DÉMÊLER LES ACTIONS DES ARTICULATEURS EN JEU LORS DE LA PRODUCTION DE PAROLE AVEC LE LOGICIEL C.H.I.C. : ANALYSE DE SÉQUENCES DE RADIOGRAPHIES DE LA TÊTE.

Julie BUSSET[1] et Martine CADOT[2]

TITLE

C.H.I.C. software for untangling the various interventions of the articulators during speech production: Analysing radiographic sequences of the head.

RÉSUMÉ

L'analyse d'images issues de ciné-radiographies d'une personne en train de parler présente des difficultés. La première de ces difficultés, d'ordre représentationnel, est que les mouvements des articulateurs (mâchoire, langue, etc.) à l'origine de la parole se situent dans un espace multidimensionnel complexe du fait des interdépendances mécaniques multiples et dynamiques. La deuxième, d'ordre technique, est que ces données sont issues d'annotations des images réalisées en plusieurs temps, lieux, et selon plusieurs techniques, manuelles ou semi-automatiques. Nous montrons dans cet article une utilisation de CHIC qui produit, à partir des données nettoyées et projetées dans un espace de dimension réduite, une représentation synthétique des liens entre les points des articulateurs : le graphe implicatif qui en résulte, montre, sans introduire de connaissances a priori, une séparation claire entre les structures articulatoires, groupées pour certaines, éclatées pour d'autres.

Mots-clés : *analyse implicative, fouille de données, cinéradiographie, modèle articulatoire, production de parole, échelonnement multidimensionnel, algorithme INDSCAL.*

ABSTRACT

Cine-radiographic image sequences of a speaker are a challenging material for data analytics. First, the articulary movements of the speech-generating elements (tongue, maxillary, ...) are lying in a complex multidimensional space, because of the numerous and dynamical interdependences of mechanical origin. Second, these data come from images whose annotations originate in diverse places and times, using several techniques, whether manual or semi-automatic. In this paper we use CHIC for producing a synthetic representation of the links between articulary points, starting from cleaned data projected in a low-dimensional space: the resulting implicative graph shows a clear separation between articulary structures, whether fragmented or put together.

Keywords : *Statistical Implicative Analysis, data mining, Cineradiography, articulary model, speech production, multidimensional scaling (MDS), Individual Difference Scaling (INDSCAL)*

[1] Université de Nancy1/LORIA, Campus scientifique, BP 239, F-54506 Vandoeuvre-lès-Nancy Cedex, julie.busset@loria.fr

[2] Université de Nancy1/LORIA, Campus scientifique, BP 239, F-54506 Vandoeuvre-lès-Nancy Cedex, martine.cadot@loria.fr

1 Introduction

La production de parole est un phénomène complexe qui peut être abordé de multiples façons. Disposant d'un jeu de données formé de quatre séquences de radiographies annotées d'un locuteur en train de prononcer six phrases, nous avons décidé d'aborder ce phénomène par la statistique exploratoire : créer une représentation dans laquelle ne figureraient que les articulateurs dont le mouvement est nécessaire pour la prononciation de la phrase. Par exemple, on sait bien que les mouvements de la langue sont nécessaires pour la parole, mais qu'en est-il de l'os hyoïde ? Bouge-t-il sans raison, ou a-t-il un rôle dans la production de sons et lequel ? A notre connaissance, il n'y a pas encore de réponse à cette question.

Représenter les seuls articulateurs en action pour produire une phrase était une tâche difficile, notamment parce que le lien entre mouvements des articulateurs et production d'un son n'est pas univoque : des mouvements différents peuvent produire le même son (si un articulateur a son mouvement gêné, les autres peuvent compenser), et des mouvements identiques peuvent produire des sons différents (les articulateurs ne sont pas les seuls « acteurs » de la parole).

Cet article expose comment nous avons résolu en partie la tâche en trouvant une représentation non pas pour chaque phrase, mais pour chaque séquence. Il est composé de 4 parties. La première est un exposé des éléments dont nous disposons pour cette tâche : des connaissances sur la production de parole, des données sur les séquences d'images, nettoyées, transformées pour être représentées par CHIC[3]. La deuxième partie montre le principe de notre méthodologie sur un petit jeu de données ayant déjà servi d'exemple dans des ouvrages de statistique. La troisième partie analyse les représentations produites par CHIC sur les 4 séquences, et la quatrième partie contient les conclusions et perspectives.

2 La problématique et les données

La première partie de cette section décrit le cadre général de notre étude en reprenant de façon condensée le chapitre 3 du mémoire de stage de Master 2 réalisé par Julie Busset en 2006-2007 au sein de l'équipe PAROLE du LORIA. La deuxième partie positionne le problème particulier que nous traitons dans cet article, et la troisième partie les données que nous utilisons, leur acquisition et leur transformation.

2.1 Production de la parole

Nous ne mettons dans cette partie que les quelques éléments que nous jugeons indispensables pour pouvoir mesurer l'importance de l'apport de CHIC dans la représentation des mouvements du conduit vocal lors de la parole. Nous renvoyons le lecteur intéressé par plus de détails au document (Busset, 2007) d'où sont tirés ces éléments, ainsi qu'aux nombreux ouvrages théoriques traitant de ce domaine (Haton et al., 2006).

[3] Le lecteur intéressé par une description de CHIC pourra consulter le site indiqué dans les références

2.2 Les organes de production

La parole est une modulation du flux d'air qui provient des poumons. En passant par les cordes vocales, ce débit d'air génère une onde voisée ou sonore qui est envoyée dans le conduit vocal (Figure 1), avant de sortir à travers les lèvres et le conduit nasal. Les différents sons proviennent essentiellement des déformations du conduit vocal, représenté de façon très simplifiée, pour un adulte, par un tube de 17 cm de longueur avec une section transversale qui varie entre 0 (constriction maximale) et 20 cm². Les principaux articulateurs de ce conduit vocal sont la mâchoire, la langue, les lèvres, le vélum (voile du palais) et le larynx.

