

VisArtico : visualiser les données articulatoires obtenues par un articulographe

Slim Ouni^{1,2} Loïc Mangeonjean²

(1) LORIA - UMR 7503, 54506 Vandoeuvre-lès-Nancy Cedex

(2) Université de Lorraine

Slim.Ouni@loria.fr

RESUME

Dans cet article, nous présentons VisArtico, un logiciel de visualisation de données articulatoires obtenues par un articulographe, l'AG500. Ce logiciel permet de visualiser les positions des capteurs et de les animer simultanément avec l'acoustique : l'utilisateur a la possibilité de visualiser le contour de la langue et des lèvres. Il permet également de trouver le plan midsagittal du locuteur, et déduire la position du palais, si cette information est absente lors de l'acquisition. De plus, VisArtico offre la possibilité d'étiqueter phonétiquement les trajectoires. D'autres fonctionnalités sont également décrites. L'objectif est de fournir un outil efficace de visualisation de données articulatoires qui peut être utile à toute personne étudiant la production de la parole.

ABSTRACT

VisArtico : visualizing articulatory data acquired by an articulograph

In this paper, we present VisArtico, visualization software of articulatory data acquired by an articulograph, AG500. The software allows displaying the positions of the sensors that are simultaneously played with the speech signal. It is possible to display the tongue contour and the lips contour. The software helps to find the midsagittal plane of the speaker and find the palate contour. In addition, VisArtico allows labeling phonetically the articulatory data. Our main goal is to provide an efficient tool to visualize articulatory data for researchers working in speech production field.

MOTS-CLES : Articulographe, visualisation, production de la parole, conduit vocal, EMA.

KEYWORDS : Articulograph, visualization, speech production, vocal tract, EMA.

1 Introduction

L'étude de la production de la parole se fait aussi bien au niveau acoustique qu'au niveau articulatoire. Contrairement aux techniques d'acquisition acoustique, les techniques d'acquisition articulatoire sont des tâches délicates qui rendent la collecte de données difficile. L'imagerie aux rayons-X a d'abord été utilisée durant plusieurs décennies (Ghazali, 1977 ; Bothorel et al., 1986). Mais son utilisation a été fortement réduite, voire même interdite dans certains pays, compte tenu des risques qu'elle représente pour la santé. Aujourd'hui, d'autres techniques, comme l'imagerie à résonance magnétique (IRM), l'utilisation de techniques par échographie (US) ou encore l'électromagnétographie (EMA), sont utilisées. Les données obtenues par ses techniques permettent de suivre le mouvement de la langue, de la mâchoire et des lèvres. Certaines techniques offrent une bonne résolution spatiale (IRM), et une bonne résolution temporelle (EMA).

Dans ce papier, nous nous intéressons à la visualisation des données articulatoires obtenues par l'EMA. C'est une technique d'acquisition de données articulatoires consistant à suivre la position de petites bobines électromagnétiques collées sur les articulateurs de la parole. La position de ces bobines est calculée en mesurant les courants électriques produits par leurs déplacements lorsqu'elles sont soumises à plusieurs champs magnétiques de faible intensité, généralement trois, dont les caractéristiques dépendent du temps (Perkell et al. 1992 ; Hoole, 1996). Cette technique ne présente aucun danger connu pour la santé des sujets. Nous nous intéressons en particulier à la visualisation des données obtenues par l'articulographe 3D, AG500 (Carstens Medizinelektronik GmbH, Allemagne) qui permet de mesurer le déplacement simultané de 12 capteurs à une fréquence de 200 Hz. Les capteurs sont collés généralement sur la langue, les lèvres, et sur une des incisives. Trois capteurs sont généralement utilisés pour annuler les effets du mouvement de la tête et situer les données articulatoires par rapport à un repère fixe. En effet, le locuteur peut avoir un mouvement assez libre dans le cube de l'articulographe. Après un post-traitement des données, le logiciel Calpos, fourni avec l'articulographe permet d'obtenir en sortie des données avec 5 degrés de liberté pour chaque capteur (3 coordonnées spatiales en x, y et z et 2 coordonnées angulaires). Les chercheurs doivent par la suite utiliser leurs propres moyens (scripts, programmes, etc.) pour interpréter ces données en fonction de l'objectif de l'étude. Malheureusement cela limite énormément l'exploitation de l'articulographe pour les chercheurs qui n'ont pas forcément de connaissances avancées en informatique. Par exemple, les phonéticiens sont très intéressés par les données articulatoires, mais en l'absence d'un logiciel adéquat pour visualiser ces données, l'utilisation de l'articulographe reste très réduite.

