

Contrôle lingual en production de parole chez l'enfant de 4 ans : une méthodologie associant étude articulatoire et modélisation biomécanique

Guillaume Barbier¹ Pascal Perrier¹ Lucie Ménard² Louis-Jean Boë²

(1) GIPSA-lab, Département Parole et Cognition, UMR 5216 CNRS, Grenoble, France

(2) Laboratoire de Phonétique, Université du Québec à Montréal, Montréal, Canada
prenom.nom@gipsa-lab.grenoble-inp.fr, menard.lucie@uqam.ca

RESUME

Une méthodologie originale est proposée pour l'étude de l'émergence du contrôle moteur de la parole chez l'enfant de 4 ans. Nous associons les résultats d'une analyse articulatoire et acoustique d'un corpus de parole contenant voyelles isolées et séquences voyelle-consonne-voyelle à des simulations obtenues avec un modèle biomécanique bi-dimensionnel du conduit vocal. L'analyse de la variabilité des réalisations vocaliques d'une répétition à une autre dans un même contexte renseigne sur la précision du contrôle. L'analyse de la variabilité contextuelle renseigne sur les processus de planification mis en œuvre dans la production de la coarticulation anticipatoire. Enfin, la mise en relation des données et des résultats obtenus avec le modèle biomécanique permet d'inférer des patrons de commandes motrices pour les principaux sons et les séquences associant ces sons. Des résultats préliminaires sont présentés.

ABSTRACT

Tongue control in 4-years-old children's speech production: A methodology combining articulatory study and biomechanical modeling

An original methodology is proposed for the study of the emergence of speech motor control in 4-years-old children. We combine the results of an analysis of an articulatory and acoustic speech corpus containing isolated vowels and vowel-consonant-vowel sequences with simulations obtained using a two-dimensional biomechanical model of the vocal tract. The analysis of the variability of vowels across repetitions in a single context provides information on the accuracy of the control. The analysis of the contextual variability gives insights into the planning process used in the production of anticipatory coarticulation. Finally, the linking of data and results obtained with the biomechanical model enables to infer patterns of motor commands for the main sounds and sequences of sounds. Preliminary results are presented.

MOTS-CLES : Production de parole, développement, contrôle moteur, échographie linguale, modèle biomécanique de langue.

KEYWORDS: Speech production, development, speech motor control, ultrasound tongue imaging, biomechanical tongue model.

1 Introduction

Ce travail s'inscrit dans un projet visant à comprendre comment l'enfant de 4 ans, dans un conduit vocal différent de celui de l'adulte, contrôle sa langue (mais aussi sa mandibule et ses lèvres) pour produire les sons de sa langue maternelle. Nous présentons notre méthodologie fondée sur l'association de l'analyse de données articulatoires et acoustiques et d'interprétation de simulations réalisées avec un modèle biomécanique bi-dimensionnel du conduit vocal. Après un rapide résumé des connaissances sur l'évolution du contrôle lingual en parole au cours de la croissance, nous exposons notre protocole expérimental de recueil de données articulatoires chez l'enfant, et nous décrivons le modèle de langue développé. Nous exploitons ensuite ce modèle pour estimer les commandes motrices correspondant à la production des voyelles [i], [a] et [u] chez un enfant de 4 ans. Nous terminons par les perspectives de ce travail et les premières conclusions que nous en tirons.

2 Le développement du contrôle lingual

De nombreuses études se sont intéressées au développement de la production de la parole, en étudiant les productions des nouveau-nés et des jeunes enfants. Ces études ont permis de dégager les grandes étapes du développement phonétique et ont montré que c'est au cours de la première année de vie que se mettent en place les bases fondamentales de la communication parlée. Aux alentours de 7-8 mois, l'enfant produit des proto-syllabes, l'unité rythmique de base de la communication parlée, au sein de laquelle sont coordonnés gestes glottiques et supra-glottiques. À ce stade appelé babillage canonique, l'enfant ne disposerait pas du contrôle volontaire de la langue. Puis vers 8-10 mois, les lieux d'articulation deviennent plus diversifiés (babillage varié), on assiste alors à l'émergence du contrôle volontaire de la langue. Vers 12-13 mois, les premières lexies sont produites (pseudo-mots), les lieux d'articulation sont alors variés. Cette variété témoigne de l'existence d'un contrôle indépendant de la langue dans le plan antéro-postérieur et de la mandibule dans le plan vertical (Canault, 2007).

