



HAL
open science

Modélisation Articulaire de la Main en Langue Française Parlée Complétée: le cas de la clé digitale

Pablo Sacher, Denis Beautemps, Coriandre Emmanuel Vilain

► To cite this version:

Pablo Sacher, Denis Beautemps, Coriandre Emmanuel Vilain. Modélisation Articulaire de la Main en Langue Française Parlée Complétée: le cas de la clé digitale. JEP 2008 - 27e Journées d'Etudes sur la Parole, Jun 2008, Avignon, France. pp.Article 1679. hal-00288293

HAL Id: hal-00288293

<https://hal.science/hal-00288293>

Submitted on 16 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation Articulaire de la Main en Langue Française Parlée Complétée : le cas de la clé digitale

Pablo Sacher, Denis Beutemp, Coriandre Vilain

Grenoble Image Parole Signal Automatique, département Parole & Cognition
46, av. Félix Viallet, 38031 Grenoble, Cedex 1, France

ABSTRACT

Cued Speech (CS) (in French "Langue Française Parlée Complétée") is a manual code that complements lipreading to enhance speech perception from visual input (Cornett [3]). In the framework of CS synthesis, the handshape modelling is a main topic. Thus this contribution discusses different statistical models used for the prediction of the handshape. An original experimental setup made of an opto-track is presented.

Keywords: Cued Speech, Handshape modelling, PCA

1. Introduction

Ce travail s'inscrit dans un cadre général de modélisation pour la synthèse de la Langue Française Parlée Complétée (code LPC). Le code LPC est un complément manuel à la lecture labiale utilisé pour communiquer avec les personnes sourdes de tradition oraliste. Version en langue Française du Cued Speech (Cornett [3]), il est conçu pour désambiguïser les formes de lèvres similaires ([p, b, m] ou [i, e]) à l'aide d'un code manuel. Une position de la main code pour un groupe de voyelles, une configuration de la main pour un groupe de consonnes. L'association d'une position et d'une configuration code pour un groupe de syllabes de type Consonne-Voyelle (CV). Les différents modules du système de synthèse sont constitués d'un modèle de l'articulation labiofaciale et d'un modèle de main sur lesquels sont appliqués des modèles d'apparence. La contribution à ce programme présentée dans cet article est centrée sur la modélisation de l'articulation de la main utilisée pour coder les consonnes du code LPC à partir d'une analyse statistique de la forme de la main (clés digitales), la forme de la main étant capturée par un dispositif expérimental original ne faisant intervenir que trois caméras.

2. Méthodologie expérimentale

Afin de mettre en place le protocole complet permettant de générer un modèle articulaire des clés digitales de la main en LPC, l'un d'entre nous s'est chargé de prononcer et de coder avec la main un corpus composé de séquences de type /maCamaCama/ avec $C = \{ [p, k, s, b, l, g] \}$. Les consonnes ainsi codées couvrent 7 des 8 clés du code LPC.

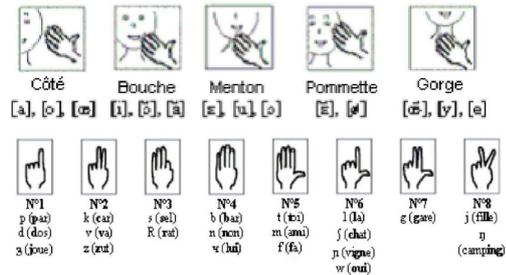


Fig. 1: Liste des positions et des configurations manuelles du code LPC. Attina [4]

2.1. Enregistrement du mouvement de la main

Le mouvement de la main est enregistré au moyen du système de capture du mouvement Optotrak 3020 (Northern Digital Imaging - <http://www.ndigital.com/>). Ce système de mesure actif permet de donner - avec une précision de l'ordre du millimètre - la position 3D de marqueurs infrarouges placés dans l'espace. Dans notre étude, nous avons utilisé un ensemble de 20 marqueurs que nous avons disposés sur la main du codeur de la façon suivante : 1 marqueur est placé aux articulations et sur l'ongle de chaque doigt (soit 3 marqueurs pour le pouce et 4 marqueurs par doigt pour les autres doigts), un dernier marqueur est placé sur le dos de la main (voir figure 2)



Fig. 2: Positionnement des marqueurs infrarouges sur la main du codeur.

