

J. P. BENZÉCRI

CH. FÉGHALI

**L'articulation des voyelles : analyse sur
microordinateur et interprétation des résultats en
comparaison avec ceux issus d'une analyse ternaire**

Les cahiers de l'analyse des données, tome 8, n° 2 (1983),
p. 159-180

http://www.numdam.org/item?id=CAD_1983__8_2_159_0

© Les cahiers de l'analyse des données, Dunod, 1983, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

L'ARTICULATION DES VOYELLES : ANALYSE SUR MICROORDINATEUR ET INTERPRÉTATION DES RÉSULTATS EN COMPARAISON AVEC CEUX ISSUS D'UNE ANALYSE TERNAIRE

[ARTICULATION VOYELLES]

par J.P. Benzécri (1)
et Ch. Féghali (2)

Comme l'atteste la longueur de son titre, cet article poursuit de multiples fins, Expliquons sa genèse !

Dans son n° 10 (juillet 1981), la revue *Behaviormetrika* (Boite Postale 5030 ; Tokyo International ; Tokyo 100-31) publie un article de J.B. Kruskal intitulé "Multilinear Models for Data Analysis". Le tableau de données d'après lequel est illustré cet article (cf. § 1) nous a paru offrir matière à des analyses factorielles et à des C.A.H. susceptibles de fournir des résultats sommaires mais intéressants pour la phonétique articulatoire (cf. § 3.4). D'autre part, (cf. § 2) la taille du tableau permettait de publier complètement les listages de sorties : c'était l'occasion d'avoir recours à un ensemble de programmes écrits en langage BASIC et exécutables sur un micro-ordinateur. Enfin le thème propre de l'article cité était l'analyse des tableaux à triple entrée (tableaux parallélépipédiques) par une procédure spécifique appelée PARAFAC CALDECOMP : l'analyse des correspondances multiples étant un souci majeur des statisticiens, il valait la peine de comparer les résultats présentés par J.B. Kruskal avec ceux de l'analyse des divers tableaux de correspondance binaire que nous avons construits à partir du tableau ternaire initial (cf. § 3).

1 Le tableau des données : L'observation aux rayons X a permis de mesurer exactement sur des clichés la position de la langue relativement au palais et au pharynx. D'où le tableau ternaire suivant que J.B. Kruskal publie en renvoyant pour plus de détails à une revue d'acoustique (cf. Harshman, R.A. Ladefoged, P. ; & Goldstein, L. in Journal of the Acoustical Society of America ; 62 pp 693-707 ; 1977).

I = Un ensemble de 10 phonèmes vocaliques de la langue anglaise (dans sa prononciation aux U.S.A.).

J = Un ensemble de 13 positions le long de la langue.

K = Un ensemble de 5 sujets (lesquels, à en juger par l'illustration de l'article de J.B. K., pourraient être des femmes).

$k(i,j,k)$ = largeur du tractus vocal au niveau de la position j quand le sujet s prononce le son i.

N.B. Les lecteurs intéressés par les programmes d'analyse des données en langage BASIC peuvent s'adresser à Ch. Féghali (laboratoire de statistique) Université Pierre et Marie Curie. (4 place Jussieu - 75230 PARIS CEDEX 05).

(1) Professeur de statistique. Université Pierre et Marie Curie.

(2) Docteur 3° cycle, Chercheur-Informaticien C. N. R. S.

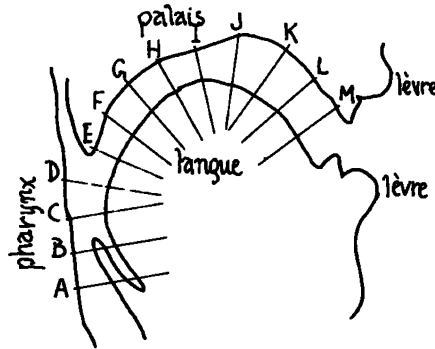
Harshman et coll. précisent que les sujets devaient prononcer des phrases de la forme suivante :

"Say *h* (vowel) *d* again" (i.e. répétez le mot *h* (voyelle) *d*) ; les voyelles étant :

/ i, ɪ, eɪ, , x, a, o, à, , u/ ,

comme dans "heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd".