Si le contrôle de la langue semble émerger vers 10 mois, celui-ci n'est pas semblable à celui de l'adulte et il continue de s'affiner jusqu'à la fin de l'adolescence (Smith, 2010). Dans le cas de la production de voyelles, les données sur cette évolution ne sont cependant pas unanimes. Ce contrôle s'affinerait jusqu'à 3 ans pour certains auteurs (Vihman, 1998), et jusqu'à 10 ans pour d'autres (Kent, 1976). Ces conclusions non concordantes soulèvent quelques questions fondamentales : comment définir et mesurer la maturité du contrôle ? Est-ce lorsque l'enfant est capable de contraster suffisamment différentes voyelles, ou lorsque ces voyelles sont acoustiquement stables ?

Les données acoustiques ne reflètent qu'indirectement des mouvements des articulateurs, et ne suffisent pas, à elles seules, à comprendre les mécanismes sous-jacents à la production de parole, à savoir le contrôle moteur des articulateurs. Or l'immense majorité des études développementales se fonde sur des données acoustiques, étant donnée la grande difficulté d'accès à des données articulatoires chez l'enfant, pour des raisons éthiques et techniques. En effet, parmi les outils de mesure articulatoire, peu semblent adaptés à l'étude des enfants : la cinéradiographie est irradiante ; l'IRM dynamique ou l'EMA sont des dispositifs relativement lourds pour de jeunes enfants.

Mais parmi les dispositifs permettant le recueil de données articulatoires, un outil nous semble très prometteur pour l'étude des productions enfantines : l'échographie, non invasive, non dangereuse et ne nécessitant qu'un dispositif expérimental léger.

3 Une étude articulatoire des productions enfantines

Devant la nécessité de comprendre comment s'affine le contrôle moteur pour la production de la parole après la fin de la première année de vie, nous avons décidé de mener une étude articulatoire chez des enfants de 4 ans. Nous faisons l'hypothèse qu'à cet âge, l'enfant est capable de produire la totalité des phonèmes de sa langue maternelle : il nous sera donc possible d'étudier comment l'enfant produit ces phonèmes. Cet âge nous semble également l'âge minimum auquel une interaction prolongée avec l'enfant est possible (environ 20 minutes), ce qui permet de recueillir un corpus suffisant.

3.1 Présentation du dispositif expérimental

Afin de recueillir des données articulatoires des productions enfantines, nous avons utilisé l'échographie (en deux dimensions) dans le plan sagittal médian, en disposant la sonde échographique sous le menton du sujet. Ce dispositif permet d'observer les mouvements de la langue en temps réel et à 30 images par seconde. Les données correspondent à la distance séparant le contour lingual de la sonde émettrice de l'onde ultrasonore. Afin de placer ces données dans un référentiel tête-sonde constant, il faut soit s'assurer que la sonde reste fixe par rapport à la tête, soit repérer, puis corriger, les déplacements de la sonde par rapport à la tête et de les replacer dans un référentiel unique par un jeu de rotations/translations. La plupart des chercheurs ont opté pour la première solution, grâce à un dispositif composé d'un casque maintenant fixes la tête du sujet et la sonde échographique. Mais un tel dispositif est trop contraignant pour de jeunes enfants et nous avons préféré adopter un dispositif plus léger. Le système HOCUS (*Haskins Optically Corrected Ultrasound System*, Whalen *et al.*, 2005) permet de corriger les mouvements de tête grâce à la mesure de la position de capteurs optiques (dispositif Optotrak[®]) placés à la fois sur la tête du sujet et sur la sonde échographique. Ce système est bien adapté aux études développementales, puisqu'il n'impose pas de contrainte de mouvement à l'enfant, mais corrige les données une fois recueillies. Comme nous utilisons un tel dispositif pour corriger les données échographiques, nous avons choisi d'ajouter des capteurs afin d'exploiter pleinement l'Optotrak[®]. Nous avons ainsi ajouté un capteur sur la mandibule, qui permet de découpler l'action de la langue de celle de la mandibule, et 4 capteurs autour des lèvres, qui permettent de donner une idée de l'étirement (horizontal), de l'ouverture (verticale), et de la protrusion (avancement) des lèvres. L'expérience s'est déroulée au Laboratoire de Phonétique de l'UQAM, et 4 sujets ont participé à cette expérience : 2 filles (4 ans et 5 mois ; 4 ans et 10 mois) et 2 garçons (4 ans et 6 mois ; 4 ans et 10 mois) dont la langue maternelle est le français québécois. Les données acoustiques ont également été enregistrées via un microphone et ont été échantillonnées à 44.1kHz. Le corpus que nous avons recueilli est composé de voyelles isolées [i], [a], et [u] et de séquences Voyelle-Consonne-Voyelle (VCV), réparties en 3 blocs. Pour obtenir, de la part des enfants, qu'ils produisent ces séquences, répétées chacune 8 fois, sans trop les ennuyer, nous avons présenté cette expérience sous la forme d'un jeu de marionnettes, dans lequel on demande à l'enfant d'énoncer le nom de la