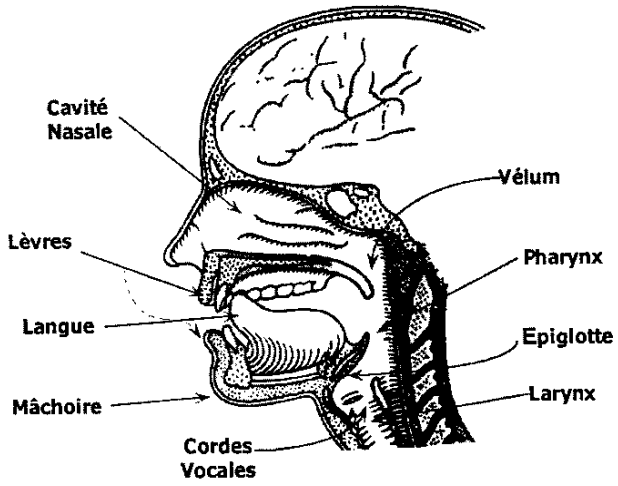

FIGURE 1 – *Schéma de l'anatomie du conduit vocal (d'après Flanagran 1972)*

En français, la plus petite unité de son est appelée « phonème » et ils se divisent essentiellement en « voyelles » et « consonnes ».

2.2.1 Production des voyelles et des consonnes du français

Les voyelles se caractérisent par l'absence ou la présence de nasalité (quand le voile du palais s'abaisse, ce qui met en parallèle les cavités buccale et nasale), le degré d'ouverture du conduit vocal, la position de la constriction principale du conduit vocal et la position des lèvres. Il y a 12 voyelles orales émises seulement par la bouche (par ex. dans les mots l**i**t, pr**é**, b**e**lle, p**a**tte, p**â**te, l**u**, p**eu**, l**eu**r, le, l**ou**p, l**o**t, l**o**tte) et 4 voyelles nasales (par ex. dans les mots br**in**, br**un**, l**en**t, l**on**g).

Les consonnes se caractérisent par le « voisement » (sonores ou sourdes), le mode d'articulation (occlusif, nasal, fricatif, glissant ou liquide), la position de la constriction principale (lèvres, dents, voile du palais). Par exemple la consonne **t** du mot « **t**ous » fait partie de la classe des consonnes occlusives et sonores (Haton et al. 2006), comme le **c** de « **c**as » et le **p** de **p**ère, ce qui se traduit par la succession des 3 étapes acoustiques suivantes :

1. un silence correspondant au blocage de l'air dans le conduit vocal ;
2. une explosion (« burst »), due au relâchement brusque de l'air ;

3. une transition vers la voyelle qui suit.

2.2.2 La co-articulation

Nous avons vu dans la sous-section précédente que la prononciation de la consonne **p** de **p**ère nécessitait trois étapes, la troisième étant une transition vers la voyelle suivante. Cette transition entre 2 phonèmes fait partie d'un phénomène plus général de « coarticulation », qui signifie que deux phonèmes qui se suivent subissent tous deux une distorsion[4]. L'étude de la coarticulation est un domaine de recherche très actif qui a donné lieu à plusieurs articles et thèses au sein de l'équipe Parole du LORIA.

2.3 Des articulateurs à la parole avec CHIC

Dans cet article, nous montrons comment CHIC permet de démêler automatiquement, c'est-à-dire sans introduire de connaissances a priori, les actions des divers articulateurs du conduit vocal pendant la production de parole. Toutefois, la connaissance du domaine intervient avant et après. Avant de proposer les données en entrée au logiciel, il faut les transformer en des matrices de type « sujets x variables » dans le format attendu du logiciel, avec les étiquettes appropriées, tout en conservant au mieux leur nature multidimensionnelle, notamment proximités dans l'espace et succession dans le temps. Et après les avoir traitées par CHIC, il convient de retourner aux données afin d'interpréter les résultats du logiciel en s'aidant de la connaissance du domaine. Nous nous limiterons dans cet article à quelques interprétations visant à montrer que les résultats trouvés par CHIC ont du sens dans le domaine de la parole.

2.4 Les données, leur acquisition et leur transformation

Le corpus a été enregistré dans les années 90 afin d'étudier la coarticulation de la langue française. Les données sont formées de 4 séquences, que nous appellerons H1, H2, H3 et H4, composés de radiographies successives d'une même personne en train de prononcer des « phrases » courtes de 6 mots. Pour H3 et H4, la séquence est simple[5] (/aku/, /iku/, /uku/, /atu/, /itu/ et /utu/), mais pour H1 et H2 les 6 phrases sont plus complexes ; elles commencent et finissent par les mêmes phonèmes /se dø si/ et /ylt**ɛʁ**/ avec au milieu une consonne non labiale[6] de plus à chaque fois. La première phrase est /se dø siylt**ɛʁ**/ et la sixième est /se dø si**kst** **sk**ylt**ɛʁ**/[7]. Les phrases sont prononcées normalement dans H1 et H3, plus vite dans H2 et H4.

2.4.1 Acquisition

Les images ont été annotées une à une d'abord manuellement puis de façon semi-automatique. Dans un premier temps, les contours des articulateurs ont été indiqués par

[4] Termes repris d'une définition de Wikipedia (voir section Références)

[5] Le « slash » indique que la notation n'est pas celle du français, mais de la phonétique internationale : le /u/ est l'écriture phonétique du « ou » du français, alors que /y/ est l'écriture phonétique du « u » du français.

[6] Labiale : qui se prononce avec les lèvres

[7] Le symbole /ø/ désigne le « eu » de « p**eu** », /ɛ/ désigne le « è » de « b**e**lle », et /ʁ/ désigne le « r » du français, comme « pè**r**e »

autant de clics sur la radiographie que de points visibles sur l'image (Figure 2), les coordonnées 2D de ces points étant stockées dans un fichier.