Pour chacune des voyelles, on a choisi une seule image comme figurant sa prononciation par chacun des sujets : choix délicat, guidé pour les voyelles /ɪ, e, x, a, / par la stabilité du 2-ème formant ; ou, lorsque la prononciation présentait constamment les caractères d'une diphtongue, choix fixé à 30 ms (millisecondes) environ après la première consonne *h*. Du point de vue de l'analyse des correspondances il aurait été particulièrement souhaitable de posséder les descriptions des positions successives de la langue lors de la prononciation d'une diphtongue.



I. {i, ɪ, e, ε, ae, a, o, ω, u}

J. {A, B, C, D, E, F, G, H, I, J, K, L, M}

S. {1, 2, 3, 4, 5}

[Articulation Voyelles] § 1 :
Explication des données d'après Harshman & coll.

On doit encore souligner combien il est difficile de définir avec précision des positions le long de la langue ; les mesures de largeur elles-mêmes, sont définies par Harman et coll., en fonction de leur objectif propre, qui est l'étude des formes de la langue : il ne s'agit pas à proprement parler de largeur (dimension sagittale), mais de ce que serait la largeur si la paroi supérieure du tractus vocal était immobile ; condition qui est réalisée en avant (palais dur), mais non en arrière (voile du palais et pharynx). Enfin l'espace compris entre les lèvres a été mesuré, mais ces mesures ne sont ni publiées ni analysées par les auteurs cités.

Pour l'analyse des correspondances, le tableau ternaire $I \times J \times S$ sera considéré comme un tableau binaire $((I \times S), J)$ avec

$$k((i, s), j) = k(i, j, s)$$

la ligne (i, s) décrivant par un profil sur J la forme du tractus vocal du sujet s prononçant le son i . A ce tableau, s'ajoutent les deux tableaux de marge binaire $I \times J$ et $S \times J$; définis par :

$$k(i, j) = \Sigma\{k(i, j, s) \mid s \in S\} ;$$

$$k(s, j) = \Sigma\{k(i, j, s) \mid i \in I\}.$$

En bref le tableau $k(i, j)$ donne pour chaque son i un profil moyen du tractus vocal calculé sur les 5 locuteurs, et le tableau des $k(s, j)$ donne pour chaque locuteur s le profil moyen de son tractus vocal quand il prononce l'ensemble des 10 sons considérés. Nous n'avons pas analysé le tableau de marge $I \times S$: lequel donne pour chaque son i prononcé par un sujet s ; une largeur moyenne sur toute la longueur du tractus vocal (plus précisément le long de la langue) : cette analyse prendrait de l'intérêt s'il y avait un plus grand nombre de sujets : dans la suite on se bornera à signaler pour quels sons chaque sujet réalise le maximum et le minimum de la largeur moyenne de son tractus vocal.

On a effectué trois analyses factorielles : chacune prend pour tableau principal l'un des trois tableaux $(I \times J)$; $((I \times S) \times J)$; ou $(S \times J)$; les deux autres étant alors en supplémentaires. De plus on a effectué des C.A.H. sur $I \times S$ et J d'après le tableau rectangulaire $((I \times S) \times J)$.

2 Listage commenté des résultats

2.1 Analyse factorielle : On reproduit sur trois pages successives:

1°) La liste, sous forme conversationnelle des demandes du programmeur : on remarquera particulièrement le tableau des caractéristiques des graphiques.

2°) Le tableau des valeurs propres et taux ; le tableau des facteurs pour l'ensemble des lignes principales : ensemble $I \times S$, selon les notations du § 2; les réalisations des voyelles sont désignées par un chiffre (numéro du sujet) ; suivi de la lettre V (initiale de voyelle) ; puis du numéro de la voyelle : depuis $1 = i$, jusqu'à 10 (chiffre \emptyset) = u .

