l'espace des fréquences (les fréquences de résonance d'un conduit vocal sont inversement proportionnelles à sa longueur, mais les rapports des résonances restent constants).

À partir de générations systématiques, on obtient ainsi avec un modèle articulatoire (le 2D, en l'occurrence) les EVM au cours de la croissance. Ils sont quasiment identiques (en Bark) : le bébé, s'il avait les mêmes capacités de contrôle articulatoire que l'adulte, pourrait produire, dès la naissance, le même espace vocalique. L'EVM constitue la donnée de base pour l'élaboration des prototypes vocaliques (F1 F2 F3) et consonantiques (F2 F3 pour les plosives) définis par les paramètres articulatoires, la coupe sagittale, la fonction d'aire, la fonction de transfert et ses formants, le son lui-même. Bien connus pour les adultes, les modèles articulatoires ont permis de nouvelles propositions pour les enfants (Ménard, 2002).

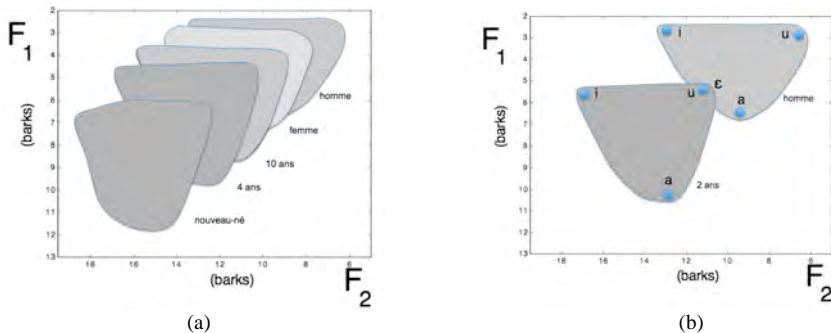


Figure 12a. Les espaces vocaliques maximaux de l'homme, de la femme, d'un enfant de 10 ans, de 4 ans et d'un nouveau-né dans le plan F1 F2 avec des échelles en Bark qui rendent compte de l'espace perceptif.

Figure 12b. Dans l'espace F1 F2 les voyelles [ε] de l'adulte et [u] de l'enfant de deux ans ont les mêmes valeurs.

8. L'utilisation de stimuli synthétiques pour tester des modèles de contrôle et de perception

La perception bébé-adulte

Au cours de la croissance, avec la variation de la longueur du conduit vocal, l'espace vocalique (et consonantique) se déplace en fréquence, ce qui entraîne des décalages formantiques. Pour donner un exemple : les formants du [u] de l'enfant de 2 ans sont les mêmes que ceux de la voyelle [ε] de l'adulte (figure 12b). Dans une intercompréhension enfant-adulte réussie, s'opère donc une normalisation perceptuelle qui permet à chacun d'eux de récupérer les cibles vocaliques. Ce processus ne peut pas fonctionner uniquement sur les espaces traditionnels F1 F2, F2 F3. Quels sont donc le(s) paramètre(s) normalisateur(s) mis en jeu dans ce phénomène ? La synthèse articulatoire permet de générer des stimuli en jouant sur F0 et sur la qualité vocalique, c'est-à-dire en produisant des voyelles avec un conduit vocal d'enfant et des cordes vocales d'adulte et inversement, avec toutes les possibilités intermédiaires. Toute une série de travaux (Ménard & Boë, 2000; Ménard, 2002; Ménard & al., 2002, 2004, 2007 et à paraître) ont montré (avec des voyelles du français et de l'anglais) que les perceptions de l'aperture et de l'arrondissement sont corrélées respectivement aux paramètres (exprimés en Bark) F1-F0, F2-F1 (ou F3-F2) et F'2 (un paramètre global tenant compte de F2, F3 et F4). Pour le degré d'aperture, les frontières entre voyelles fermées, mi-fermées et mi-ouvertes s'établissent respectivement à 2 et 4 Barks. La nature même de ces paramètres normalisateurs confère une robustesse au mécanisme de perception malgré la variabilité induite par la croissance. En fait c'est la hauteur de la voix qui joue, combinée aux valeurs formantiques, le rôle important de cette normalisation. Avec la fréquence moyenne de vibration des cordes vocales F0, qui nous l'avons vu est corrélée à la longueur du conduit vocal par l'intermédiaire de l'âge (figure 10b), l'auditeur peut opérer un recalage fréquentiel. Ces paramètres normalisateurs permettent à leur tour de définir des gabarits acoustiques qui guident le locuteur dans le choix des stratégies articulatoires à produire afin d'atteindre une cible vocalique bien perçue. Compte tenu des changements importants dans la morphologie du conduit vocal au cours de la croissance, ces stratégies se modifient de l'enfance à l'âge adulte. Produire une voyelle nécessite donc un réajustement constant des organes articulatoires afin de compenser les changements morphologiques.