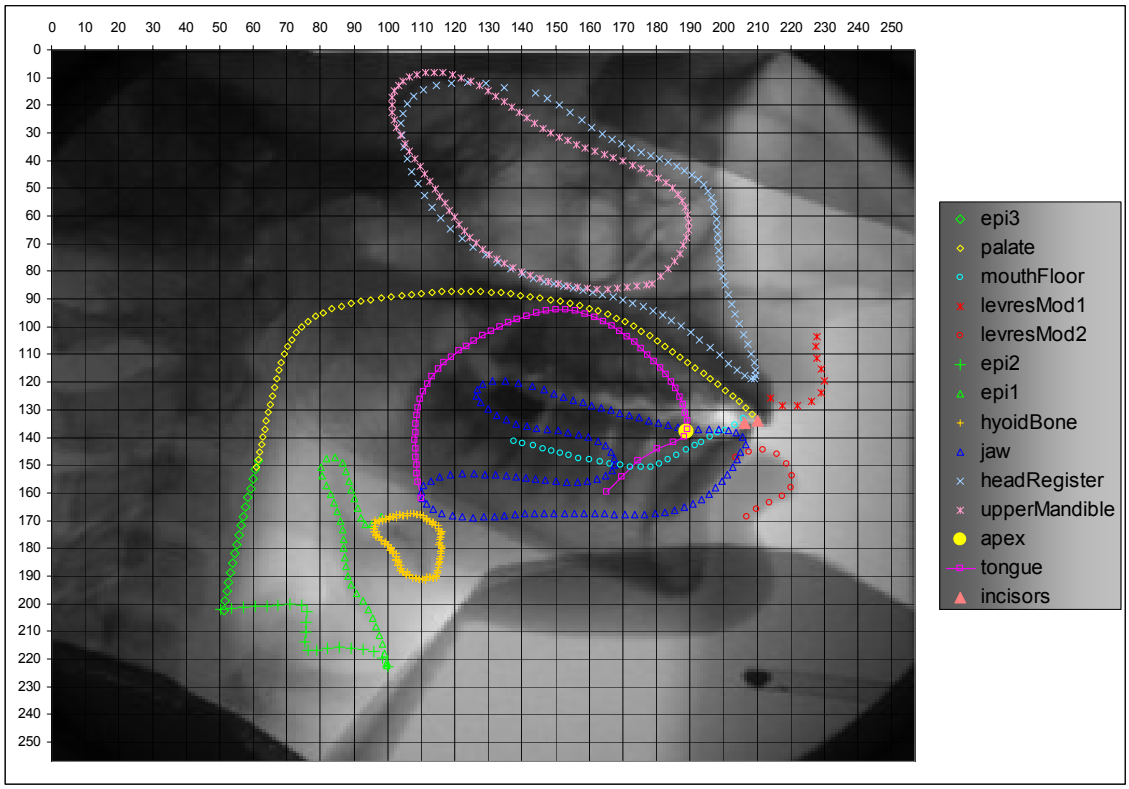

FIGURE 2 – *En couleur les points des contours des articulateurs annotés sur la radiographie Hirsch1_156*

Puis un logiciel a été créé afin de les annoter de façon semi-automatique en se calant sur quelques images de référence. Ce logiciel a connu de nombreuses versions. Par exemple, dans une de ses versions, pour annoter une nouvelle image, l'annotateur doit choisir une image de référence, puis le fichier de ses contours est lu et une partie de ceux-ci (les articulateurs non déformables, comme l'os hyoïde en jaune foncé, la mâchoire en bleu foncé, le haut de la tête en bleu clair et en rose clair) sont projetés sur l'image à annoter, il ne reste plus qu'à les déplacer à la souris (rotations et translations) pour les adapter au mieux à l'image saisie. Toutefois les lèvres inférieure et supérieure (rouge), les 3 parties de l'épiglotte (vert) sont trop déformables pour pouvoir utiliser cette méthode. Quant à la langue, elle est en plus difficile à repérer car en grande partie cachée. Plusieurs techniques ont été testées pour aider à la détermination des contours déformables (Busset, 2011).

2.4.2 Nettoyage

Nous disposions de 1000 fichiers de coordonnées de contour environ, autant que d'images. Nous les avons lus un à un pour stocker les informations sur chaque contour (intitulé, coordonnées des points) dans un tableur. Le nettoyage a consisté à ne garder qu'un nombre fixe de points par contour (Figure 3, tracés en pointillés), et à effectuer des rotations et translations pour recaler toutes les images entre elles à partir de points de référence. Nous avons gardé les points de référence dans les données car il leur

restait une légère variabilité[8]. Et seuls les fichiers présentant tous les contours sans erreurs ont été conservés, soit 671.