3°) Le tableau des facteurs pour les lignes supplémentaires : ensembles I des voyelles (de VOY1 à VOY \emptyset) et ensemble S des sujets (de IND1 à IND5) ; le tableau des facteurs pour les colonnes principales (ensemble des positions de A à M le long de la langue. Enfin un petit graphique du plan 1×2 : les points

superposés sont signalés ligne par ligne ; avec pour identifier chaque ligne le dernier identificateur qui y est effectivement imprimé. Au § 4, on trouvera le calque d'un graphique à plus grande échelle du plan 1×2 ; complété d'annotations manuscrites.

2.2 Classification automatique : Le listage publié occupe 6 pages:

1°) Résultats de la C.A.H. sur l'ensemble J des positions le long de la langue. On notera qu'avec l'histogramme des indices de niveau, on a non seulement l'indice de niveau $I(N)$ et le taux $TX(N)$ afférent au dipole formé par le centre de gravité des deux successeurs immédiats de N ($A(N)$ et $B(N)$) ; mais aussi l'inertie totale de la classe N: $I.TOT$; laquelle est toujours supérieure à $I(N)$; et est en particulier égale à l'inertie totale du nuage (ou somme des indices de niveau) si N est au sommet de la hiérarchie (Ici : $I.TOT = 159$, sur la ligne $N = 25$, correspond à "Somme des indices de niveau = .158771, en tête du listage). Une autre particularité est que, afin de rendre l'arbre plus facile à lire, on a donné aux noeuds un niveau proportionnel non à $I(N)$ (ou à $TX(N)$) mais à la racine carrée de cet indice : ce qui accroît l'importance relative des dénivellations de la base de l'arbre.

2°) L'histogramme de la C.A.H. sur l'ensemble $I \times S$ (voyelles sujets).

3°) La composition des classes de cette C.A.H. .

4°) Les dessins de l'arbre (faits suivant ces mêmes conventions que pour la C.A.H. sur J).

5°) Les aides à l'interprétation, en analyse factorielle (programme FACOR) pour une partition en 5 classes ; qui a été retenue parce qu'elle groupe les voyelles prononcées par les divers sujets, de façon à peu près ordonnée ; sans qu'il soit possible, par une subdivision plus fine, de mettre en évidence des classes plus homogènes.

6°) Pour chacun des quatre noeuds N retenus, on a figuré comme deux histogrammes, les abscisses des individus des deux classes $A(N)$ et $B(N)$, sur la ligne joignant les centres de gravité de ces classes : il est remarquable que ces histogrammes n'empiètent qu'exceptionnellement. Ce qui signifie qu'il est possible de séparer les deux classes $A(N)$ et $B(N)$ par un hyperplan perpendiculaire à la ligne joignant leurs centres de gravité. De plus (en désignant simplement par N, $A(N)$, $B(N)$ les centres de gravité des classes) on note que l'hyperplan de séparation n'est pas l'hyperplan perpendiculaire à $A(N) B(N)$ en N ; mais plutôt l'hyperplan médiateur de $A(N) B(N)$; en d'autres termes, les histogrammes (histogramme supérieur afférent à la classe $A(N)$; et histogramme inférieur, afférent à la classe $B(N)$) se touchent au milieu de $A(N)$, $B(N)$; ou encore, dans l'espace, un individu de la classe N appartient à $A(N)$ ou à $B(N)$ selon qu'il est plus proche du centre de gravité de l'une ou l'autre classe. Evidemment, les classes étant identifiées par voie ascendante, il ne peut s'agir que d'une règle approchée, souffrant des exceptions. Toutefois, chaque fois qu'en C.A.H. on a utilisé des histogrammes comme aide à l'interprétation des noeuds, cette règle approchée est apparue satisfaisante : la raison est, selon nous, que la C.A.H. visant à minimiser la variance intraclasse, elle fournit des résultats approchant de l'affectation au centre le plus proche ; laquelle comme on le sait depuis E. Diday, permet d'atteindre itérativement un minimum (local) de la variance intra-classe.

