

Développement et mise en œuvre de marqueurs fiduciaires pour l'imagerie IRM du conduit vocal en vue de la modélisation articulatoire de la parole

Pierre Badin¹ Arielle Koncki¹ Julián Andrés Valdés Vargas¹

Laurent Lamalle² Christophe Savariaux¹

(1) GIPSA-Lab (DPC / ICP), UMR 5216, CNRS – Université de Grenoble, France

(2) SFR1 RMN Biomédicale et Neurosciences (Unité IRM Recherche 3 Tesla), INSERM – CHU de Grenoble, France

Pierre.Badin@gipsa-lab.grenoble-inp.fr, konckia@minatec.inpg.fr

RESUME

L'IRM permet de caractériser la forme et la position des articulateurs de la parole, mais pas de suivre l'évolution des points de chair, car il n'existe pas de repères associés de manière fiable aux tissus hautement déformables de ces articulateurs. Or ces informations sont intéressantes pour la connaissance des propriétés biomécaniques de ces organes ainsi que pour la modélisation des relations entre modalités de mesure telles que l'IRM et l'articulographie électromagnétique. Nous avons donc attaché aux articulateurs d'un locuteur des marqueurs fiduciaires constitués de polymères non toxiques et visibles à l'IRM, et enregistré un corpus d'images IRM médiosagittales. Les contours des articulateurs et les coordonnées des marqueurs déterminés manuellement ont été analysés. Nous avons observé une déviation des marqueurs de l'ordre de 0.6 à 1.5 cm par rapport à une hypothèse d'élasticité répartie de manière uniforme. Nous avons par ailleurs montré que les marqueurs peuvent prédire les contours des articulateurs avec une explication de la variance autour de 85 %, et une erreur RMS entre 0.08 et 0.15 cm, à comparer avec 74 à 95 %, et 0.07 à 0.14 cm, pour les modèles articulatoires originaux.

ABSTRACT

Development and implementation of fiduciary markers for vocal tract MRI imaging and speech articulatory modelling

MRI allows to characterize the shape and position of speech articulators, but not to track the evolution of flesh points, since there are no markers reliably associated with the highly deformable tissues of these articulators. This information is however interesting for the knowledge of the biomechanical properties of these organs as well as for modelling the relations between measurement modalities such as MRI or the electromagnetic articulography. We have therefore attached to a speaker's articulators fiduciary markers made of non toxic polymers visible by MRI, and recorded a corpus of MRI midsagittal images. The articulators' contours and the markers' coordinates manually determined have been analysed. We have observed a departure from the hypothesis of uniformly distributed elasticity ranging from 0.6 to 1.5 cm. Besides, we have shown that the markers can predict the articulators' contours with a variance explanation around 85 %, and an RMS error from 0.08 to 0.15 cm, compared to 74 à 95 %, and 0.07 à 0.14 cm for the original articulatory models.

MOTS-CLES : Modèle articulatoire, point de chair, IRM, marqueur fiduciaire.

KEYWORDS : Articulatory model, flesh point, MRI, fiduciary marker.

1 Introduction

Il existe de nombreux modèles articulaires pour la parole (*cf.* Bailly, Badin, Revéret & Ben Youssef (in press)). Les modèles géométriques et les modèles fonctionnels représentent les contours ou surfaces géométriques des articulateurs, que ce soit dans le plan médiosagittal ou en volume. Les modèles biomécaniques simulent les propriétés de déformation des tissus déformables qui composent les articulateurs : les nœuds de leurs réseaux bi- ou tri-dimensionnels peuvent être associés à des points de repère physiologiques – que l'on appellera *points de chair* (anglais : *flesh points*) – de ces tissus (Brunner, Fuchs & Perrier (2011)).