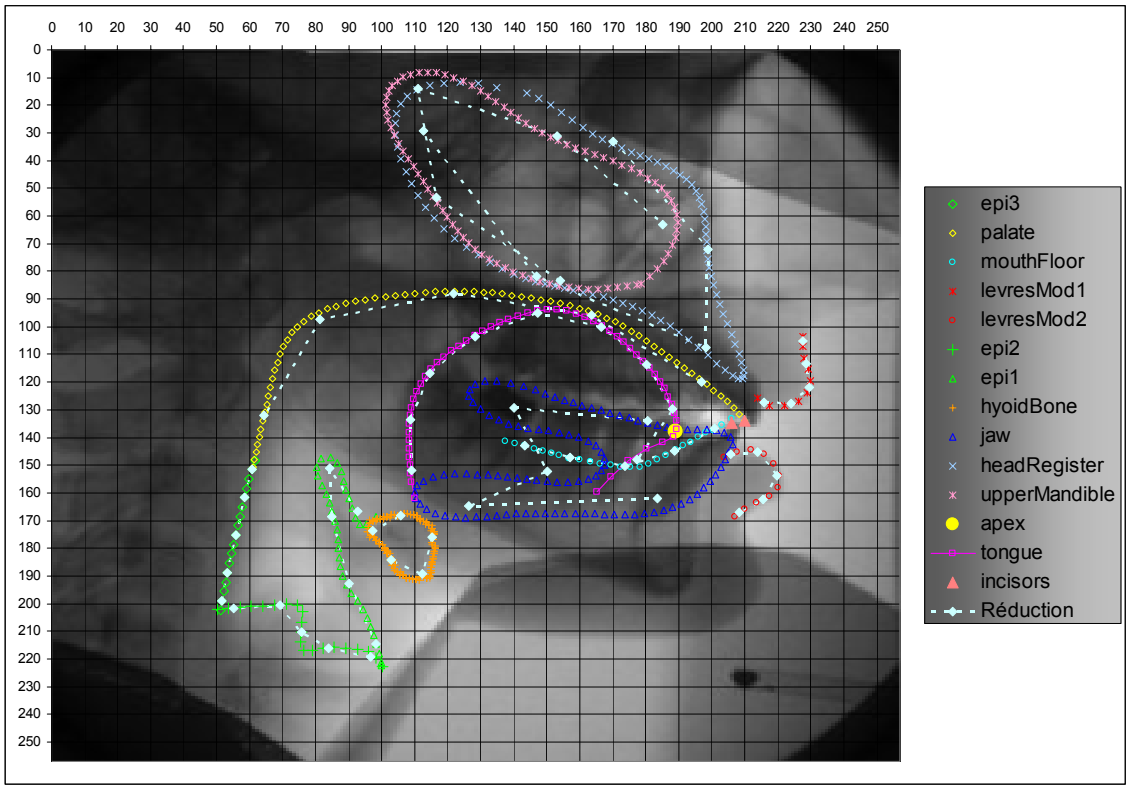

FIGURE 3 – *En pointillés les courbes « réduction », formées de 5 points pour chaque contour de Hirsch1_156, sauf la langue (tongue) qui en a 9*

A la fin de cette étape, nous disposions de 4 hyper-matrices de 71 colonnes (les contours), de p lignes (p étant le nombre d'images conservées), et de profondeur 2 (coordonnées x et y).

2.4.3 Pré-traitement avec INDSCAL

Pour pouvoir traiter ces hyper-matrices, il convenait de les transformer en matrices, en essayant de ne pas perdre les liaisons spatiales et/ou temporelles. Pour garder les premières, nous avons choisi de remplacer les coordonnées x et y par les distances entre points pris deux à deux, puis nous avons transformé les distances en coordonnées dans un espace de dimension réduite avec l'algorithme INDSCAL (Young, 1972), décrit sommairement dans la section suivante sur les données colas de Schiffman et al. (1981). Deux types de distances étaient possibles, soit des distances entre les points d'une même image, soit les distances entre les coordonnées sur les images d'un même point. Pour le premier type de distance, l'algorithme INDSCAL convergeait dès 2 dimensions. Les résultats étant d'autant plus intéressants que le nombre de dimensions était élevé, nous avons choisi le deuxième type de distance et nous avons obtenu en sortie de cet algorithme 4 matrices donnant les coordonnées des 71 points des contours dans un espace réduit de 25 dimensions pour H2, H3, H4 et 30 dimensions pour H1.

[8] Rappelons que ces radiographies de la tête ont été faites en 2D sur des personnes en 3D en train de parler, et qu'il est difficile de parler sans tourner du tout la tête !

3 INDSCAL et CHIC sur un jeu d'essai

Les méthodes d'échelonnement multidimensionnel (MDS pour MultiDimensional Scaling) ont pour but de transformer des données multidimensionnelles issues d'échelles (scores allant de 1 à 5 par exemple) en des points situés dans un espace métrique de dimension réduite (Tournois et Dickes,1993). Parmi ces méthodes, nous avons choisi l'algorithme INDSCAL (« Multidimensional Scaling for Individual Differences » de Caroll (1972) qui peut traiter des tableaux de données 3 voies (objets x objets x sujets), par exemple autant de matrices de dissimilarités[9] entre objets qu'on a de sujets, et produire en sortie non seulement les positions des objets dans un espace métrique de dimension réduite, mais également les positions des sujets, appelées poids, sur une hyper-sphère de même dimension (Schiffman et al., 1981). Nous utilisons l'algorithme INDSCAL sur un petit jeu d'essai de 10 objets et 10 sujets, puis nous essayons a posteriori de retrouver les deux groupes de sujets, d'abord directement sur l'hyper-sphère, puis en utilisant CHIC.

3.1 Les données « colas »

Nous prenons les données que Schiffman et al. (1981) ont utilisé pour illustrer diverses procédures de MDS. Elles sont issues d'une expérience de dégustation de dix boissons différentes à base de cola. Les données se présentent sous forme de dix matrices de dissimilarités, une par sujet, avec dans chaque matrice les différences de goût entre toutes les paires de boissons exprimées par le sujet sous forme de nombres entre 0 et 100. Indépendamment de ces données, on sait que seuls les sujets 1, 4, 5, 6 et 9 sont « PTC tasters » c'est à dire sensibles au phenylthiocarbamide.

3.2 Le traitement par INDSCAL

La procédure INDSCAL de Caroll (1972) produit pour chaque sujet un vecteur de longueur 1 et d'origine O le centre du repère. Les extrémités des vecteurs sujets sont donc sur une sphère de centre O et de rayon 1. Les coordonnées des vecteurs (Tab. 1) étant des poids compris entre 0 et 1, les vecteurs sont tous dans l'octant positif (dim1>0, dim2>0 et dim3>0) de l'espace 3D choisi par Schiffman et al. (1981).

TABLEAU 1 – *Vecteurs poids individuels des sujets dans l'espace à 3 dimensions obtenu par INDSCAL*

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
dim1	0,69	0,32	0,32	0,75	0,9	0,88	0,45	0,54	0,74	0,41
dim2	0,51	0,75	0,76	0,36	0,09	0,11	0,66	0,64	0,38	0,72
dim3	0,51	0,58	0,56	0,55	0,43	0,46	0,61	0,54	0,56	0,56

3.3 L'interprétation des résultats d'INDSCAL

Pour pouvoir grouper au mieux les sujets présents sur une sphère, on calcule leurs éloignements deux à deux par l'angle entre leurs deux vecteurs dans l'espace 3D (Tableau 2), comme spécifié dans (Schiffman et al. 1981).

[9] Les dissimilarités sont des écarts comme les distances, mais moins « rigoureux » que ces dernières : ils ne respectent pas nécessairement l'inégalité triangulaire d(A,C)≤d(A,B)+d(B,C).