N.B. Dans le présent listage, l'histogramme donne la position des individus désignés chacun par son sigle, ~~par une autre~~ portion du même programme, conçue pour un grand nombre d'individus, l'histogramme est simplement formé de colonnes de sigles *, dont chacun désigne un nombre d'individus qui peut être choisi par l'utilisateur.

QUEL EST LE TITRE DE L'ANALYSE ?
VOYELLES ET LANGUE

QUEL EST LE NOM DU FICHIER ?
VOX

VOICI LE NOM DES LIGNES

| | | | | | | | | | | |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|---------|
| 101(1) | 102(2) | 103(3) | 104(4) | 105(5) | 106(6) | 107(7) | 108(8) | 109(9) | 110(10) | 111(11) |
| 202(12) | 203(13) | 204(14) | 205(15) | 206(16) | 207(17) | 208(18) | 209(19) | 200(20) | 301(21) | 302(22) |
| 303(23) | 304(24) | 305(25) | 306(26) | 307(27) | 308(28) | 309(29) | 300(30) | 401(31) | 402(32) | 403(33) |
| 404(34) | 405(35) | 406(36) | 407(37) | 408(38) | 409(39) | 400(40) | 501(41) | 502(42) | 503(43) | 504(44) |
| 505(45) | 506(46) | 507(47) | 508(48) | 509(49) | 500(50) | 601(51) | 602(52) | 603(53) | 604(54) | 605(55) |
| 1006(56) | 1007(57) | 1008(58) | 1009(59) | 1000(60) | 1101(61) | 1102(62) | 1103(63) | 1104(64) | 1105(65) | |

QUEL EST LE NOMBRE DE LIGNES SUPPLEMENTAIRES ? 15

QUEL NUMERO DE LIGNE ? 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65

VOICI LE NOM DES COLONNES

| | | | | | | | | | | |
|-------|-------|------|------|------|------|------|------|------|-------|-------|
| A(1) | B(2) | C(3) | D(4) | E(5) | F(6) | G(7) | H(8) | I(9) | J(10) | K(11) |
| L(12) | M(13) | | | | | | | | | |

QUEL EST LE NOMBRE DE COLONNES SUPPLEMENTAIRES ? 0

COMBIEN DE FRATELUS VOULEZ-VOUS ? ?

VOULEZ-VOUS LA SORTIE DU TABLEAU DE L'ANALYSE ? 1

VOULEZ-VOUS GARDER LE NOMBREU FICHIER ? 0

QUEL CODE A LA SORTIE DU TABLEAU DE L'ANALYSE DES LIGNES ? 1

QUEL CODE A LA SORTIE DU TABLEAU DE L'ANALYSE DES COLONNES ?