Quelques logiciels qui permettent de visualiser les données EMA existent, comme EMATOOLS (Nguyen, 2000) ou JustView (Carstens, 2006). Malheureusement, certains logiciels ne sont plus maintenus ou disponibles, ou bien il faut les utiliser avec d'autres logiciels, comme MATLAB, par exemple. Ce dernier nécessite un effort de programmation de script, pour arriver à visualiser correctement les données.

Pour ces raisons, nous proposons le logiciel VisArtico qui permet de visualiser des données articulatoires obtenues par un articulographe. Le logiciel a été pensé de telle façon que les données fournies par l'articulographe soient directement utilisées. VisArtico permet de visualiser les gestes articulatoires simultanément avec l'acoustique correspondante. Ce logiciel ne permet pas uniquement de visualiser les capteurs mais il enrichit l'information visuelle en indiquant clairement et graphiquement les données relatives à la langue, aux lèvres et à la mâchoire. Dans la suite de cet article, nous présentons le logiciel et ses principales fonctionnalités.

2 Interface utilisateur de *VisArtico*

2.1 Description de l'interface

VisArtico étant un logiciel de visualisation, l'interface utilisateur en est une composante importante. Nous présentons dans la suite, une description de l'interface graphique du logiciel (voir figure 1). L'interface est composée de trois modules :

(1) une représentation spatiale 3D des capteurs

La position de chaque capteur dans l'espace ainsi que leur orientation sont représentées dans un repère tridimensionnel qui est celui de l'articulographe. L'utilisateur a ensuite la possibilité de relier certains capteurs entre eux par des segments ou des splines, si la disposition des capteurs exploite la troisième dimension. Il est également possible de montrer toutes les positions des capteurs sous forme de nuage de points. Cette vue permet d'exploiter pleinement l'AG500 pour représenter les données en 3D. Rappelons que les versions précédentes, AG100 et AG200 ne permettent de représenter les données que dans un plan.

(2) une représentation midsagittale des capteurs

La représentation midsagittale permet de présenter dans un repère 2D une vue midsagittale du conduit vocal qui laisse apparaître le contour de la langue, des lèvres, et de la mâchoire. La forme de la mâchoire est approximative et a pour unique but une meilleure interprétation des données. La forme du palais peut également être affichée si l'information est disponible. Enfin, une vue de face des lèvres est aussi disponible pour observer le degré d'ouverture des lèvres. Cette vue offre une représentation des données qui intéresse une grande partie de chercheurs dans le domaine de la production de la parole, comme les phonéticiens par exemple.

(3) une représentation temporelle des trajectoires

Ce module permet de visualiser les trajectoires articulatoires d'un ou de plusieurs capteurs selon toutes les informations disponibles (position et orientation) ainsi que le signal acoustique et l'étiquetage phonétique correspondants. Les trajectoires peuvent être affichées dans un seul ou plusieurs panels. Cette représentation temporelle permet également de sélectionner des segments de temps. Cette sélection peut se faire au niveau des phonèmes si l'étiquetage phonétique est disponible.

VisArtico permet d'animer les trois vues simultanément. Le déplacement du curseur dans le module temporel change instantanément la vue spatiale correspondante.

2.2 Etiquetage

Ajouter de l'information phonétique est très utile pour une meilleure interprétation des gestes articulatoires. Visartico permet l'étiquetage des données articulatoires. L'utilisateur peut le faire directement par l'interface utilisateur du logiciel, ou bien il peut importer un étiquetage réalisé par un autre logiciel comme Winsnoori (Laprie, 2002), ou Praat (Boersma, 2001), ou encore un format propre très simplifié qui est

représenté par un temps de début, un temps de fin et le symbole phonétique correspondant. Nous envisageons d'ajouter d'autres formats provenant d'autres logiciels qui sont très utilisés par la communauté de la communication parlée. Il est possible d'ajouter plusieurs niveaux d'étiquetage. Par exemple, nous pouvons avoir un étiquetage au niveau des phonèmes, des mots, des phrases, etc.