Le contrôle des réalisations pré-vocaliques à 4 mois et du babillage à 7 mois

Entre 3 et 5 mois, les bébés sont capables d'une imitation vocalique précoce. Si on leur présente en boucle sur un écran le visage et le son d'un adulte prononçant les voyelles-cible [i a u], les enregistrements de leurs vocalisations peuvent être étiquetés en trois macro-classes, des pseudo [i a u] (Kuhl & Meltzoff, 1996). Les productions analysées plaident en faveur d'un comportement imitatif pour le groupe des plus âgés (4-5 mois). Les nourrissons sont donc globalement capables d'imiter sélectivement les voyelles audio-visuelles, dans la mesure de leurs capacités de contrôle articulaire. En utilisant le modèle articulaire de croissance et ces données acoustiques, les travaux de Serkhane *et al.* (2002, 2007) montrent que cette imitation précoce, due pour l'essentiel à la stimulation acoustique, peut être atteinte avec les mêmes résultats acoustiques dans un apprentissage de type robotique bayésienne avec moins de 100 essais. L'inférence articulaire (*le recadrage articulaire*) montre que les formes moyennes de langue tourneraient autour de la configuration neutre avec une tendance à s'étendre vers celles du type [u], quel que soit l'espace considéré. Les lieux moyens de la constriction intra-orale sont surtout antérieurs, la zone du [u] semblant marquer une limite postérieure du recul de la langue, tandis que l'aire de cette constriction serait relativement large. Globalement, les valeurs moyennes d'aire aux lèvres seraient faibles.

En ce qui concerne le babillage, les simulations de Serkhane (2002, 2007) montrent bien le rôle majeur de la mandibule pour le babillage canonique, elles sont en grande partie compatibles avec les données disponibles, compte tenu de la variabilité morphologique.

Dans un schéma développemental vraisemblable, la première étape qui suit la maîtrise du cycle mandibulaire de babillage serait le contrôle des contacts, c'est-à-dire la capacité d'ajouter à la porteuse-mâchoire une commande réquisitionnant spécifiquement la langue ou les lèvres pour contrôler le contact labial, apical ou vélaire. La phase de babillage pourrait permettre, dans le cas de cycles mandibulaires de grande amplitude, de découvrir les propriétés de compliance (c'est-à-dire de comportement souple à s'adapter à l'environnement) de la langue contre le palais ou des lèvres l'une contre l'autre et ainsi d'*augmenter l'espace d'exploration articulatoire-acoustique* par de nouvelles commandes (celle de la pointe de la langue par exemple). Cette exploration accrue permettrait de spécifier des cartes de lien production-perception élargies, probablement capables, dans un mécanisme d'inversion, de gérer le passage progressif de la proto-syllabe vers des syllabes contrôlées en contenu, contrastant geste labial et geste lingual quelle que soit la pré-configuration du conduit vocal. L'idée serait tout simplement de chercher à apprendre des patrons de coarticulation adultes, tels que ceux décrits par Sussman *et al.* (1999), et de montrer qu'ils passent précisément par un découplage progressif de la partie vocalique et de la partie consonantique (Vilain & *al.*, 1999).