COMBIEN VOULEZ-VOUS DE GRAPHIQUES ? 4

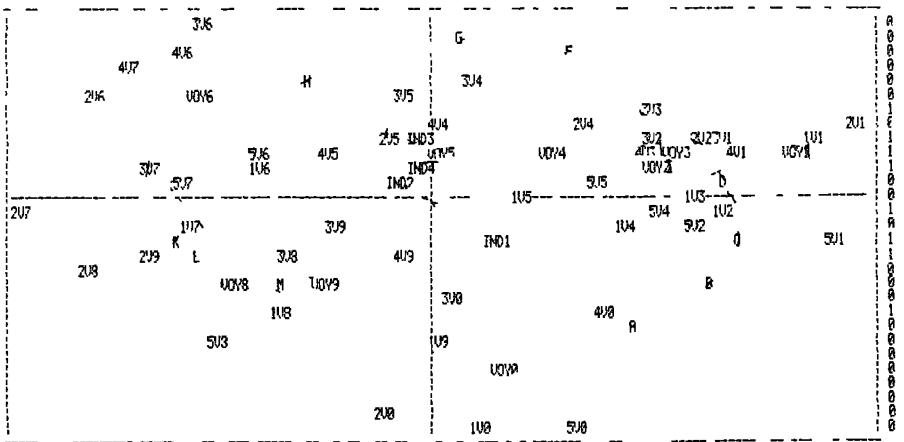
CARACTERISTIQUES DES GRAPHIQUES

| CARACTERISTIQUES | GRAP 1 | GRAP 2 | GRAP 3 | GRAP 4 |
|-------------------------------------|--------|--------|--------|--------|
| NUMBRE DE CARACTERES GRAPHIQUES | 4 | 4 | 4 | 4 |
| AXE HORIZONTAL | 1 | 1 | 2 | 2 |
| AXE VERTICAL | 2 | 3 | 3 | 4 |
| TYPE DE REPRESENTATION DES LIGNES | 1 | 1 | 1 | 1 |
| TYPE DE REPRESENTATION DES COLONNES | 2 | 2 | 2 | 2 |
| POINTS SUPERPOSES | 1 | 1 | 1 | 1 |
| DIMENSION DES GRAPHIQUES | 1 | 1 | 1 | 1 |
| COLONNE 9 | 1 | 1 | 1 | 1 |

| ISUPP | IDEN | GLT | POS | INR | #1F | COR | CTR | #2F | COR | CTR | #3F | COR | CTR | #4F | COR | CTR | #5F | COR | CTR | #6F | COR | CTR | #7F | COR | CTR |
|-------|------|------|-----|-----|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11 | UOV1 | 1808 | 98 | 193 | 557 | 987 | 231 | 53 | 9 | 13 | -20 | 1 | 11 | 2 | 0 | 1 | 25 | 2 | 78 | 4 | 0 | 3 | 1 | 0 | 0 |
| 12 | UOV2 | 999 | 95 | 71 | 541 | 933 | 84 | 48 | 13 | 7 | 14 | 2 | 5 | -3 | 0 | 1 | -4 | 0 | 3 | -7 | 1 | 14 | 0 | 0 | 0 |
| 13 | UOV3 | 1808 | 99 | 88 | 371 | 975 | 104 | 52 | 19 | 13 | -24 | 4 | 15 | 7 | 0 | 3 | 3 | 0 | 1 | 14 | 1 | 39 | 0 | 0 | 0 |
| 14 | UOV4 | 998 | 99 | 22 | 175 | 857 | 23 | 69 | 132 | 23 | 17 | 8 | 7 | 2 | 0 | 0 | 1 | 0 | 0 | -3 | 0 | 3 | -1 | 0 | 1 |
| 15 | UOV5 | 967 | 104 | 4 | -2 | 1 | 0 | 69 | 694 | 23 | -37 | 217 | 39 | 14 | 28 | 14 | 9 | 14 | 11 | 11 | 22 | 32 | 4 | 2 | 5 |
| 16 | UOV6 | 1808 | 119 | 129 | -783 | 873 | 174 | 157 | 143 | 142 | 21 | 2 | 13 | -7 | 0 | 6 | 0 | 0 | 0 | 2 | 0 | 2 | -7 | 0 | 26 |
| 17 | UOV7 | 1808 | 106 | 165 | -494 | 991 | 199 | 41 | 7 | 9 | 8 | 0 | 2 | 13 | 1 | 12 | 15 | 1 | 27 | 2 | 0 | 1 | 4 | 0 | 7 |
| 18 | UOV8 | 999 | 98 | 82 | -337 | 854 | 86 | -138 | 144 | 91 | 0 | 0 | 0 | 5 | 0 | 2 | -1 | 0 | 0 | 8 | 0 | 13 | 2 | 0 | 2 |
| 19 | UOV9 | 995 | 94 | 32 | -198 | 664 | 26 | -132 | 328 | 80 | 22 | 9 | 11 | -1 | 0 | 0 | 5 | 1 | 4 | -7 | 1 | 17 | 2 | 0 | 1 |
| 18 | UOV8 | 1808 | 88 | 58 | 98 | 186 | 6 | -288 | 879 | 337 | 0 | 0 | 0 | -25 | 8 | 42 | -23 | 6 | 57 | 9 | 1 | 16 | -1 | 0 | 2 |
| 111 | IND1 | 992 | 184 | 16 | 82 | 588 | 9 | -74 | 414 | 58 | 22 | 34 | 22 | -28 | 31 | 55 | 7 | 4 | 12 | -9 | 8 | 41 | 7 | 1 | 6 |
| 112 | IND2 | 996 | 218 | 18 | -69 | 359 | 8 | 16 | 19 | 3 | -98 | 823 | 441 | -9 | 7 | 14 | -9 | 7 | 24 | 4 | 1 | 6 | 0 | 0 | 0 |
| 113 | IND3 | 995 | 198 | 13 | 37 | 172 | 2 | 77 | 546 | 55 | 51 | 244 | 133 | 9 | 7 | 10 | -23 | 53 | 138 | 9 | 7 | 32 | 9 | 7 | 53 |
| 114 | IND4 | 993 | 221 | 6 | -32 | 253 | 2 | 47 | 512 | 23 | 0 | 0 | 0 | -25 | 157 | 183 | 11 | 28 | 38 | 3 | 3 | 5 | -3 | 4 | 17 |
| 115 | IND5 | 997 | 187 | 19 | 76 | 359 | 8 | -80 | 482 | 59 | 26 | 42 | 33 | 53 | 171 | 368 | 16 | 16 | 56 | -6 | 3 | 21 | -7 | 4 | 41 |