D'un autre côté, les données articulaires à partir desquelles ces modèles peuvent être développés ou qui peuvent servir à les valider sont fournies soit par les méthodes d'imagerie médicale classique (IRM, scanner densitométrique) soit par des méthodes spécifiques à la parole comme l'articulographie électromagnétique (EMA, *cf.* Kaburagi, Wakamiya & Honda (2005)). Cette dernière permet de déterminer les coordonnées d'une quinzaine de points de chair auxquels sont attachées les petites bobines réceptrices du champ électromagnétique de l'articulographe, avec une bonne résolution temporelle de quelques centaines de Hz, mais un nombre limité de points. À l'inverse, l'IRM fournit des images médiosagittales ou volumiques de bonne résolution spatiale (de l'ordre du pixel / mm), mais avec une résolution temporelle de quelques dizaines de Hertz au maximum pour l'IRM dynamique. L'autre inconvénient de l'IRM classique est le manque d'information sur le comportement des points de chair, comme l'illustre la Figure 1, où l'on voit parfaitement le contour médiosagittal de la langue, mais où il est très difficile de localiser la pointe de la langue. L'IRM cinématique *tatouée* (anglais : *tagged cine-MRI*) constitue une approche intéressante de ce point de vue parce qu'elle fournit potentiellement le déplacement de n'importe quel point de la coupe analysée (Parthasarathy, Prince, Stone, Murano & NessAiver (2007)). Cependant le tatouage des images IRM est éphémère (rémanence de l'ordre de la seconde), ce qui est une condition rédhibitoire à son utilisation pour des modèles statistiques nécessitant l'enregistrement de plusieurs dizaines d'articulations avec des positions de marqueurs identiques, sans compter la nécessité de répéter de manière fiable la même séquence quelques dizaines de fois (Parthasarathy *et al.* (2007)).



FIGURE 1: Image d'articulation /k/.

Nous sommes donc confrontés à un double problème : d'une part un manque de connaissance sur les propriétés d'élasticité et de déformation locale des tissus des articulateurs, et d'autre part la difficulté de lier par la modélisation les points de chairs et les contours pour piloter les modèles à bonne résolution spatiale à partir de données à bonne résolution temporelle issues de l'EMA (Badin, Tarabalka, Elisei & Bailly (2010)). Les principaux objectifs de notre étude étaient donc : (1) la mise en œuvre de marqueurs fiduciaires visibles à l'IRM et fixés de manière constante

pendant toute la séance d'enregistrement IRM, permettant ainsi d'obtenir à la fois des données de contours de bonne résolution spatiale et un certain nombre de points de chair associés ; (2) l'analyse des relations entre ces deux types de données et la construction des modèles linéaires qui les mettent en correspondance.

2 Conception et mise en œuvre de marqueurs fiduciaires visibles à l'IRM

Un marqueur fiduciaire est un objet utilisé dans le champ de vision d'un système d'imagerie et qui apparaît dans l'image produite, pour une utilisation en tant que point de référence ou de mesure. Nous avons donc décidé de mettre au point des marqueurs fiduciaires visibles à l'IRM afin de pouvoir déterminer sur les mêmes images les contours des articulateurs – que ce soit en 2D ou en 3D – et les coordonnées des points de chair auxquels sont attachés ces marqueurs. Le cahier des charges de ces marqueurs était donc : (1) ne pas être toxiques ; (2) émettre assez de signal pour être visibles à l'IRM avec les protocoles généralement utilisés pour l'imagerie de l'articulation tout en conservant des tailles compatibles avec une articulation à peu près normale de la parole, comme pour l'EMA; (3) pouvoir être collés avec la colle *Cyano Veneer* à base de cyanoacrylate ou le ciment dentaire *Fuji I* à base d'acide polyacrylique.

Après divers essais infructueux de conditionnement en gels ou en gélules de substances susceptibles de donner du signal telles que le jus de myrtille ou différentes huiles, nous avons utilisé les gels décoratifs *GelGems®*: constitués de polymères résistants thermoplastiques qui contiennent des huiles minérales, ils donnent un bon signal à l'IRM. La fiche toxicologique du fournisseur *Design Ideas* indique qu'ils ne sont pas toxiques, même s'ils ne doivent pas être avalés dans un usage normal.

Les plaques de différentes épaisseurs de ces gels peuvent être très facilement découpées en petits blocs parallélépipédiques. Divers essais sur fantôme pour trouver un compromis entre taille des blocs – et donc perturbation de l'articulation – et visibilité ont débouché sur des blocs de 4 mm de hauteur dans le sens perpendiculaire à la surface de l'articulateur, et de 5×5 mm dans la direction de la surface (l'épaisseur de coupe médiosagittale IRM étant de 4 mm, il était important que la dimension transverse soit au moins aussi grande, en particulier au cas où les mouvements des articulateurs du locuteur ne se feraient pas exactement de manière symétrique et ne maintiendraient pas les marqueurs dans le plan médiosagittal). Ces marqueurs ont été ensuite testés sur le locuteur PB pour lequel sont déjà disponibles des données IRM (Badin & Serrurier (2006)) et des données EMA (Badin *et al.* (2010)). L'un des objectifs de cette étude étant de tester les relations entre les points de chair – qui peuvent être déterminés par EMA pour de la parole courante – et les contours médiosagittaux des articulateurs, nous avons collé les marqueurs dans le plan médiosagittal en des points similaires à ceux utilisés pour les bobines EMA, avec quelques points supplémentaires, comme illustré à la Figure 2 : quatre sur la langue (pointe, lame, corps, arrière), et six sur le profil extérieur médiosagittal (limites peau / vermillon pour les lèvres inférieure et supérieure, pointe et creux du nez, menton et creux du menton).