marionnette à chaque fois que celle-ci apparaît (présentation par paires). Le premier bloc est composé des voyelles isolées [i] [a] [u] et des séquences [iku] [iki] [aka] [uku]. Le second bloc est composé des séquences [isu] [isi] [asa] [usu] [ifu] [ifi] [afa] [u/u]. Le dernier bloc est composé des séquences [itu] [iti] [ata] [utu] [ipu] [ipi] [apa] [upu].

3.2 Analyse des données

La première étape vers l'analyse des données (ou prétraitement) est d'effectuer la synchronisation entre l'image échographique et la position des capteurs, qui permettra la correction des mouvements de tête. Concernant l'analyse acoustique, le signal a d'abord été sous-échantillonné à 12 kHz pour permettre une analyse plus fine dans la bande fréquentielle des formants des voyelles (300Hz-5kHz). L'analyse acoustique a été effectuée grâce à un programme de traitement du signal écrit en MATLAB®. Ce programme utilise en parallèle les méthodes de détection de formants à partir d'une analyse spectrale par prédiction linéaire (LPC, i.e. *Linear Predictive Coding*): une première méthode recherche les maxima locaux du spectre LPC dans des bandes de fréquences prédéfinies pour les formants des voix d'enfants, et mémorise leurs fréquences dans un ordre croissant ; la deuxième fait de même mais avec les fréquences des pôles du filtre LPC. Cette dernière mesure permet de s'assurer qu'aucun pôle n'est manqué dans le cas où deux pôles proches se confondent dans un seul maximum spectral. Concernant l'analyse articuloire, les formes linguales correspondant aux voyelles isolées et aux séquences VCV ont été extraites sous la forme de séquences d'images. Ces séquences ont ensuite été traitées à l'aide du logiciel *EdgeTrak* qui permet l'extraction semi-automatique du contour supérieur de la langue pour chaque image. Les contours ainsi extraits ont été ensuite replacés dans un référentiel unique grâce au système HOCUS. En plus des tâches de parole, il a été demandé aux sujets d'avalier un liquide, ce qui permet de détecter le contour du palais.

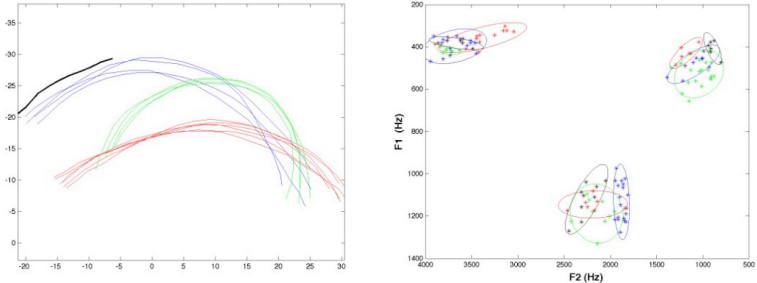


FIGURE 1 – Résultats articulatoires (à gauche) et acoustiques (à droite). À gauche, pour un même sujet : la partie antérieure du palais dur (en noir), quelques exemples de réalisation des voyelles [i] en bleu, [a] en rouge, et de la voyelle [u] en vert. À droite, dans le plan F1-F2, les ellipses de dispersion de chaque sujet (une couleur par sujet) lors de la production de voyelles isolées.