| I | PP | IDEN | GLT | POS | INR | #1F | COR | CTR | #2F | COR | CTR | #3F | COR | CTR | #4F | COR | CTR | #5F | COR | CTR | #6F | COR | CTR | #7F | COR | CTR |
|-----|----|------|-----|-----|------|-----|-----|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11 | A | 1808 | 92 | 87 | 717 | 662 | 78 | -281 | 269 | 188 | 77 | 39 | 133 | -58 | 17 | 164 | 26 | 4 | 72 | 31 | 7 | 283 | 14 | 1 | 61 | |
| 12 | AR | 999 | 74 | 183 | 442 | 891 | 118 | 149 | 181 | 79 | 29 | 4 | 16 | -13 | 1 | 9 | 6 | 0 | 3 | 41 | 8 | 267 | -27 | 4 | 217 | |
| 13 | AC | 996 | 69 | 186 | 485 | 96 | 124 | -69 | 28 | 16 | -11 | 1 | 3 | 32 | 4 | 51 | -16 | 1 | 27 | 38 | 4 | 128 | 7 | 0 | 12 | |
| 14 | AD | 997 | 78 | 95 | 473 | 978 | 112 | 24 | 3 | 2 | -13 | 1 | 4 | 53 | 16 | 166 | -70 | 4 | 76 | 7 | 0 | 2 | 26 | 3 | 175 | |
| 15 | AE | 995 | 73 | 64 | 543 | 842 | 64 | 134 | 129 | 64 | -28 | 3 | 9 | 39 | 11 | 75 | -2 | 0 | 1 | -35 | 9 | 282 | -11 | 1 | 40 | |
| 16 | AF | 992 | 71 | 46 | 222 | 478 | 27 | 227 | 482 | 169 | -42 | 18 | 33 | -24 | 6 | 32 | 2 | 0 | 1 | -28 | 4 | 64 | -21 | 5 | 122 | |
| 17 | AG | 999 | 69 | 31 | 75 | 17 | 1 | 255 | 898 | 216 | -13 | 3 | 4 | -67 | 65 | 225 | 6 | 0 | 3 | 19 | 5 | 51 | 0 | 0 | 0 | |
| 18 | AH | 998 | 69 | 35 | -288 | 573 | 21 | 185 | 427 | 114 | 48 | 29 | 41 | -8 | 1 | 4 | 38 | 11 | 72 | 21 | 6 | 65 | 29 | 10 | 218 | |
| 19 | AI | 995 | 69 | 87 | -431 | 939 | 98 | 79 | 31 | 28 | 65 | 21 | 76 | 45 | 10 | 99 | 18 | 2 | 25 | 7 | 0 | 7 | -7 | 0 | 3 | |
| 110 | IJ | 999 | 81 | 183 | -438 | 953 | 128 | -33 | 6 | 5 | 72 | 25 | 187 | 37 | 7 | 78 | 8 | 0 | 6 | -5 | 0 | 7 | -17 | 2 | 97 | |
| 111 | IK | 997 | 98 | 188 | -418 | 952 | 117 | -65 | 25 | 19 | 24 | 3 | 13 | -3 | 0 | 1 | 32 | 6 | 111 | 4 | 0 | 4 | -8 | 0 | 28 | |
| 112 | IL | 999 | 88 | 88 | -376 | 892 | 96 | -94 | 57 | 38 | -55 | 28 | 71 | -35 | 9 | 86 | -56 | 28 | 324 | 0 | 0 | 0 | 10 | 1 | 34 | |
| 113 | IM | 1808 | 84 | 55 | -246 | 576 | 39 | -137 | 188 | 77 | -149 | 214 | 486 | 14 | 2 | 12 | 54 | 28 | 285 | 0 | 0 | 0 | 2 | 0 | 1 | |