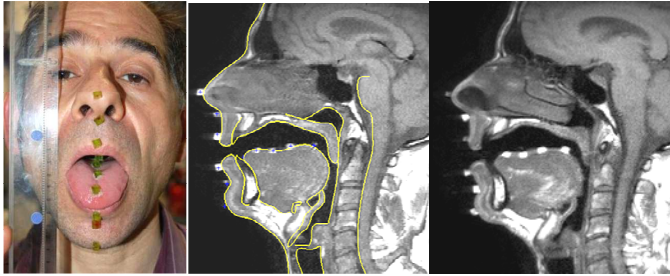


FIGURE 2 : Illustration des positions des marqueurs et des tracés de contours et marqueurs pointés (gauche : photo de face ; milieu : coupe médiosagittale avec contours (jaune) et marqueurs (croix bleues) ; droite : coupe sagittale adjacente complémentaire.

Les enregistrements IRM ont été réalisés à l'aide d'un imageur PHILIPS ASHIEVA 3T TX. La mise en place des marqueurs étant délicate, et les mouvements des articulateurs du locuteur n'étant pas forcément exactement symétriques par rapport au plan médiosagittal, les marqueurs peuvent se retrouver imparfaitement alignés dans le plan médiosagittal. Ainsi, pour pallier le risque de ne pas retrouver tous les marqueurs sur la même image, deux coupes jointives ont été acquises pour chaque articulation : l'une dans le plan médiosagittal et l'autre dans un plan sagittal adjacent où les marqueurs apparaissent le mieux (voir Figure 2). Chaque image est une coupe de 4 mm d'épaisseur, avec un champ de vue de 256×256 mm, et une résolution de 1mm par pixel. La durée d'acquisition était de l'ordre de 8 secondes par image.

On notera sur la Figure 2 que les taches blanches correspondant aux marqueurs semblent en quelque sorte poinçonnées dans la surface de la langue. Les marqueurs de la langue et notamment les antérieurs sont les plus sujets à cet effet, alors que ceux du visage semblent bien posés en surface. Cet effet coïncide avec les articulations pour lesquelles un sillon central existe. Un effet de volume partiel pourrait expliquer cette observation : on intègre les signaux d'origine microscopique sur une épaisseur de 4 mm, la composition étant inhomogène sur cette épaisseur (par exemple, langue creusée en U en coupe coronale). Des expériences complémentaires pour essayer de comprendre ce qui se passe mécaniquement seraient à envisager par la suite.

3 Données de contours et marqueurs

Afin de permettre la construction de modèles articulatoires linéaires (*cf.* Badin & Serrurier (2006)), nous avons enregistré un corpus français complet comprenant toutes les voyelles orales et nasales /a e e i y u o ø ɔ œ ã ã ã ɔ̃/, ainsi que les consonnes /p t k f s ʃ m n ʁ l/ soutenues pendant 16 secondes dans les mêmes contextes vocaliques VCV (voyelle-consonne-voyelle). Un sous-ensemble restreint aux contextes /a e e i u o/ a été utilisé pour la présente étude. Suivant la procédure décrite dans Badin & Serrurier (2006), les contours des différents organes déformables ont été édités manuellement à l'aide de courbes splines à partir de l'image médiosagittale, et les contours rigides ont été positionnés manuellement par rototranslation. Les structures crâniennes ont servi à aligner les contours de toutes les articulations sur le même repère, en permettant ainsi de compenser les mouvements de la tête du locuteur dans la direction sagittale. Les coordonnées des centres des taches produites par les marqueurs ont été marquées manuellement, en utilisant