Ces données (dont l'analyse exhaustive est en cours) serviront premièrement à tester la stabilité du contrôle, en mesurant la variabilité articulatoire et acoustique des répétitions de chaque voyelle dans un même contexte. Dans le cadre de nos hypothèses sur

l'immaturation du système de contrôle moteur chez l'enfant de 4 ans, nous nous attendons à observer une grande variabilité. Cette variabilité serait la conséquence attendue de l'immaturation des représentations internes, liant commandes motrices et buts acoustiques (cf. Ménard *et al.*, 2008). Ces données considérées pour chaque voyelle dans des contextes différents serviront, dans un second temps, à mesurer la coarticulation anticipatoire intra et inter syllabique. Des représentations internes incomplètes, couvrant mal le domaine de variabilité de la production de la parole, devraient être associées à une moindre influence du contexte dans une séquence de phonèmes.

4 Un modèle biomécanique de la langue de l'enfant

Afin d'accéder à un niveau d'analyse supplémentaire, et de comprendre comment l'enfant contrôle sa langue afin de produire ces différents phonèmes, nous avons élaboré un modèle biomécanique de langue de l'enfant, permettant, pour une configuration articulaire donnée, d'estimer les commandes motrices sous-jacentes, c'est-à-dire l'information que le Système Nerveux Central doit envoyer au système périphérique - les muscles - pour atteindre une configuration articulaire et donc un produit acoustique.

4.1 Le modèle de langue adulte existant

Le modèle de langue pour l'enfant de 4 ans que nous avons élaboré repose sur un modèle de la langue de l'adulte déjà existant (Payan & Perrier, 1997). Ce modèle biomécanique bi-dimensionnel de langue de l'adulte exploite la Méthode des Éléments Finis (MEF) pour modéliser les tissus mous de la langue. Cette méthode permet de résoudre numériquement les équations traduisant la déformation d'un corps déformable (la langue) sous l'action de forces extérieures. La structuration de la langue en un ensemble d'éléments permet aussi de définir des propriétés biomécaniques locales, et donc, en particulier, d'effectuer un partage entre tissus passifs et tissus musculaires actifs. Dans ce modèle, les éléments sont de forme quadrilatérale, les équations d'élasticité continue sont discrétisées en chacun des nœuds, et les propriétés physiques de la langue sont prises en compte par deux constantes d'élasticité. Ce maillage MEF est ensuite inséré dans l'architecture du conduit vocal d'un locuteur.

Ce modèle prend en compte les muscles *généioglosse* (découpé en *généioglosse antérieure* et *postérieure*, pour répondre au fait que l'ensemble du muscle *généioglosse* n'a pas la même influence sur les déplacements de la langue), *hyoglosse*, *styloglosse*, *longitudinalis supérieur* et *inférieur*, et possède donc 7 degrés de liberté, qui rendent compte des déformations linguales dans le plan sagittal. Ce modèle est contrôlé selon l'Hypothèse du Point d'Équilibre (Feldman, 1986). Selon cette hypothèse, la force générée par un muscle dépend à la fois de la longueur de ce muscle (paramètre externe au SNC, et donc non contrôlable) et du niveau d'activation neuronale (paramètre contrôlé par le SNC) qui intervient sur la longueur du muscle. Plus un muscle est court par rapport à sa longueur au repos, plus il génère de la force (selon une relation non linéaire). Dans ce modèle, les muscles sont directement contrôlés en termes de longueur, et non en termes de force. Le passage d'un équilibre mécanique du système - dans notre cas une cible phonémique - en un autre s'effectue par un changement à vitesse constante des valeurs de longueur des 7 muscles, et correspond à une déformation non linéaire de tout le système.