APE HORIZONTALE (1) ; LARGEUR = 1,74335
 APE VERTICALE (2) ; HAUTEUR = ,632201
 NOMBRE DE POINTS REPRESENTES = 78



DESCRIPTION DES POINTS SUPERPOSES DU GRAPHIQUE

- LES POINTS SUIVANTS
- 313 IE
 - 215 IO13
 - UOV4 IO2
 - 317 IU17
 - 217 IO13
 - 511 IO15
 - IL IO13
 - 118 IO17

2.3 Première interprétation : Nous nous bornerons à noter ici l'accord entre les résultats apparus dans le plan 1×2 , et les structures bien connues des phonéticiens : quant aux différences interindividuelles et à l'axe 3, nous y reviendrons au § 3, en confrontant l'analyse des correspondances binaires avec la méthode PARAFAC CANDECAMP destinée spécifiquement aux tableaux ternaires.

Notons d'abord que les axes 1 et 2 issus respectivement des analyses où le tableau principal est $I \times J$ et $(I \times S) \times J$, ne diffèrent pas sensiblement entre eux ; et que les axes 3 diffèrent peu. La seule différence d'une analyse à l'autre concerne les valeurs propres et les taux : en effet on a selon les cas :

$\lambda_\alpha = \text{Inertie}(N(I))$ sur l'axe α ; $\lambda'_\alpha = \text{Inertie}(N(I \times S))$ sur l'axe α ;

$$\tau_\alpha = \lambda_\alpha / \text{Itot}(N(I)) \quad ; \quad \tau'_\alpha = \lambda'_\alpha / \text{Itot}(N(I \times S)) .$$

L'analyse fondée sur le tableau $S \times J$ elle-même, ne fournit pas des axes très différents, bien que soit très faible l'inertie de $N(S)$ (nuage des sujets caractérisés chacun par la forme moyenne de son tractus vocal).

Quant aux termes d'inertie il est tout à fait remarquable que pour le tableau principal $I \times J$ on ait : $\tau_1 + \tau_2 = 99,5\%$. Cela signifie que les 12 profils du tractus vocal associés à l'ensemble I des 10 voyelles (ces profils étant ici chacun défini comme une moyenne sur 10 sujets) ne dépendent quasi rigoureusement que de deux paramètres articulatoires ; qu'on peut identifier aux facteurs 1 et 2. Ce résultat même s'il s'accorde avec la représentation classique du triangle des voyelles (cf. fig. § 3.3), représentation bidimensionnelle, n'était aucunement prévisible : il conviendrait d'étendre la présente analyse aux profils articulatoires des consonnes, pour savoir si ceux-ci se placent dans le plan défini par les profils des voyelles.