l'image sagittale adjacente lorsque ces taches n'étaient pas assez visibles, sous l'hypothèse que cette image n'est pas trop différente de l'image médiosagittale. Le contour de la langue a été tracé depuis la jonction avec l'épiglotte jusqu'au contact de l'apex avec la mandibule, et quand c'était possible en incluant le contour de la cavité sublinguale jusqu'à l'attachement du frein de la langue à la mandibule. Puisqu'il n'y avait pas de marqueur sur l'extrémité de la langue, ce point a été repéré et tracé séparément par l'opérateur. Les contours de langue sont ré-échantillonnés par 150 points équirépartis entre la racine et la pointe de la langue et sont similaires à des points de chair dans l'hypothèse d'une élasticité uniforme tout le long du contour ; de même les lèvres ont été ré-échantillonnées par 100 points chacune.

Pour chaque articulation et chaque contour, nous avons ensuite déterminé les projections des centres des marqueurs sur les contours, et les indices des points les plus proches de ces projections, afin d'évaluer le glissement entre les véritables points de chair et les points équirépartis. La Figure 3 montre, sur les tracés de la langue et des lèvres moyennés sur le corpus, les points correspondants à ces indices pour chaque articulation : on voit apparaître les zones associées à chaque marqueur ; l'étendue de ces zones est assez importante – de 0.6 à 1.5 cm –, ce qui montre que l'hypothèse d'élasticité uniforme n'est pas vérifiée. Par ailleurs, nous avons observé que la distance entre les marqueurs et les contours peut aller jusqu'à 0.4 cm, plus grande pour la lèvre supérieure, et pour les deux marqueurs à l'arrière de la langue.

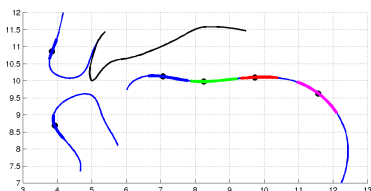


FIGURE 3 : Zones affiliées à chaque marqueur. Les étendues associées suivant le contour moyen sont de 1.2, 1.4, 1.1, et 1.5 cm pour les marqueurs linguaux (de l'apex vers la racine) et de 0.6 et 0.8 cm respectivement pour les lèvres supérieure et inférieure.

Il faut noter que la position des points rééchantillonnés sur les contours dépend en particulier des points d'ancrage qui définissent les extrémités de ces contours. Une erreur éventuelle sur ces points se répercute, d'une manière plus ou moins forte en fonction de leur distance, sur la position des autres points échantillonnés sur ce contour. La fiabilité de la position des points d'ancrage est donc cruciale. Citons par exemple le marqueur définissant le début des contours de la lèvre inférieure qui se trouve au centre du creux du menton : sa projection sur le contour peut être assez instable. Un faible bruit sur la position du pointage du marqueur peut entraîner une erreur de position du projeté importante et donc avoir des conséquences non négligeables sur l'échantillonnage. Il en est de même pour la langue : les extrémités sont définies par l'expert, mais peuvent être sujettes à des erreurs d'interprétation, en particulier au niveau de la pointe de la langue.

4 Modèles articulatoires et relations entre contours et marqueurs

Nous avons ensuite établi des modèles articulatoires par Analyse en Composantes Principales (ACP) guidée, suivant la méthode de Badin & Serrurier (2006) réduite au plan médiosagittal. Pour la langue, les composantes *JH*, *TB*, *TD*, *TTH* et *TTV* correspondent respectivement à l'influence de la mâchoire, au déplacement du corps de la langue, à la forme arrondie ou plate du corps, et aux mouvements horizontaux et verticaux de l'apex. Pour les lèvres, les mesures de protrusion (*UL_pro* et *LL_pro*) et de hauteur (*UL_hei* et *LL_hei*) de chaque lèvre ont été utilisées comme composantes complémentaires à la mâchoire. Enfin, le mouvement principal du velum dans une direction haut-arrière / bas-avant contrôlé par *VH* est complété par un petit mouvement horizontal (*VS*). Le Tableau 1 indique la variance expliquée cumulée pour chacune des composantes, et l'erreur quadratique moyenne de reconstruction des contours. La Figure 4 illustre les nomogrammes associés à chacune des composantes des divers articulateurs. Ces données sont compatibles avec les modèles précédents sur le même locuteur (Badin & Serrurier (2006)), même si nous avons noté de petites différences pouvant vraisemblablement être attribuées à la différence de systèmes de coordonnées. Dans l'ensemble, la reconstruction est précise, avec une variance expliquée entre 74 et 95 %, et une erreur quadratique moyenne (RMSE) entre 0.07 et 0.14 cm. Les moins bonnes performances sur la lèvre inférieure pourraient être attribuées à l'imprécision du marqueur situé au creux entre la lèvre inférieure et le menton.