On a indiqué par une étoile dans le tableau 2 les trois plus petits angles : entre les sujets s2 et s3 (0,02 radians), entre les sujets s4 et s9 (0,02 radians) et entre les sujets s5 et s6 (0,03 radians). Mais la lecture approfondie de ce tableau est assez fastidieuse, et ne débouche pas directement sur la constatation de l'existence de deux groupes. Nous essayons alors des représentations graphiques.

TABLEAU 2 – *Angles (en radians) entre les 10 vecteurs sujets pris 2 à 2.*

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
s1	0	0,45	0,46	0,17	0,48	0,45	0,3	0,2	0,15	0,36
s2	0,45	0	0,02*	0,59	0,92	0,89	0,16	0,25	0,57	0,09
s3	0,46	0,02*	0	0,6	0,92	0,9	0,18	0,25	0,58	0,1
s4	0,17	0,59	0,6	0	0,33	0,3	0,43	0,36	0,02*	0,5
s5	0,48	0,92	0,92	0,33	0	0,03*	0,75	0,68	0,35	0,83
s6	0,45	0,89	0,9	0,3	0,03*	0	0,72	0,65	0,32	0,8
s7	0,3	0,16	0,18	0,43	0,75	0,72	0	0,11	0,41	0,09
s8	0,2	0,25	0,25	0,36	0,68	0,65	0,11	0	0,33	0,15
s9	0,15	0,57	0,58	0,02*	0,35	0,32	0,41	0,33	0	0,48
s10	0,36	0,09	0,1	0,5	0,83	0,8	0,09	0,15	0,48	0

Pour observer plus facilement les proximités des vecteurs, on les a projetés en dimension 2 dans les 3 plans possibles (Figure 4). La longueur des vecteurs projetés est dans tous les cas devenue inférieure à 1, ce qui fait que leur extrémité se situe non pas sur le cercle de centre 1 mais à l'intérieur de la portion de disque située dans le premier quadrant On a marqué par une forme carrée les 5 sujets qui sont « PTC tasters, et par un cercle les 5 autres sujets. On constate sur les graphiques qu'il n'y a pas de chevauchement entre les deux groupes, mais qu'ils ne sont pas suffisamment distants entre eux pour qu'on puisse affecter les sujets au bon groupe sans introduire de connaissances a priori.

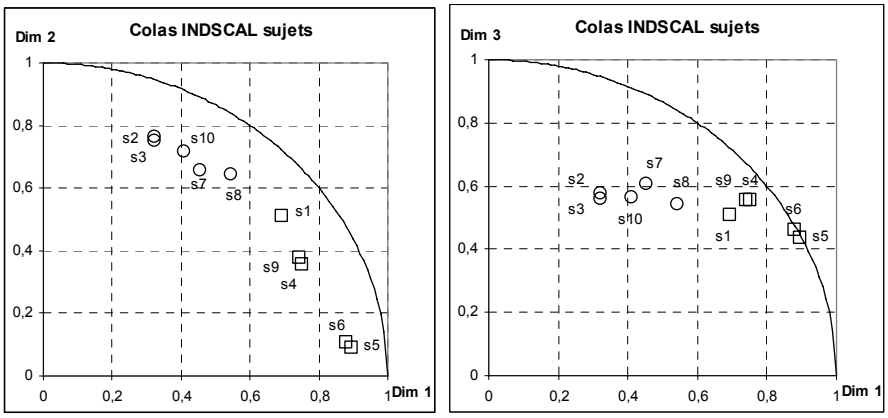

FIGURE 4 – *Projection des vecteurs sujets dans 2 des 3 plans des axes*

3.4 La représentation graphique avec CHIC

Le logiciel CHIC, après lecture des données du tableau 1, permet d'en avoir d'autres représentations, par « arbre cohésitif », et par « graphe implicatif ». Pour le premier, nous n'avons pas fixé de valeurs de paramètres, mais utilisé celles par défaut, et nous avons obtenu 2 hiérarchies bien séparées correspondant aux deux groupes de sujets

(Figure 7, gauche). Le tableau des indices de cohésion montre cette même séparation (Figure 5).