De leur côté en utilisant le modèle articulaire 3D, Kröger *et al.* (2006ab) ont développé un modèle neuronal de contrôle de la parole pour simuler l'apprentissage des articulations proto-vocaliques et proto-consonantiques du bébé au cours de la phase de babillage (Oller & *al.*, 1999). Ce modèle est inspiré des travaux pilote de Guenther *et al.* (2006). Le modèle comporte des cartes auditives, somato-sensorielles et motrices, associées à un modèle de production qui délivre un feed-back tactile, proprioceptif et auditif (cartes de Kohonen auto-organisées). Dans le cas des productions protovocaliques les possibilités articulaires sont limitées à trois gestes de base (haut-avant, bas-arrière, bas) et à deux gestes labiaux (arrondi, étiré) associés aux 3 premiers formants ; les articulations proto-consonantiques sont des réalisations CV et VC décrites par des transitions formantiques et associées à 3 groupes de paramètres articulaires correspondant à l'occlusion (labiale, apicale ou dorsale) avec sa représentation somatosensorielle (*high level motor*). La boucle de *feedback* neuronale de la production de la parole est ainsi modélisée avec succès en utilisant un réseau de neurones avec des représentations sensibles et motrices relativement simples et des associations

appropriées. Les auteurs soulignent que la représentation somatosensorielle est importante pour la définition de base des gestes linguaux proto-vocaliques et des cibles proto-consonantiques et qu'une association directe auditive-motrice permet d'obtenir de bons résultats pour modéliser les articulations proto-vocaliques.

Des percepts régularisés par l'action

Replaçons ces résultats par rapport aux théories proposées pour la perception de la parole qui peuvent se classer en trois grandes lignées. Les tenants de la théorie auditive (Nearey, 1997 ; Massaro, 1987) avancent que l'invariance est à rechercher dans les caractéristiques physiques du signal produit. Ceux qui sont en faveur de la théorie motrice (Lieberman & Mattingly, 1985; Fowler & Rosenblum, 1991) plaident pour une spécification articulatoire de l'invariance (les gestes réellement produits ou planifiés) ; les objets traités par la perception sont dans ce cas des gestes et non des sons. En quelque sorte les théories motrices envisagent la perception de la parole sans l'audition : *speech is special* (Whalen & Liberman, 1987; Whalen, 2006). Alors que production et perception sont dans les deux cas traités de manière indépendante, la troisième voie (Schwartz & al., 2002, 2007) postule que l'invariance réside dans une négociation entre locuteur et auditeur. Avec la *Perception-for-Action-Control Theory* (PACT) la perception est *mise en forme (shaped)* à la fois par un traitement auditif et par la prise en compte du processus moteur. La PACT présuppose que la perception de la parole permet non seulement à l'auditeur de suivre les vocalisations de son ou sa partenaire pour pouvoir les comprendre, mais aussi les imiter pour apprendre. En d'autres termes, la perception permet à un auditeur de spécifier le contrôle de ses productions à venir en tant que futur locuteur. Cette approche propose un processus intégré qui combine une mise en forme perceptive avec un mécanisme d'inversion permettant à l'auditeur de récupérer le contrôle articulatoire en relation avec la compréhension du but perceptif.

Les résultats des simulations menées avec des modèles articulatoires (Ménard & al.; Serkhane & al.; Kröger & al.) confirment bien le rôle primordial des mécanismes perceptuels dans le contrôle des gestes articulatoires et s'opposent, par là même, aux principes d'une théorie motrice en faveur de l'invariance des gestes. D'abord, les simulations menées avec un modèle articulatoire montrent que la croissance non uniforme du conduit vocal entraîne une modification des cartes sensori-motrices. Ensuite ils révèlent que le contrôle moteur n'atteint le stade mature qu'à la fin de l'enfance (vers 10-12 ans). Or les catégories phonologiques des enfants sont perçues par les adultes bien avant ce stade. Pour communiquer, l'enfant exploite donc les moyens dont il dispose, moyens différents de l'adulte, en encodant des représentations phonologiques communes à celles de sa communauté linguistique. Par imitation le jeune locuteur construit ses propres cartes sensori-motrices et maintient cette cartographie par la *feedback* sensoriel. L'action est ici contrainte par la perception.