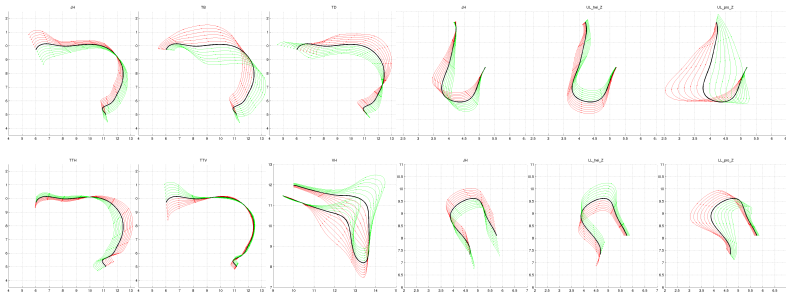


FIGURE 4 : Nomogrammes articulatoires de langue, lèvres et velum pour des variations des prédicteurs de -3 à +3 par pas de 0.5. La moyenne apparaît en noir épais, les contours pour les valeurs négatives en vert, et pour les valeurs positives en rouge.

Par ailleurs, une ACP standard a montré qu'il faut 5 composantes pour expliquer 97.2 % de la variance des 150 points de la langue, de même que pour expliquer 98.9 % de la variance des projections des 4 marqueurs, ce qui est encourageant pour la prédiction des contours à partir des marqueurs. Dans une approche complémentaire, nous avons donc développé des modèles articulatoires contrôlés par les coordonnées des marqueurs (complétées par celle de l'incisive inférieure). Pour chaque organe – langue et lèvres – ces coordonnées sont modélisées par ACP ; les prédicteurs ainsi obtenus – décorrélés entre eux – sont imposés comme paramètres de contrôle de modèles de contours obtenus par régression linéaire multiple. Les performances générales de ces modèles indiquées sur la dernière ligne du TABLEAU 1, sont proches de celles

des modèles originaux, avec une variance expliquée autour de 86 %, et une RMSE entre 0.08 et 0.15 cm, même si une certaine dégradation est notable. La FIGURE 5 illustre certains résultats. Pour le /f/, la compression de la lèvre inférieure, vraisemblablement associée un phénomène non-linéaire, est mal représentée. L'apex de langue pour le /l^h/ est également mal reconstruit, tandis que le /k^a/, de forme plus régulière, est très bien reconstruit.

Tongue	Cumvar	RMSE	UpperLip	Cumvar	RMSE	LowerLip	Cumvar	RMSE	Velum	Cumvar	RMSE
JH	14,6	0,43	JH	7,6	0,21	JH	25,3	0,23	VH	84,5	0,08
TB	54,8	0,31	UL_pro	70,6	0,12	LL_pro	50,2	0,19	VS	89,6	0,07
TD	82,0	0,20	UL_hei	79,8	0,10	LL_hei	73,7	0,14			
TTH	90,8	0,14									
TTV	95,4	0,10									
From Mrks	88,7	0,15		85,7	0,08		86,5	0,10			

TABLEAU 1: Variance expliquée pour les modèles articulatoires (Cumvar : variance expliquée cumulée en % ; RMSE en cm)

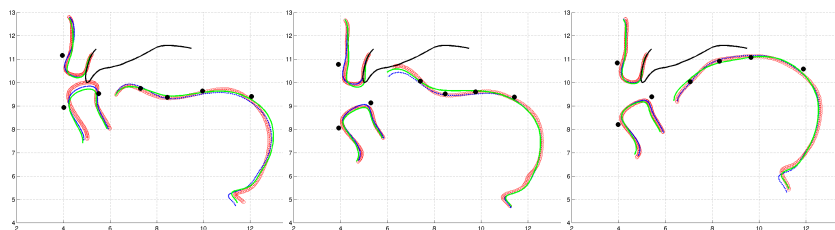


FIGURE 5 : Contours originaux (rouge), modélisés (bleu), et reconstruits (vert) à partir des marqueurs (points noirs) (de gauche à droite: /f^h/, /l^h/, /k^a/).