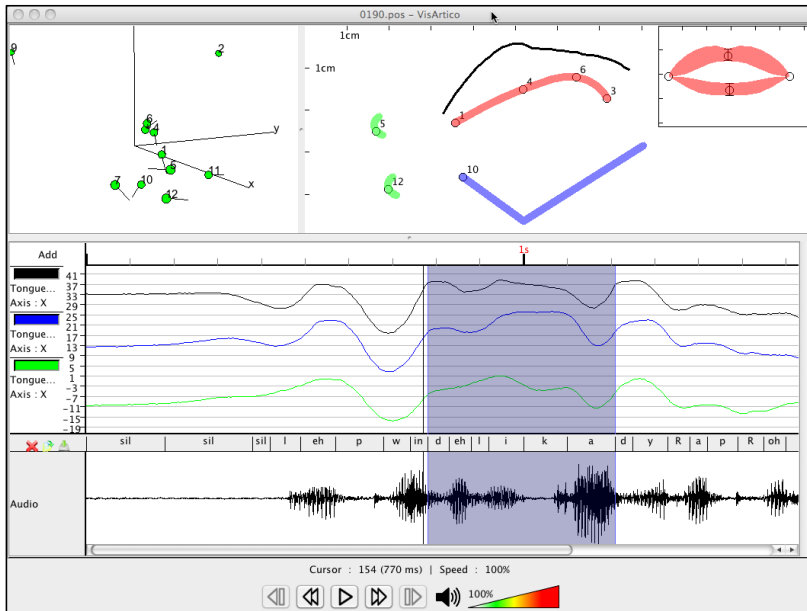


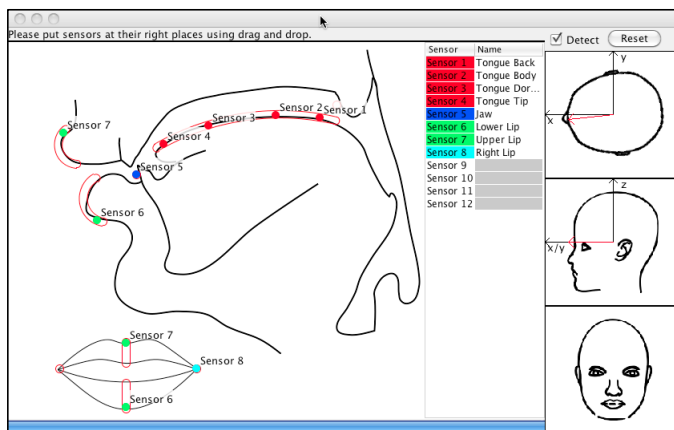
Figure 1 – Interface graphique de VisArtico. 3 vues sont disponibles : une vue 3D, une vue 2D midsagittale, et une vue temporelle représentant les trajectoires articulaires et le signal acoustique.

2.3 Configuration

Lorsque les données sont chargées pour la première fois par le logiciel, ce dernier ne connaît pas la configuration des capteurs, c.-à-d. quels sont les capteurs mis sur la langue, ceux sur les lèvres, etc. L'outil de configuration, représenté par la figure 2, permet de faire cette association entre capteurs et articulateurs. Les articulateurs proposés sont : la langue, les lèvres, la mâchoire et le voile du palais. La forme de la mâchoire est approximative. Le contour de langue est réalisé en connectant les capteurs de la langue, soit par des segments, soit par une spline, pour approcher la forme réelle de la langue. Le logiciel peut déterminer la forme des lèvres à partir de 2 capteurs (par interpolation) et jusqu'à 4 capteurs (deux capteurs au niveau des commissures, un capteur au niveau de la lèvre supérieure et un au niveau de la lèvre inférieure).

Le module de configuration permet de déterminer le plan midsagittal du locuteur, ce qui permettrait une meilleure interprétation de la vue midsagittale. En effet, lors de l'acquisition, on s'arrange pour que les capteurs de la langue soient placés sur le plan midsagittal du locuteur. Néanmoins, l'orientation du locuteur dans le cube de l'articulographe peut ne pas être bien aligné avec le repère de l'articulographe. VisArtico utilise les positions des capteurs de la langue, pour déterminer le plan midsagittal. Il s'agit de déterminer la droite directrice du nuage de points de ces capteurs. Par la suite, le logiciel effectue une rotation par rapport à cette droite. Dans la figure 2, les trois têtes (représentant les vues) montrent le résultat de cette correction. La flèche rouge indique l'orientation du plan midsagittal.

L'utilisation de cet outil de configuration se fait une seule fois pour une acquisition donnée. Le logiciel se rappellera de la configuration pour les autres sessions d'une



même acquisition.

Figure 2 – Module de configuration. Il permet d'indiquer l'emplacement de chaque capteur au niveau du conduit vocal et de déterminer le plan midsagittal du locuteur.