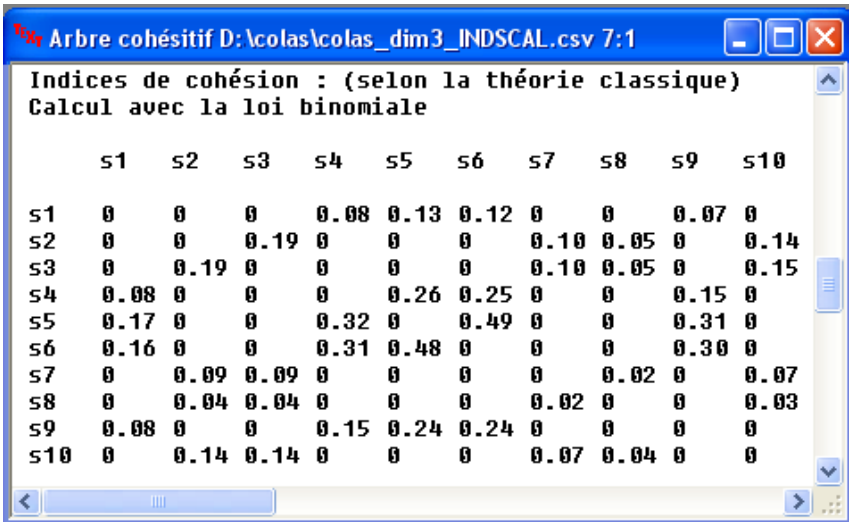

FIGURE 5 – *Indices de cohésion des données colas, copie d'écran de CHIC*

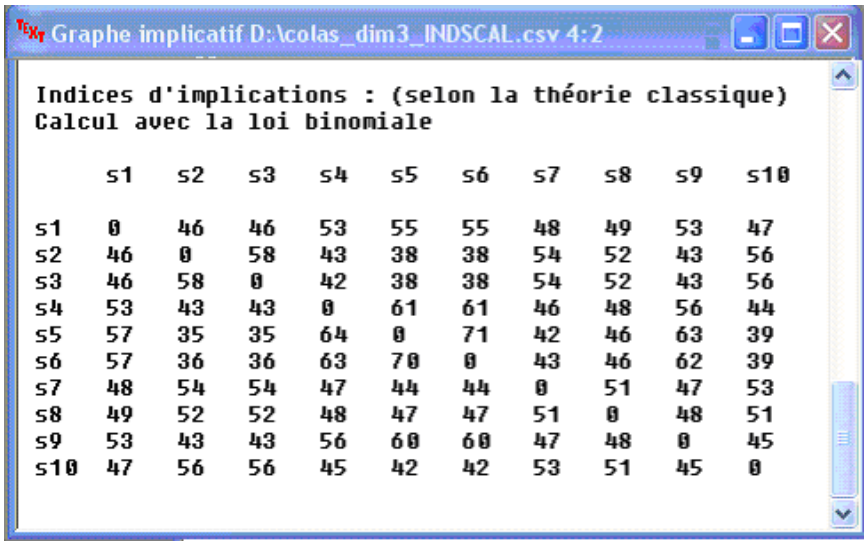

FIGURE 6 – *Indices d'implication des données colas, copie d'écran de CHIC*

Pour le graphe implicatif (Figure 7, droite), nous avons choisi l'affichage automatique, mais avec le seuil par défaut (90), aucun sujet ne s'affichait. Au fur et à mesure que nous le diminuions, des sujets apparaissaient sur deux arbres séparés, un arbre pour chaque groupe, tous les sujets se trouvant dans le bon groupe. C'est seulement avec la valeur de seuil de 50 que tous les sujets sont apparus[10], ce qui a donné 2 arbres rectilignes comme on peut le voir dans la figure 7, chacun correspondant au bon groupe.

3.5 Bilan pour les données colas

On peut remarquer sur ce petit exemple la force de représentation de l'arbre cohésitif comme du graphe implicatif produits par CHIC. A partir de la matrice des

[10] On peut voir dans la figure 6 que le seuil d'implication de 50 sépare bien les sujets en 2 groupes.

coordonnées de 10 sujets sur une sphère de rayon 1 (boule unité dans un espace euclidien à 3 dimensions), CHIC produit deux graphiques qui nous fournissent une séparation beaucoup plus nette entre les deux groupes de sujets que les 2 méthodes de la section précédente, respectivement l'examen d'un tableau d'angles et la projection de la sphère sur les 3 plans de dimension 2.

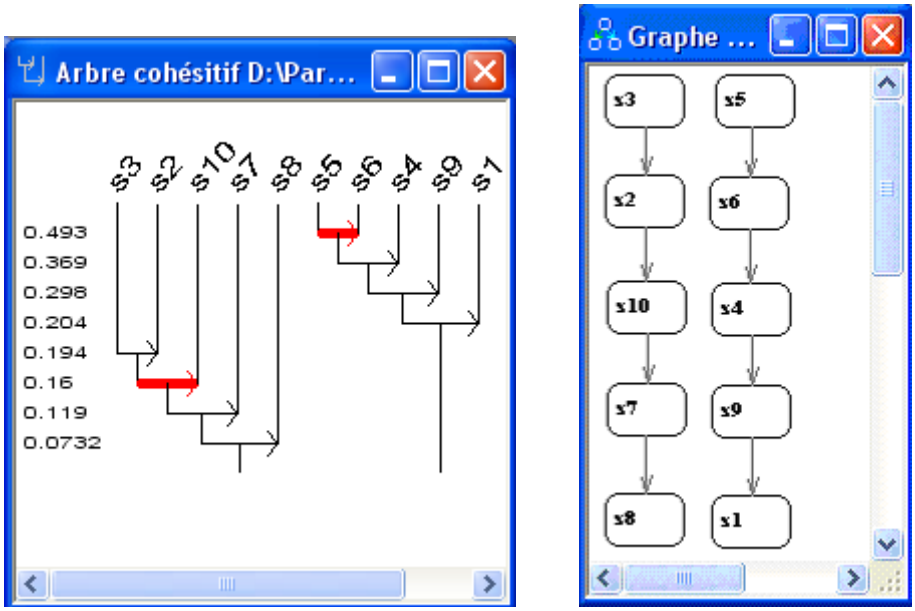

FIGURE 7 – *Arbre cohésitif et graphe implicatif produits automatiquement par CHIC sur les données colas*

Ayant constaté que CHIC nous permet de retrouver aisément les 2 groupes, nous l'appliquons dans la section suivante aux matrices construites dans la section précédente, qui sont les coordonnées des 71 points sur des hyper-sphères de rayon 1 dans des espaces à 25 ou 30 dimensions.

4 Utilisation de CHIC pour les données parole.

Nous avons entré dans CHIC les quatre séries de matrices obtenues par INDSCAL en commençant par la dernière, appelée H4, car c'est la plus simple et la plus courte (le locuteur prononce successivement et rapidement les 6 mots /aku/, /iku/, /uku/, /atu/, /itu/ et /utu/, avec une courte pose entre chacun. L'arbre cohésitif s'est bien dessiné, mais il s'est avéré plus difficile à lire que pour les données colas, les 71 étiquettes étant écrites sur la même ligne se chevauchaient. Une fois la surface de travail agrandie, les étiquettes se lisaient très bien (Figure. 12, en Annexe), mais les groupes se détachaient moins bien que dans le graphe implicatif. C'est donc ce dernier que nous avons privilégié, avec un seuil de 80.

4.1 Le graphe implicatif de H4

On a recopié dans la figure 8 le graphe implicatif de la série H4. Il a été dessiné automatiquement, puis un peu tassé en hauteur en remontant à la souris des étiquettes verticalement afin de faciliter la lecture.