9. Conclusion et perspectives

L'ontogenèse du conduit vocal renvoie à toute une série de questions qui, pour être spécialisées, peuvent être recoupées dans un cadre pluridisciplinaire. Ce conduit s'inscrit dans une charpente osseuse qui subit un remodelage très important sous le contrôle des gènes architecturaux, les gènes HOX. Les études sur le fœtus permettent d'éclairer les premiers semestres après la naissance, période cruciale pour l'apprentissage de la parole. Par les transformations qu'elle engendre, la croissance du conduit vocal constitue donc un paradigme expérimental de choix pour sa modélisation (qui s'oriente vers une approche biomécanique), les premières manifestations de la parole, les relations production-perception et la nature de l'invariance échangée entre les protagonistes de la communication.

Mais il ne faut pas oublier que la sphère oro-faciale est le siège de la respiration, succion, déglutition et mastication, fonctions qui s'installent progressivement au cours de la gestation et après la naissance. Vraisemblablement depuis plusieurs centaines de milliers d'années, ce carrefour a été détourné de ses fonctions vitales par l'Homme anatomiquement moderne et réutilisé pour la mise en place de la production de la parole. Les études de ces fonctions vitales et premières ont été systématiquement déconnectées de celles de la production de la parole. De récents travaux ont révélé toute la richesse d'une possible comparaison (MacNeilage, 1998 ; Hiiemae & al., 2002 ; Serrurier & al., 2008) : les réalisations gestuelles de la parole, au niveau du conduit vocal (coordination avec le larynx mise à part), ne seraient qu'un sous-ensemble de celles de la mastication/déglutition. Une manière nouvelle d'aborder l'émergence et la phylogenèse de la parole dans l'espèce humaine pour laquelle les crânes et les mandibules fossiles sont bien les seuls témoins architecturaux du conduit vocal de nos ancêtres.

<louis-jean.boe@gipsa-lab.inpg.fr>, <menard.lucie@uqam.ca>, <peterbirkholz@gmx.de>, <bkroeger@ukaachen.de>, <pierre.badin@gipsa-lab.inpg.fr>, <g-captier@chu-montpellier.fr>, <canaultmelanie@yahoo.fr>, <nkielwa@free.fr>

Remerciements. Cet article est une synthèse de recherches qui sont le fruit de relations et de collaborations pluridisciplinaires entre spécialistes d'anatomie, d'anthropométrie physique, d'orthodontie, d'infographie et de parole. Elles ont été stimulées en partie par le projet *CogSpeech : From Communication by Orofacial Gestures in primates to human Speech* (resp. J.-L. Schwartz) dans le cadre européen *Origin of Man, Language and Languages* et l'ANR *SkullSpeech* (resp. P. Perrier et L.-J. Boë). Des travaux complémentaires ont été associés. Nous remercions H. Vorperian pour ses images IRM d'enfants qui sont les toutes premières et J.-L. Schwartz pour ses apports théoriques fondamentaux.

Références

- Badin, P. & Fant, G. (1984). Notes on vocal tract computations. *STL QPSR* 2-3, 53-108.
- Benoît, R. (2001). Development Biology, craniofacial genetics. *Edgewise J.* 44, 9-40.
- Benoît, R. (2008). Contrôle génétique du développement vertébro-cranio-facial. Aujourd'hui et au cours de l'évolution. Les recherches génétiques portant sur le développement de l'individu peuvent-elles éclairer les mécanismes de l'évolution ? *Bulletin de l'Académie Nationale de Chirurgie Dentaire*, 51-122.
- Birkholz, P. (2005). 3D-Artikulatorische Sprachsynthese. *Logos*, Berlin.
- Birkholz, P., Jackèl, D., Kröger, B.J. (2006). Construction and control of a three-dimensional vocal tract model. *Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'06*, Toulouse, 873-876.
- Birkholz, P., Kröger, B.J. (2006). Vocal tract model adaptation using magnetic resonance imaging. *7th Int. Seminar on Speech Production (ISSP'06)*, Ubatuba, Brazil, 493-500.
- Browman, C.P., Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252. Voir aussi : <http://www.haskins.yale.edu/research/gestural.html>
- Boë, L.J. (1999). Modeling the growth of the vocal tract vowel spaces of newly-born infants and adults. Consequences for ontogenesis and phylogenesis. *Int. Congress of Phonetic Sciences* 3, 2501-2504.
- Boë, L.J., Captier, G., Granat, J., Deshayes, M.J., Heim, J.L., Birkholz, P. & al. (2008). Skull and vocal tract growth from fetus to 2 years. *8th Int. Seminar on Speech Production (ISSP'08)*, Strasbourg.
- Boë, L.J., Deshayes, M.J., Granat, J., Benoît, R., Kielwasser, N. (2007). Délimitation de zones osseuses du crâne de la face, des vertèbres cervicales et de l'os hyoïde à lumière des gènes du développement (gènes HOX et non HOX). Corroboration sur un échantillon d'enfants autour de 6 ans. *XIV^e Colloque de la Société de Biométrie Humaine*, Paris.
- Boë, L.J., Granat, J., Badin, P., Autesserre, D., Pochic, D., Zga, N., Henrich, N., Ménard, L. (2006). Skull and vocal tract growth: From newborn to adult. *XVIth Int. Congress of Phonetic Sciences, ID 1426*, Saarbrücken, Germany.