4.2 Adaptation du modèle à l'anatomie de l'enfant

Afin de transformer la géométrie du modèle biomécanique et de rendre compte des différences morphologiques entre le conduit vocal d'adulte et celui de l'enfant, nous avons adapté la méthode proposée par Winkler *et al.*, 2011 de transformation du modèle entre deux adultes. Il nous faut pour cela disposer de données anatomiques sur l'enfant. Nous avons donc utilisé des données radiographiques de la tête et du cou effectuées dans le plan sagittal (fournies par l'American Association of Orthodontists et Daniel Lieberman, Harvard University). Nous avons sélectionné 7 radiographies d'enfants de 4 ans à 4 ans et 10 mois, sans pathologie ni malocclusion. Le sujet F02794, de sexe féminin et âgé de 4 ans et 6 mois nous servira de locuteur de référence. Sur ces radiographies, les contours qui déterminent la géométrie du conduit vocal ainsi que les insertions extrinsèques des muscles linguaux ont été repérés. Le pointage de ces insertions est très important, puisqu'il définit l'orientation des fibres musculaires (donc l'orientation de la force exercée) et leur longueur (donc la force générée). Le repérage des insertions extrinsèques des muscles linguaux a été validé par le Dr. Captier (spécialiste d'anatomie pédiatrique au Laboratoire d'Anatomie de Montpellier).

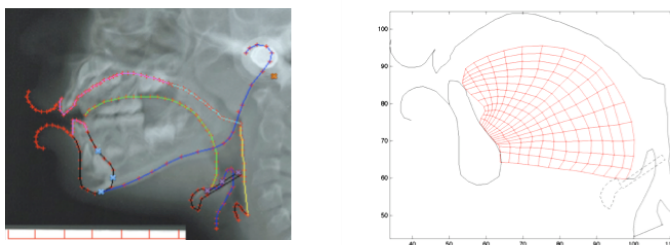


FIGURE 2 – Repérage des contours définissant la géométrie du conduit vocal et des insertions musculaires (à gauche). Le maillage MEF inséré dans le conduit vocal d'un enfant (à droite).

L'orientation des muscles linguaux chez l'enfant est différente de celle présente chez l'adulte. De plus, la longueur des muscles est différente. Nous nous attendons donc à observer des patrons musculaires différents entre les enfants et les adultes pour la production des mêmes voyelles. À ce modèle de langue nous avons ajouté un modèle de mandibule selon les principes proposés par Zandipour (2006), afin de permettre la simulation de la configuration linguale y compris en cas d'ouverture mandibulaire.

4.3 Exploitation du modèle

Nous avons ensuite généré un grand nombre de configurations articulatoires correspondant à la production de voyelles. À partir de la position neutre des articulateurs, 10000 configurations articulatoires ont été générées en échantillonnant l'espace des variables de commandes selon la méthode de Monte-Carlo intégrant l'hypothèse d'une distribution uniforme des paramètres de commande autour d'une position neutre des articulateurs. De cette étape résulte 8800 simulations, soit 8800 configurations articulatoires correspondant à des voyelles, qui couvrent l'espace articulatoire (environ 1200 simulations ont été écartées pour des raisons de contact avec

les dents ou avec le palais). Le modèle est également doté d'un modèle acoustique (Badin et Fant, 1984), qui nous permet d'obtenir le spectre associé à une configuration, dont nous ne conserverons que les valeurs des trois premiers formants (F1, F2, et F3) pour chacune des 8800 configurations articulatoires.

5 Confrontation du modèle aux données expérimentales

Afin de valider le modèle biomécanique de langue de l'enfant, mais aussi d'estimer les commandes motrices sous-jacentes à la production de voyelles chez l'enfant, nous avons comparé les simulations produites par le modèle aux configurations linguales observées en échographie. Une mise à l'échelle a été nécessaire, notre modèle de langue étant légèrement plus grand que la langue de l'enfant étudié. Pour procéder à une comparaison qualitative, nous avons approximativement replacé le contour de palais extrait de l'échographie sur le contour de palais du modèle (la morphologie des deux enfants étant différente, le contour de palais ne colle évidemment pas parfaitement). Le palais est donc utilisé comme référentiel fixe pour comparer les données.

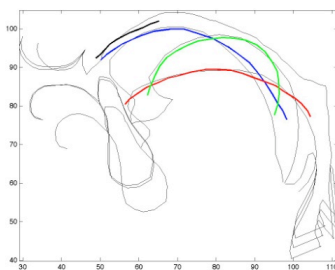


FIGURE 3 – Comparaison qualitative de 3 simulations du modèle (en noir) et de 3 configurations articulatoires observées en échographie chez un sujet, correspondant aux voyelles [i] en bleu, [a] en rouge, et [u] en vert.