Ces marqueurs ainsi placés définissent par leurs coordonnées le maillage 3D de la main. Ils sont reliés à l'unité de traitement du système Optotrak (System Control Unit, SCU). La position de chaque marqueur est mesurée par triangulation au moyen d'un système composé de 3 caméras infra rouges alignées dans un boîtier métallique relié à la SCU. Cette dernière est reliée à un PC d'acquisition (voir figure 3). La fréquence

d'échantillonnage est de 120 Hz. Les données composées des coordonnées X,Y,Z de chaque diode dans un repère fixe lié aux caméras sont ensuite enregistrées sur le PC d'acquisition.

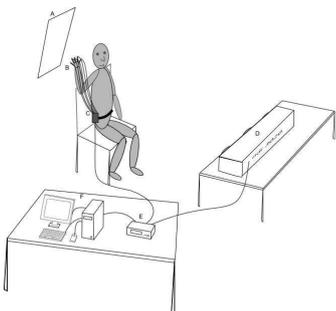


Fig. 3: Schéma du dispositif expérimental.

2.2. Problème lié au masquage des diodes

Du fait de la disposition des caméras alignées dans un boîtier, il est nécessaire que les diodes soient placées de façon telle que leur cône d'émission atteigne les 3 caméras. Un problème se pose alors lorsque les clés digitales entraînent un masquage de certaines diodes, par exemple lorsqu'un ou plusieurs doigts sont pliés. C'est le cas de toutes les clés hormis la clé 5 codant le groupe de consonnes [m,t,f]. Nous avons choisi de pallier ce problème au moyen d'un miroir disposé de telle sorte qu'il réfléchit le cône de lumière infrarouge des diodes situées sur les doigts pliés. Les caméras reçoivent alors un signal identique à celui qui aurait été émis par une diode image de celle tournée vers le miroir. La position réelle de ces diodes peut ensuite être calculée par symétrie par rapport au plan du miroir. Pour calculer ces positions avec précision, nous avons utilisé 3 diodes supplémentaires destinées à mesurer les coordonnées du plan du miroir. De la sorte, en appelant les trois points du miroir A, B, C , il est possible de calculer sa normale \vec{n} . Soit D le point d'une diode et D' sa projection sur le plan, on sait que $D\vec{D}' = k\vec{n}$, donc, le symétrique D_s de D par rapport au plan du miroir peut se déduire par $D\vec{D}_s = -k\vec{n}$ avec $k = C\vec{D} \cdot \vec{n}$. Ensuite, le signe du produit scalaire entre $C\vec{D}_s$ et $C\vec{O}$ (où O est le centre de l'Optotrack) nous renseigne sur le point (D ou D_s) à choisir. Un exemple de trajectoire symétrisée et non symétrisée est présentée figure 4. Le dispositif expérimental complet est schématisé sur la figure 3.

Il n'en demeure pas moins que symétriser les trajectoires des diodes masquées ne permet pas de pallier la totalité des problèmes liés au dispositif expérimental. En effet, la transition entre l'émission directe, et l'émission réfléchiée dans le miroir de la diode masquée, n'est pas instantanée : il existe une zone de l'espace où une diode donnée émet *à la fois* vers le miroir et vers l'Optotrack, ce qui conduit ce dernier à localiser la diode de façon erronée. Ainsi, il existe des points qui ne font partie ni de la trajectoire directe, ni de la trajectoire symétrisée, nous les appellerons des points aberrants. Un exemple de ces points est montré dans la figure 5.

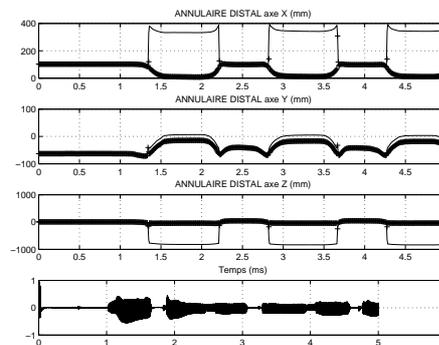


Fig. 4: Exemple de trajectoire symétrisée (+) par rapport à la trajectoire non symétrisée (lignes).