5 Conclusions et perspectives

Nous avons développé avec succès des marqueurs fiduciaires constitués de polymères non toxiques visibles à l'IRM qui peuvent être attachés en différents points de chair des articulateurs. Une base d'images IRM a été acquise pour un locuteur pour un corpus représentatif du français, et les contours des articulateurs ainsi que les marqueurs associés ont été tracés. Nous avons constaté que l'hypothèse d'uniformité d'élasticité des organes le long de leurs contours n'est pas complètement vérifiée, avec des zones de glissement des marqueurs par rapport aux contours de l'ordre de 0.6 à 1.5 cm. Nous avons par ailleurs montré que les coordonnées des marqueurs et de la mandibule peuvent prédire les contours des articulateurs avec une explication de la variance autour de 85 %, et une erreur RMS entre 0.08 et 0.15 cm, à comparer avec 74 à 95 %, et 0.07 à 0.14 cm, pour les modèles articulatoires originaux.

Dans le court terme, il est nécessaire d'améliorer la technique décrite : tester des marqueurs plus petits et définir plus précisément les points d'ancrage. Ces résultats confirment ensuite la validité de l'approche qui consiste à piloter des têtes parlantes à partir de données articulatoires.

(Badin *et al.* (2010)). Ils confirment aussi les études antérieures sur les relations entre points de chair et contours d'articulateurs tout en les étendant : Badin, Baricchi & Vilain (1997) ont travaillé à partir d'un court film cinéradiographique et donc d'une quantité de données limitée, sans traiter les lèvres, et n'ont pas abordé le problème de l'élasticité non homogène des contours ; les travaux de Kaburagi & Honda (1994) à partir d'imagerie ultrasonique et d'articulographie électromagnétique étaient limités à la partie de la langue visible par la sonde échographique, et n'abordaient pas non plus le problème d'élasticité.

Il sera également nécessaire de traiter d'autres locuteurs pour confirmer les résultats préliminaires présents, et il serait intéressant d'utiliser les marqueurs intermédiaires pour construire des modèles articulatoires qui prennent mieux en compte le déplacement des points de chair. Enfin, ces nouvelles données pourraient être confrontées à des modèles biomécaniques de la langue et des lèvres.

Remerciements

Nous remercions la société *Design Ideas* qui a gracieusement mis à notre disposition un lot varié d'éléments en gel *Gelgems®* et nous a fourni les fiches de toxicologie du matériau. Ce travail a été en partie financé par le projet ANR-08-EMER-001-02 ARTIS.

Références

- Badin, P., Baricchi, E. & Vilain, A. (1997). Determining tongue articulation: from discrete fleshpoints to continuous shadow. In *5th EuroSpeech Conference*, vol. 1, pp. 47-50. Rhodes, Greece, University of Patras, Wire Communication Laboratory, Patras, Greece.
- Badin, P. & Serrurier, A. (2006). Three-dimensional modeling of speech organs: Articulatory data and models. In *IEICE Technical Report*, vol. Vol. 106, No 177, SP2006-26, pp. 29-34. Kanazawa, Japan, The Institute of Electronics, Information, and Communication Engineers.
- Badin, P., Tarabalka, Y., Elisei, F. & Bailly, G. (2010). Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, **52**(6), 493-503.
- Bailly, G., Badin, P., Revéret, L. & Ben Youssef, A. (in press). Sensori-motor characteristics of speech production. In *Audiovisual speech* (E. Vatikiotis-Bateson, G. Bailly & P. Perrier, editors), Cambridge, UK: Cambridge University Press.
- Brunner, J., Fuchs, S. & Perrier, P. (2011). Supralaryngeal control in Korean velar stops. *Journal of Phonetics*, **39**(2), 178-195.
- Kaburagi, T. & Honda, M. (1994). Determination of sagittal tongue shape from the positions of points on the tongue surface. *Journal of the Acoustical Society of America*, **96**(3), 1356-1366.
- Kaburagi, T., Wakamiya, K. & Honda, M. (2005). Three-dimensional electromagnetic articulography: A measurement principle. *The Journal of the Acoustical Society of America*, **118**(1), 428-443.
- Parthasarathy, V., Prince, J.L., Stone, M., Murano, E.Z. & NessAiver, M. (2007). Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *The Journal of the Acoustical Society of America*, **121**(1), 491-504.