- Boë, L.J., Perrier, P., Guérin, B., Schwartz, J.L. (1989). Maximal vowel space. *Eurospeech 2*, 281-284.
- Boë, L.J., Maeda, S. (1998). Modélisation de la croissance du conduit vocal. *Journées d'Études Linguistiques, La voyelle dans tous ses états*, Nantes, 98-105.
- Boë, L.J., Ménard, L., Maeda, S. (2000). Adaptation of control strategies during the vocal tract growth inferred from simulation studies with an articulatory model. *5th Seminar on Speech Production Kloster Seeon, Germany*, 277-280.
- Brosda, S. (1999). *De la variation dans le babillage canonique : l'apprentissage sensorimoteur*. Mémoire de DEA Sciences Cognitives, ICP, Institut National Polytechnique de Grenoble.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague, Mouton.
- Canault, M. (2007a). *L'émergence du contrôle articulaire au stade du babillage. Une étude acoustique et cinématique*. Doctorat de l'Université Marc Bloch-Strasbourg II.
- Canault M. (2007b). Un chemin accidenté vers l'acquisition du timing. In Vaxelaire B., Sock R., Kleiber G., Marsac F. (eds), *Perturbations et Réajustements*, Publications de l'Université Marc Bloch, 35-50.
- Canault M., Perrier P., Sock R. (2006). L'émergence du contrôle segmental au stade du babillage : une étude acoustique. *26^e Journées d'Etude sur la Parole*, Dinard, 193-197.
- Charrier, J.B. & Creuzet, S. (2007). Embryologie de la face et dysplasies otomandibulaires. *Orthodontie Française*, 78, 7-24.
- Creuzet, S., Couly, G., Bennaceur, S., Vincent, C., Le Douarin, N.M. (2002). Negative effect of Hox gene expression on the development of the neural crest-derived facial skeleton. *Development* 129, 4301-4313.
- Couly, G., Coltey, P., Le Douarin, N. (1993). The triple origin of skull in higher vertebrates. A study in quail-chick chimeres. *Development* 117, 409-429.
- Couly, G., Creuzet, S., Benaceur, S., Vincent, Le Douarin, N.M. (2002). Interactions between HOX-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebral head. *Development* 129, 1061-1073.
- Davis, B.L. & MacNeilage, P.F. (1990). The acquisition of vowels: a case study. *Journal of Speech and Hearing Research* 33, 16-27.
- Davis, B.L. & MacNeilage, P.F. (1994). Organization of canonical Babbling. A case study. *Language and Speech* 37(4), 341-355.
- Fenart, R. (2003). Crâniographie vestibulaire. Analyse morphométrique positionnelle. *Biométrie Humaine et Anthropologie* 21, 231-284.
- Fitch, W.T., Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. of the Acoustical Society of America* 106(3), 1511-1522.
- Fowler, C.A., Rosenblum, L.D. (1991). The perception of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (eds), *Modularity and the motor theory of speech perception*, Hillsdale, NJ: Erlbaum, 33-59.
- Goldstein, U.G. (1980). *An Articulatory Model for the Vocal Tracts of Growing Children*. PhD Thesis MIT. Boston, MA. <<http://theses.mit.edu>>
- Gracco, V.L. & Lofqvist, A. (1993). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *Haskins Laboratories Status Report on Speech Research* 115-116, 17-32.
- Green J.R., Moore, C.A., Reilly, K.J. (2002). The sequential development of jaw and lip control for speech. *J. of Speech Language, and Hearing Research* 45, 66-79.
- 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, 280-301.
- Hiiemae, K.M., Palmer, J.B. (2003). Tongue movements in feeding and speech. *Crit. Rev. Oral Biol. Med.* 14(6), 413-429.
- Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. of the Acoustical Society of America* 97(5), 3099-3110.
- Honda, K. & Tiede, M.K. (1998). An MRI study on the relationship between oral cavity shape and larynx position. *5th International Conference on Spoken Language Processing* 2, 437-440.