On constate que, même si les contours ne sont pas exactement semblables, ils restent assez proches. Les déformations produites par le modèle à partir d'une position neutre des articulateurs sont donc relativement réalistes. De plus, il devient alors possible d'estimer les commandes motrices sous-jacentes à la réalisation de ces voyelles chez ce sujet. Ainsi, pour réaliser la voyelle [i], on constate que c'est le muscle *généioglosse postérieure* qui est fortement recruté. Pour réaliser la voyelle [u], qui nécessite également un arrondissement des lèvres, ce sont le *styloglosse* et le *longitudinalis inférieur* qui sont les principaux acteurs. Enfin, concernant la voyelle [a], réalisée mâchoire ouverte, le *hyoglosse* est fortement recruté.

6 Conclusion

L'originalité de ce travail réside en l'utilisation de nouveaux outils - échographie linguale et modélisation biomécanique - à l'étude des productions enfantines. Ce travail a déjà permis d'estimer des commandes motrices sous-jacentes à la production des voyelles [i] [a] et [u] chez un sujet, ce qui constitue un premier pas vers la compréhension du

processus sous-jacent à la production de la parole chez l'enfant : le contrôle moteur. Une fois que les données expérimentales seront analysées dans leur totalité, elles seront comparées aux données de la littérature concernant le développement de la production de la parole - et plus spécifiquement à celles concernant la coarticulation anticipatoire chez l'enfant - et permettront de participer à dresser un état des lieux de la maturité du contrôle lingual à l'âge de 4 ans. Dans une perspective plus large, d'autres études développementales utiliseront le même paradigme afin d'étudier, à différents stades ontogéniques, le contrôle de la langue pour la production de la parole dans le but d'en comprendre l'évolution.

Remerciements

Nous tenons à remercier le Dr. Captier pour la partie anatomie, Pierre Badin concernant le modèle acoustique, Marilène C. Rousseau, Amélie Prémont et Annie Brasseur concernant la partie expérimentale, l'*American Association of Orthodontists* et Daniel E. Lieberman pour l'accès aux archives radiographiques, ainsi que le projet ANR *SkullSpeech* dans lequel cette recherche s'inscrit. Nous tenons également à remercier nos sujets.

Références

- BADIN, P. & FANT, G. (1984). Notes on vocal tract computations, *STL QPSR*, 2-3, KHT, 53-108.
- CANAULT, M. (2007). *L'émergence du contrôle articuloire au stade du babillage. Une étude acoustique et cinématique*. Doctorat de l'Université Marc Bloch-Strasbourg II.
- FELDMAN, A. G. (1986). Once more on the Equilibrium-Point Hypothesis (λ model) for motor control. *Journal of Motor Behavior*, 18 (1), 17-54.
- KENT, R. D. (1976). Anatomical and Neuromuscular Maturation of the Speech Mechanism. Evidence from Acoustic Studies. *Journal of Speech and Hearing Research*, 19, 421-447.
- MENARD, L., PERRIER, P., SAVARIAUX, C., AUBIN, J. & THIBEAULT, M. (2008). Compensation strategies for a lip-tube perturbation of French [u] : an acoustic and perceptual study of 4-year-old children. *Journal of the Acoustical Society of America*, 124, 1192-1206.
- PAYAN, Y. & PERRIER, P. (1997). Synthesis of V-V Sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, 22, 185-205.
- SMITH, A. (2010). Development of neural control of orofacial movements for speech. In Hardcastle, W. J., Laver, J. & Gibbon, F. E. (Eds.), *The Handbook of Phonetic Sciences* (2nd Ed., pp. 251-296).
- VIHMAN, M. M. (1998). Later phonological development. In J.E. Bernthal & N.W. Bankson (Eds.) *Articulation and phonological disorders*, (4th Ed., pp. 113-147).
- WHALEN, D. H., ISKAROUS, K., TIEDE, M., OSTRY, D. J., LEHNERT-LEHOULLIER, H., VATIKIOTIS-BATESON, E. & HAILEY, D. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech Language and Hearing Research*, 48 (3), 543-553.
- WINKLER, R., FUCHS, S., PERRIER, P. & TIEDE, M. (2011). Biomechanical tongue models: An approach to studying inter-speaker variability, *Proceedings of Interspeech 2011*, 273-276.
- ZANDIPOUR, M. A. (2006). *Modeling Investigation of Vowel to Vowel Movement Planning in Acoustic and Muscle Spaces*. Unpublished Doctoral dissertation, Boston University.