3 Quelques fonctionnalités

3.1 Détection du contour du palais

La visualisation du palais permet une meilleure interprétation du mouvement de la langue. Elle aide notamment à mieux voir les lieux de constriction. Lors d'une acquisition par l'articulographe, on utilise généralement l'une des sessions pour enregistrer le contour du palais. Cela se fait en dessinant le contour du palais manuellement à l'aide d'un capteur. VisArtico permet par la suite d'exploiter cette session pour récupérer le contour du palais qu'il sera possible de visualiser avec les

autres données. Il arrive parfois que les données relatives au palais ne soient pas disponibles. Le logiciel peut alors donner une approximation du contour du palais grâce à un algorithme simple que nous avons développé et qui permet de prédire le contour à partir des positions des capteurs de la langue. Pour chaque instant, on détermine d'abord le contour de la langue à partir des capteurs qui lui sont associés. Nous obtenons ainsi plusieurs contours de langue obtenus à travers toutes les sessions enregistrées. Par la suite, l'algorithme récupère la position maximale dans le plan midsagittal de chaque contour, qui devrait correspondre à un contact entre la langue et le palais. Plus le nombre de sessions est grand, plus la forme du palais est réaliste. Cette solution permet d'approcher la forme réelle du palais.

3.2 Filtrage

Les mesures obtenues par l'articulographe peuvent présenter des erreurs de mesure qui sont généralement indiquées par une valeur d'erreur aux moindres carrées (RMS) fournies par le logiciel de l'articulographe, mais les erreurs peuvent également être dues aux bruits de la machine. VisArtico propose la possibilité d'appliquer un filtre passe-bas sur une ou plusieurs trajectoires articulatoires, pour supprimer les erreurs liées au bruit. Classiquement, nous utilisons une fréquence de coupure de 20hz qui permet de lisser les trajectoires, mais le logiciel donne la possibilité de choisir d'autres fréquences de coupure. Visartico permet également à l'utilisateur de ne pas prendre en compte les segments dont l'erreur RMS est supérieure à un certain seuil et pour lesquels un lissage ne permet pas de corriger ce type d'erreur. Les segments supprimés sont interpolés linéairement.

4 Conclusion

Nous avons présenté dans cet article un logiciel de visualisation des données articulatoires obtenues par un articulographe. L'objectif est d'ouvrir la possibilité d'utiliser l'articulographe aux chercheurs qui ne sont pas nécessairement des informaticiens. Ce logiciel est disponible gratuitement pour les chercheurs qui en font la demande. Un site web est disponible présentant le logiciel et ses principales fonctionnalités ainsi qu'un manuel d'utilisateur (<http://visartico.loria.fr>). Nous envisageons d'apporter plusieurs améliorations à ce logiciel, comme par exemple la possibilité de faire plus de traitements et d'analyse des données, d'afficher le signal acoustique par un spectrogramme, et de donner plus d'informations acoustiques (comme les valeurs des formants par exemple). La conception du logiciel rend son adaptation pour d'autres articulographes, comme le récent AG501 (Carstens) ou encore Wave (NDI) facile. Nous estimons que ce logiciel est très utile pour la communauté de la communication parlée et rend l'utilisation des données articulatoires plus accessible.

BOTHOREL, A., SIMON, P., WILAND, F., ZERLING, J. -P. (1986). Cinéradiographie, des voyelles et consonnes du français. Travaux de l'institut de phonétique de Strasbourg.

BOERSMA, P. (2001). Praat, A system for doing phonetics by computer. *GLOT*

International 5:9/10, 341-345.

BERNHARD, D. (2007). Apprentissage non supervisé de familles morphologiques par classification ascendante hiérarchique. *In* (Benarmara *et al.*, 2007), pages 367–376.

CARSTENS MEDIZINELEKTRONIK, (2006). JUSTVIEW: AG500 MEASURING ENVIRONMENT DISPLAY. *Lenglern, Allemagne.*

GHAZALI, S. (1977). Back consonants and backing coarticulation in Arabic. *Thèse de Doctorat de L'université de Texas, Austin.*

HOOLE, P. (1996). Issues in the acquisition, Processing, reduction and parameterization of articulographic data. *FIPKM*, 34, 158–173.

LAPRIE, Y. (2002). *The Winsnoori User's Manual Version 1.32.*

NGUYEN, N. (2000). *A MATLAB toolbox for the analysis of articulatory data in the production of speech, Behavior Research Methods, Instruments, & Computers, vol. 32, no. 3, pp.464-467.*

PERKELL, J. S., COHEN, M. H., SVIRSKY, M. A., MATTHIES, M. L., GARABIETA, I., JACKSON, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *JASA*, 92, 3078–3096.