- Kent, R.D. & Miolo, G. (1995). Phonetic Abilities in the First Year of Life. In Fletcher, P. & Macwhinney, B. (eds), *The Handbook of Child Language*, Oxford, Blackwell, 303-334.
- Kent, R.D., Osberger, M.J., Netsell, R., Hustedde, C.G. (1986). Phonetic development in identical twins differing in auditory function. *J. of Speech and Hearing Disorders* 52, 64-75.
- Kern S. (2001). Le langage en émergence. *Approche neuropsychologique des apprentissages chez l'enfant* 61(13-1), 8-12.
- Kuhl, P.K. & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. of the Acoustical Society of America* 100, 2425-2438.
- Kröger, B.J., Birkholz, P., Kannampuzha, J., Neuschaffer-Rube, C. (2006a). Modeling sensory-to-motor mappings using neural nets and a 3D articulatory speech synthesizer. *InterSpeech 2006*, Pittsburgh, USA.
- Kröger, B.J., Birkholz, P., Kannampuzha, J., Neuschaffer-Rube, C. (2006b). Learning to associate speech-like sensory and motor states during babbling. *7th Int. Seminar on Speech Production (ISSP'06)* Ubatuba, Brazil, 67-74.
- Lalévée C. & Vilain A. (2003). Development of speech frame control: a longitudinal study of oral/nasal control. *15th Int. Congr. of Phonetic Sciences*, Barcelone, 2003-2006.
- Lee, S., Potamianos, A., Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. of the Acoustical Society of America* 105(3), 1455-1468.
- Lieberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1-36.
- Liljencrants, J. & Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrasts. *Language* 48,839-862.
- Locke, J.L. (1983). *Phonological Acquisition and Change*. New-York, Academic Press.
- MacNeilage P.F. (1994). Prolegomena to a theory of the sound pattern of the first spoken language. *Phonetica*, 184-194.
- MacNeilage P.F. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499-546.
- MacNeilage P.F. & Davis B.L. (2000a). On the origin of internal structure of word forms. *Science*, 288, 527-531.
- MacNeilage P.F. & Davis B.L. (2000b). Deriving speech from non speech: a view from ontogeny. *Phonetica*, 57, 284-296.
- MacNeilage P.F. & Davis B.L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology* 11, 696-700.
- MacNeilage, P.F. & Davis B.L. (2003). Intersyllabic and word-level regularities in early acquisition. *15th International Congress of Phonetics Sciences*, Barcelone, 383-386.
- Massaro, D.W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. London: Laurence Erlbaum.
- Ménard, L. (2002). *Production et perception des voyelles au cours de la croissance du conduit vocal: variabilité, invariance et normalisation*. Doctorat de l'Université de Grenoble 3.
- Ménard, L. & Boë, L.J. (2000). Exploring vowel production strategies from infant to adult by means of articulatory inversion of formant data. *Int. Congress of Spoken Language Processing* Beijing, China, 465-468.
- Ménard, L., Davis, B., Boë, L.J. et Roy, J.P. (à paraître). Producing American-English Vowels during Vocal-Tract Growth : A Perceptual Categorization Study of Synthesized Vowels. *Journal of Speech, Language, and Hearing Research*.
- Ménard, L., Schwartz, J.L., Boë, L.J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *J. of Language, Speech and Hearing Research* 47,1059-1080.
- Ménard, L., Schwartz, J.L., Boë, L.J., Aubin, J. (2007). Production-perception relationships during vocal tract growth for French vowels: analysis of real data and simulations with an articulatory model. *Journal of Phonetics* 35, 1-19.