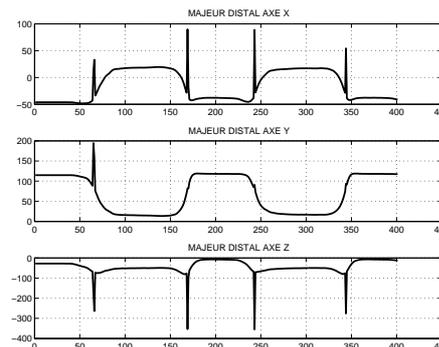


Fig. 5: Trajectoires des signaux originaux pour la séquence [mala] et mise en évidence des discontinuités

Pour résoudre le problème posé par l'existence des points aberrants, nous avons décidé de supprimer ces derniers. Pour cela, nous avons recherché l'ensemble des points situés en limite de capture directe par l'Optotrack et, en appliquant un seuil relatif à la variation de la pente aux points précédents cette limite, nous avons supprimé ceux dont la pente était significativement différente des autres. Ensuite, nous avons reconstruit les points supprimés de la courbe par interpolation.

3. Analyse des données

Rappelons que l'objectif de cette étude consiste à comparer plusieurs modèles statistiques prédisant les mouvements de la main dans une situation de codage. Nous disposons d'une série de données correspondant aux coordonnées 3D de la position de tous les capteurs de la main au cours du temps, soit $3 \cdot 20 = 60$ paramètres, et 7787 observations ; à partir de ces données il est possible de tester les différents modèles de prédiction. Étant donné que l'objet de la modélisation est centré sur l'articulation des clés digitales du code LPC, on considère donc les coordonnées relatives, c'est-à-dire en ayant eu soin de retirer le mouvement de roto-translation de la main. Ces modèles sont au nombre de deux et se distinguent par leurs hypothèses et le nombre de contraintes qu'ils supportent.

Tout d'abord, précisons que le principal outil mathématique qui sous-tend les deux modèles est l'Analyse en Composantes Principales (ACP). Il n'est pas in-

utile de rappeler que cette dernière repose à la fois sur des notions statistiques et des notions algébriques. En effet, cette analyse consiste à diagonaliser la matrice de covariance de données multidimensionnelles (à n observations et p variables). La matrice de covariance permet de mesurer les variations (linéaires) conjointes entre les variables prises deux à deux. Diagonaliser une telle matrice revient donc à chercher une nouvelle base définie par les vecteurs propres (les axes principaux) au sein de laquelle la matrice de covariance est diagonale, ce qui signifie que $\forall (i \neq j) cov(X_i, X_j) = 0$ et donc que les données projetées (les facteurs principaux) dans cette nouvelle base sont décorrélées. Par ailleurs, chaque vecteur de la nouvelle base est associé à une valeur propre de la matrice de covariance (valeur qui correspond à la variance de chaque variable exprimée dans la nouvelle base). Autrement dit, en se fixant un seuil d'explication τ de la variance, il est possible de réduire la matrice initiale de p à q variables où q est le nombre de valeurs propres expliquant la variance au seuil τ .

3.1. Modèle en développement trigonométrique

Le premier modèle que nous avons utilisé est celui présenté par Gibert et al. [2]. Il s'appuie sur un développement des coordonnées 3D en série trigonométrique de cosinus et sinus d'angles caractéristiques de la main. Ces angles (flexion/extension et abduction/adduction de toutes les phalanges et la pseudo-rotation du pouce) sont choisis de manière à rendre compte des degrés de liberté de la main (pour une revue voir Wagner [1]). Pour cette approche, il est nécessaire de préciser les hypothèses qui conduisent à un certain nombre d'approximations. D'une part, l'emplacement des capteurs définit un maillage (un graphe) dans lequel les angles sont calculés ; cela suppose que le choix de l'emplacement des capteurs permet de rendre compte de l'ensemble des configurations de la main pour le code LPC. D'autre part, que le mouvement des phalanges s'effectue sans étirement des cartilages au niveau des articulations, ce qui a pour conséquence de ne considérer qu'un mouvement dans un plan 2D.

Les angles caractéristiques sont prédits par un ensemble en nombre réduit de paramètres articulaires. Ces paramètres sont obtenus comme les facteurs principaux d'une ACP sur ces angles. Ce modèle s'articule autour de six étapes successives que nous présentons brièvement.

Étape 1. Calculer les angles caractéristiques du mouvement de la main.

Étape 2. Obtenir les paramètres articulaires de la main en réalisant une ACP sur les angles précédents.