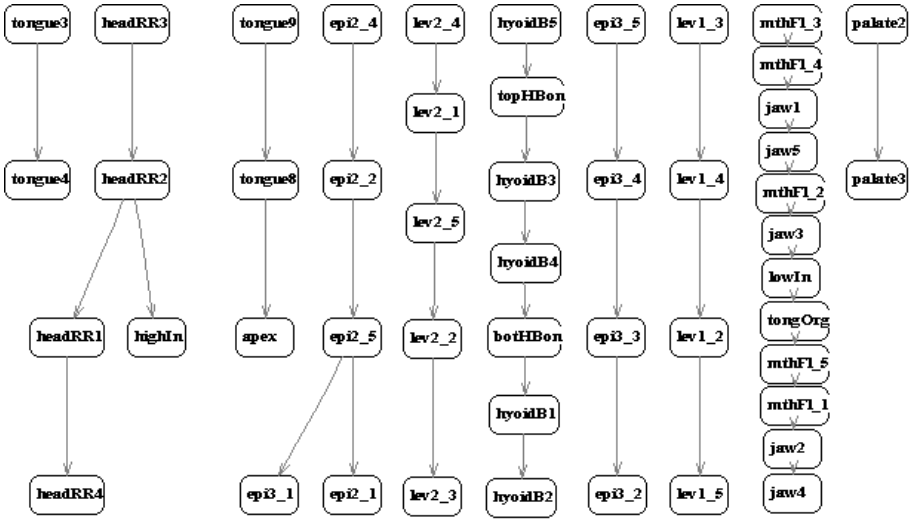

FIGURE 8 – *Le graphe implicatif de la matrice H4 (71 points sur 25 dimensions)*

En allant de la gauche vers la droite de ce graphique, on obtient les 10 groupes bien séparés qui sont assez proches des contours de départ :

1. milieu de la langue (tongue3, tongue4) ;
2. os haut de la tête (headRR1 à headRR5, sauf le dernier), et incisive supérieure (highIn);
3. extrémité de la langue (tongue9, tongue8, apex) ;
4. épiglotte horizontale (epi2_1 à epi2_5, sauf epi2_3), et début de celle du fond du palais (epi3_1) ;
5. lèvre inférieure lev2_1 à lev2_5;
6. os hyoïde (hoydB1 à hoydB5, et ses parties inférieure (botHBon) et supérieure (topHBon) ;
7. partie arrière de l'épiglotte (epi3_1 à epi3_5, sauf epi3_1) ;
8. lèvre supérieure lev1_1 à lev1_5, sauf lev1_1 ;
9. mâchoire inférieure (jaw1 à jaw5), le plancher de la langue (mthFl_1 à mthFl_5), incisive inférieure (lowIn), et balise qui a servi à recaler les images (tongOrg) ;
10. palais (palate2 et palate3)

Nous voyons que le graphe implicatif a retrouvé une grande partie des liaisons entre points que nous connaissions déjà, dues à leur appartenance à un même contour. Plus généralement, il semble mettre en évidence les points dont les mouvements sont liés lors de la production de sons particuliers. Comme les sons de la séquence H3 sont les mêmes que ceux de la séquence H4, mais produits moins vite, nous allons maintenant comparer le graphe implicatif de H3 à celui de H4.

4.2 Le graphe implicatif de H3

Le graphe implicatif de H3 est en figure 9.

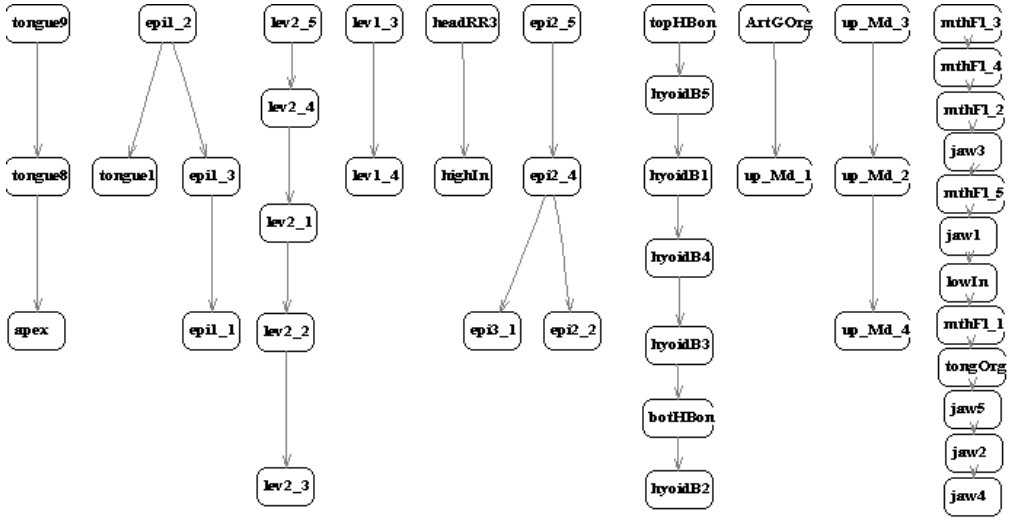

FIGURE 9 – *Le graphe implicatif de la matrice H3 (71 points sur 25 dimensions)*

De la gauche vers la droite, voici les groupes (on reprend les numéros de groupes de H4, sauf quand ce sont de nouveaux groupes) :

1. le groupe 3 ;
2. un nouveau groupe, 11, qui est le début de la partie 1 de l'épiglotte, à l'avant, ainsi que la racine de la langue ;
3. le groupe 5 ;
4. le groupe 8 très réduit ;
5. le groupe 2 très réduit ;
6. le groupe 4 un peu réduit ;
7. le groupe 6 ;
8. un nouveau groupe, 12, formé d'une balise qui a servi à recaler les images (ArtGOrg) et du début de la mâchoire supérieure (up_Md_1) ;
9. un nouveau groupe, 13, du milieu de la mâchoire supérieure (up_Md_2 à up_Md_4) ;
10. le groupe 9.

On peut conclure de cette comparaison que les deux graphes ont une structure très proche avec sept groupes sur dix en commun, avec parfois un peu moins de points dans ceux de H3 que dans ceux de H4. Le graphe implicatif parait une représentation non seulement lisible, mais stable des mouvements du conduit vocal à l'origine de la phrase prononcée dans H3 et H4.

4.3 Les graphes implicatifs de H2 et H1

Dans la séquence H2, la phrase prononcée est formé également de 6 « mots » mais difficilement prononçables, dont seul le centre varie en se compliquant de plus en plus. La séquence H1 est composée des mêmes mots, prononcés plus lentement. On a représenté en figure 10 le graphe implicatif de H2, et en figure 11 le graphe implicatif de H1. Le premier contient 7 groupes et le second en contient 8, les 7 groupes sont

communs aux deux graphes, avec parfois moins de points dans les groupes de H1 que ceux de H2.