- Ménard, L., Schwartz, J.L., Boë, L.J., Kandel, S., Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *J. of the Acoustical Society of America* 111(4), 1892-1905.
- Nearey, T.M. (1997). Speech perception as pattern recognition. *J. of the Acoustical Society of America* 101, 3241-3254.
- Oller, D.K., Eilers, R.E., Neal, A.R., Schwartz, H.K. ; (1999). Precursors to speech infancy : the prediction of speech and language disorders. *J. of Communication Disorders* 32, 223-245.
- Oller, D.K. & MacNeilage, P.F. (1983). Development of speech production. Perspectives from natural and perturbed speech. In MacNeilage, P.F. (ed.), *The Production of Speech*, New York, Springer, 91-108.
- Munhall K.G. & Jones J.A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences* 21, 524-525.
- Peterson, G.E., Barney, H.L. (1952). Control methods used in a study of vowels, *J. of the Acoustical Society of America* 24, 175-184.
- Rossato S., Badin P., Bouaouini F. (2003). Velar movements in French: an articulatory and acoustical analysis of coarticulation. *15th International Congress of Phonetic Sciences*, Barcelone, 3141-3144.
- Savariaux, C., Boë, L.J., Perrier, P. (1997). How can the control of the vocal tract limit the speaker's capability to produce the ultimate perceptive objectives of speech ? *EuroSpeech '97*, Rhodes, Greece, 1063-1066.
- Savariaux, C., Perrier, P., Orliaguet, J.P. (1995). Compensation stratégies for the perturbation of the rounded vowel u using a lip-tube: A study of the control space in speech production. *J. of the Acoustical Society of America* 98, 2428-2442.
- Savariaux, C., Perrier, P., Orliaguet, J.P., Schwartz, J.L. (1999). Compensation strategies for the perturbation of French u using a lip tube. II. Perceptual analysis. *J. of the Acoustical Society of America* 106, 381-393.
- Schwartz, J.L., Boë, L.J., Vallée, N., Abry, C. (1997). The dispersion-focalization theory of vowel systems. *J. of Phonetics* 25, 255-286.
- Serkhane, J.E. (2005). *Un bébé androïde vocalisant : Étude et modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole*. Doct. Sciences Cognitives, INPG, Grenoble.
- Serkhane, J., Schwartz, J.L., Boë, L.J., Davis, B., Matyear, C. (2002). Motor specifications of a baby robot via the analysis of infant's vocalizations. *ICSLP2002*.
- Serkhane, J.E., Schwartz, J.L., Boë, L.J., Davis, B.L., Matyear, C.L. (2007). Infants' vocalizations analyzed with an articulatory model: A preliminary report. *J. Phonetics* 35(3), 321-340.
- Serrurier, A., Barney, A., Badin, P., Boë, L.J., Savariaux, C. (2008). Comparative articulatory modelling of the tongue in speech and feeding. *Int. Seminar on Speech Production, ISSP08*, Strasbourg.
- Sussman, H.M., Duder, C., Dalston, E., Cacciatore, A. (1999). An acoustic analysis of the development of CV coarticulation: a case study. *Journal of Speech Language and Hearing Research* 42, 1080-1096.
- Vihman, M.M. (1992). Early syllables and the construction of phonology. In Ferguson, C.A., Menn L., Stoel-Gammon, C. (eds.), *Phonological development. Models, research, implications*, Timonium, York Press, 393-422.
- Vilain A. (2001). *Apport de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole*. Doct. Sciences du Langage, Université Stendhal, Grenoble.
- Vorperian, H.K., Kent, R.D., Lindstrom, M.J., Kalina, C.M., Gentry, L.R., Yandell, B.S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *J. of the Acoustical Society of America* 117(1), 338-350.
- Whalen, D.H. & Liberman, A.M. (1987). Speech perception takes precedence over nonspeech perception. *Science* 237, 169-171.
- Whalen, D.H., Benson, R.R., Richardson, M.L., Swainson, B., & al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *J. of the Acoustical Society of America* 119, 575-581.