Étape 3. Réaliser une régression linéaire entre les angles originaux et les paramètres articulaires pour obtenir des coefficients de régression entre les deux ($coef f_{ang}$).

Étape 4. Prédire les angles à partir des coefficients de régression et des paramètres articulaires.

Étape 5. Réaliser une régression linéaire entre les cosinus et sinus des angles prédits et les coordonnées 3D centrées du maillage de la main ($coef f_{pos}$).

Étape 6. Prédire les coordonnées des capteurs à par-

tir des coefficients de régression et des cosinus et sinus des angles prédits.

Remarquons toutefois que nous disposons autant de séries de $coef f_{pos}$ que de paramètres articulaires (p paramètres). En effet, nous avons calculé les p séries de $coef f_{pos}$ en effectuant une régression entre les cosinus et les sinus des angles prédits par les i^{emes} premiers paramètres articulaires, et les coordonnées 3D du maillage de la main.

Bilan Cette modélisation appliquée à nos données a l'avantage d'utiliser les angles caractéristiques (avec les hypothèses précédentes) ce qui permet de réduire de 60 à 21 le nombre de variables (les angles) soumises à l'analyse. La table 1 présente l'explication de la variance en fonction du nombre de paramètres articulaires (facteurs principaux). Ainsi, en se fixant par exemple un critère τ de 99% de la variance expliquée, on retient 8 paramètres articulaires comme prédicteur des angles caractéristiques.

Tab. 1: Variance des angles caractéristiques expliquée en fonction des paramètres articulaires (facteurs principaux de l'ACP)

Paramètre	% variance	% cumulé
P1	76,4812	76,4812
P2	11,0614	87,5426
P3	6,0649	93,6074
P4	2,4657	96,0731
P5	1,4690	97,5421
P6	0,8073	98,3494
P7	0,4565	98,8059
P8	0,4055	99,2114
P9	0,3240	99,5354

3.2. Modèle direct

Le deuxième modèle s'affranchit du calcul des angles, il repose directement sur une ACP des coordonnées 3D des capteurs et une régression linéaire entre les facteurs principaux obtenus et ces mêmes coordonnées. Cette modélisation a l'avantage de ne faire intervenir qu'une étape de prédiction. Il a l'inconvénient de ne pas réduire substantiellement la dimension de l'espace des données initiales. La table 2 présente l'explication de la variance des coordonnées 3D du maillage de la main en fonction du nombre de facteurs principaux. On remarque ainsi qu'il faut retenir 9 paramètres pour prédire 99% de la variance totale des données - en comparaison aux 8 paramètres de l'analyse en série trigonométrique.

4. Résultats

Nous avons appliqué ces deux modèles sur le même ensemble de données (les coordonnées 3D des capteurs placés sur le dos de la main et des doigts) et réalisé la comparaison de leur précision sur la prédiction des données initiales. Celle-ci est évaluée en fonction de l'erreur de reconstruction, c'est-à-dire la distance

Tab. 2: Variance des coordonnées 3D expliquée en fonction des paramètres articulatoires (facteurs principaux de l'ACP)

Paramètre	% variance	% cumulé
F1	72,7882	72,7882
F2	12,2230	85,0052
F3	5,9469	90,9521
F4	3,3634	94,3156
F5	2,8751	97,1907
F6	0,7183	97,9090
F7	0,6467	98,5557
F8	0,4377	98,9934
F9	0,2783	99,2717

euclidienne entre les positions 3D prédites et les positions 3D initiales. La figure 6 présente pour chaque en-

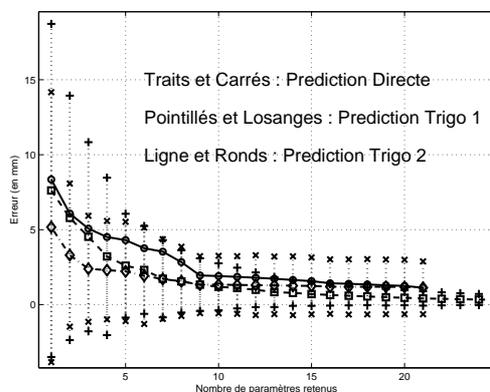


Fig. 6: Moyenne et écart-type de l'erreur (en mm) entre les prédictions et les données initiales en fonction du nombre de paramètres pour chaque modélisation (x : modèle trigonométrique ; + : modèle direct).