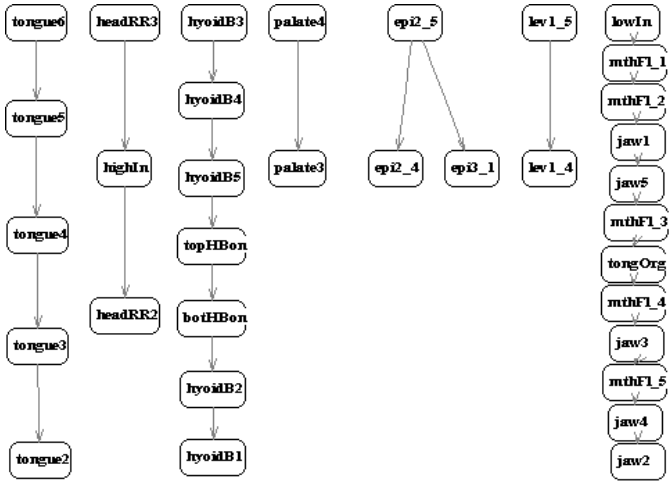

FIGURE 10 – *le graphe implicatif de la matrice H2 (71 points sur 25 dimensions)*

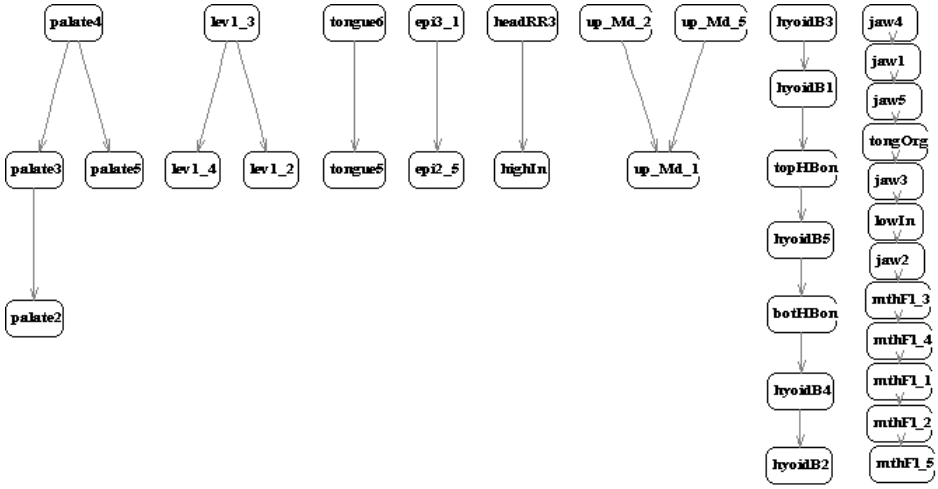

FIGURE 11 – *Le graphe implicatif de la matrice H1 (71 points sur 25 dimensions)*

On retrouve donc une différence de même type entre H1 et H2 que celle trouvée entre H3 et H4. Si on compare maintenant les groupes importants des graphes de H1 et H2 à ceux de H3 et H4, on retrouve celui de l'os hyoïde (7 points) et celui de la mâchoire inférieure (12 points) inchangés. Par contre, celui de la lèvre inférieure (5 points dans H3 et H4) n'apparaît ni dans le graphe de H1, ni dans celui de H2. Cela renforce l'interprétation que nous avons faite à la suite de la lecture des graphes de H3 et H4 : le graphe implicatif est une sorte de décomposition de la parole produite en mouvements des articulateurs, en faisant apparaître ceux qui bougent de façon concertée. Contrairement à H3 et H4, la prononciation des mots de H1 et H2 ne nécessite que très peu l'usage de la lèvre inférieure.

5 Conclusions et perspectives

CHIC a permis de réaliser une tâche complexe d'analyse de séquence d'images : démêler, identifier les "acteurs", dotés ici d'une certaine autonomie, mais aussi très liés

par des contraintes géométriques et mécaniques, dans l'espace et dans le temps, de deux "scènes" ciné-radiographiques, à savoir deux phrases prononcées par un sujet. Et ceci de façon stable, que la phrase ait été prononcée vite ou lentement.

Il reste maintenant à examiner plus finement les résultats produits, au niveau de chaque phrase, de chaque articulateur, de chacune des 25 ou 30 dimensions retenues.

Remerciements

A l'équipe Parole, son responsable Yves Laprie et ses membres, pour leur aide, et à tous les annotateurs bénévoles de ces données, de 7 à 77 ans, sans qui ce travail n'aurait pu être réalisé.

Références

[1] Busset, J. (2007). *Analyse acoustique-articulatoire des fricatives*. (Mémoire de Master 2 Ingénierie Mathématique et Outils Informatiques). Université Henri Poincaré, Nancy 1.

[2] Caroll J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling*: *Theory and applications in the behavorial sciences. Vol. 1 Theory.* New York : Academic Press.

[3] Flanagran J. L. (1972). *Speech Analysis, Synthesis and Perception*. New York : Springer-Verlag.

[4] Haton, J.-P., Cerisara, C., Fohr, D., Laprie, Y., et Smaïli, K. (2006). *Reconnaissance automatique de la parole. Du signal à l'interprétation.* Paris : Dunod.

[5] Laprie, Y., et Busset, J. (2011). Construction and evaluation of an articulatory model of the vocal tract, *19th European Signal Processing Conference - EUSIPCO-2011*, Barcelona, Espagne

[6] Schiffman, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to Multidimensional Scaling*. London : Academic Press.

[7] Tournois, J., et Dickes, P. (1993). *Pratique de l'échelonnement multidimensionnel : de l'observation à l'interprétation.* Bruxelles : De Boeck université.

[8] Article « coarticulation », http://fr.wikipedia.org/wiki/Coarticulation, 14/10/2012, 11h45.

[9] Logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive), http://www.ardm.asso.fr/CHIC.html , 14/10/2012, 11h45.

Annexe : Arbre cohésitif des 71 contours de H4