semble de paramètres de l'ACP, la moyenne et l'écart-type de l'erreur sur la reconstruction des 20 points de mesure, et ce pour les deux modèles (remarquons toutefois que nous avons choisi d'exposer une troisième courbe ("trigo 2" sur le graphe) qui présente l'erreur pour le modèle trigonométrique en utilisant les coefficients de la régression pour la totalité des paramètres articulatoires issus de l'ACP). On observe pour les deux modèles une baisse de l'erreur avec l'augmentation du nombre de paramètres avec, pour le modèle en développement trigonométrique, l'atteinte d'une erreur plancher de 1,12 mm. Nous y reviendrons par la suite. Par ailleurs, la courbe "trigo 2" est toujours au dessus de la courbe présentant l'erreur du modèle direct.

Rappelons maintenant qu'un des objectifs de la modélisation est de définir un ensemble de paramètres articulatoires de dimension suffisante mais limitée pour contrôler les points du maillage de la main. On retient d'après la figure 6 que le modèle en développement trigonométrique est le plus précis si l'on retient un nombre de paramètres inférieur ou égal à 9. Dans ce dernier cas, les erreurs moyennes pour les deux modélisations sont identiques. Rappelons que la précision de mesure du système Optotrack est de 1mm. L'er-

reur peut être réduite à zéro avec le second modèle au détriment du nombre de paramètres à retenir. L'analyse de l'erreur par point du maillage montre qu'en moyenne, pour le modèle en développement trigonométrique, c'est le point de l'extrémité du pouce qui contribue le plus à l'erreur. Toutefois, ce résultat est à nuancer, car la totalité des capteurs présents à l'extrémité des doigts montrent des variations importantes par rapport aux données (de l'ordre de 2,03mm en moyenne avec un maximum pour le pouce à 4,11mm, contre une variation de 1,08mm pour la totalité des points, et 0,76mm pour l'ensemble des points privé des points du maillage associées aux extrémités des doigts). Par ailleurs, sans retirer les points aberrants, il apparaît que le modèle trigonométrique fasse une erreur de 2,25mm si l'on retient la totalité des paramètres articulatoires - contre 0mm pour le modèle direct. Ainsi, la pertinence du modèle de développement en série trigonométrique se montre par sa capacité de filtrage, et plus généralement par le fait qu'il est moins sensible aux erreurs de mesure.

5. Conclusion

En conclusion, après avoir comparé deux modèles articulatoires d'une main en situation de codage de la Langue Française Parlée Complétée, il semble que le modèle trigonométrique présenté par Gibert [2], dans le cas où l'on souhaite minimiser le nombre de paramètres articulatoires, soit le plus performant, notamment par sa capacité à être peu sensible aux erreurs de mesure dues au dispositif expérimental. Il s'agira, dans la suite, de concevoir un dispositif expérimental qui empêche l'apparition des points aberrants (mise en série de deux Optotracks, par exemple). Par ailleurs, l'hypothèse du mouvement des phalanges dans un plan 2D est validée par le fait que l'erreur plancher de 1,12mm est de l'ordre de la précision de mesure de l'Optotrack (1mm). Enfin, en ce qui concerne le choix du modèle à adopter, si le critère de sélection consiste à résoudre la double contrainte relative à la minimisation du nombre de paramètres articulatoire à retenir, et à la minimisation de l'erreur, il apparaît que le modèle de développement en série trigonométrique est celui qui offre le compromis le plus judicieux. Toutefois, il pourrait être intéressant de rechercher des développements en série (autres que celui en sinus et cosinus) qui permettraient d'obtenir des résultats encore plus probants.

Références

- [1] Wagner C. The pianist's hand : anthropometry and biomechanics. *Ergonomics*, 31 :97-131, 1988.
- [2] Gibert G., Bailly G., Beutemps D., and Elisei F. Analysis and synthesis of the three-dimensional movements of the head, face and hand of a speaker using cued speech. *Journal of the Acoustical Society of America*, pages 1144-1153, 2005.
- [3] Cornett R-O. Cued speech. *American Annals of the Deaf*, 112 :3-13, 1967.
- [4] Attina V., Beutemps D., Cathiard M-A., and Odisio M. A pilot study of temporal organization in cued speech production of french syllables : rules for cued speech synthesizer. *Speech Communication*, 44 :197-214, 2004.