Venons à la lecture du plan 1×2 . L'ensemble J des positions (marquées d'arrière en avant par les lettres A à M) suggère une ligne continue, dont les extrémités antérieures et postérieures se plient. L'ensemble I des voyelles comprend deux branches {1-6} et {7-10} ; la première, située dans le demi-plan $F_2 > 0$, est à peu près parallèle à l'axe 1 ; la deuxième, coupant le quadrat $\{F_1 < 0, F_2 < 0\}$, fait avec l'axe 2 un angle d'environ -30° . C'est à peu près le triangle classique ; que les phonéticiens présentent d'ordinaire en lui donnant pour base le segment joignant les voyelles extrêmes (1 = i ; et 10 = u ; i.e. "ou" selon la lecture française), et pour le sommet le a (voyelle 6). Il est facile de reconstituer le profil articulatoire d'un son (c'est-à-dire de dessiner la forme de la langue!) d'après sa position dans le plan 1×2 relativement à la ligne des positions le long de la langue ; sans appliquer la formule de reconstitution des données en fonction des facteurs, on sait qu'un son se place d'autant plus près d'une position, qu'au niveau de celle-ci le tractus vocal est relativement plus large. Le a (6) est proche des positions moyennes (I, H) ; c'est en effet la voyelle d'aperture maxima ; et dans son profil, la largeur au milieu du tractus l'emporte nettement sur toutes les autres ; particulièrement celles au niveau postérieur (A, B, C) ; ces positions étant nettement opposées à la voyelle a sur l'axe 1. Au contraire le i (1) est articulé avec une largeur maxima dans les positions arrières, (A, B, C, D, E) ; la largeur minima correspond aux positions (I, J, K), opposées à la voyelle i sur l'axe 1.

Quant à l'ensemble J, la classification automatique montre des classes qui sont des segments successifs du tractus vocal : on n'attendait rien d'autre.

Sur l'ensemble I x S (profil des voyelles articulées par les divers sujets) la partie supérieure de l'arbre donne des classes qui sans être pures, ne comprennent que des voyelles assez voisines, articulées par tel ou tel sujet. Mais rien n'apparaît qui permette une discrimination entre voyelles indépendamment du sujet qui les articule : cette relative confusion est d'ailleurs bien visible sur le plan 1 x 2 ; où les sous-nuages de 5 points afférents aux voyelles se recouvrent partiellement entre eux. Ici plusieurs interprétations phonétiques se présentent. Il n'est pas sûr d'abord que la prononciation des sujets soit si soignée qu'elle permette véritablement de distinguer 10 voyelles (plus exactement 10 timbres vocaliques indépendamment du contexte ; de la longueur etc.) ; il faudrait pour plus de précision distinguer plusieurs réalisations du même son par un même sujet (cf. § 1)... et faire l'épreuve de la reconnaissance de sons isolés par un auditeur. Même à supposer qu'une prononciation soignée permette de distinguer dix timbres, il reste que les mesures par radiographie sur des parties molles (langue) sont difficiles à aligner de sujet à sujet ; et que dans la production des sons, le profil de la langue n'agit que concurremment avec d'autres dimensions non retenues ici : car d'une part, cf. § 1, les dimensions sagittales du tractus vocal dépendent de la mobilité du voile du palais et du pharynx ; et d'autre part, il y a les dimensions perpendiculaires au plan de symétrie de la tête, suggèrent que le paramètre essentiel est l'aire de la section au niveau de chaque position (de A à M).

On reviendra au § 3.4 sur les différences entre sujets (cf. figure).

3 Analyse binaire et analyse ternaire : En 20 pages, J.B. Kruskal ne peut donner qu'une introduction aux méthodes qu'il préconise, avec un aperçu des résultats obtenus sur un exemple. Il est cependant instructif de tenter une comparaison entre les travaux de

J.B. Kruskal ou Harshman & coll. et nos analyses des tableaux de correspondance binaire.

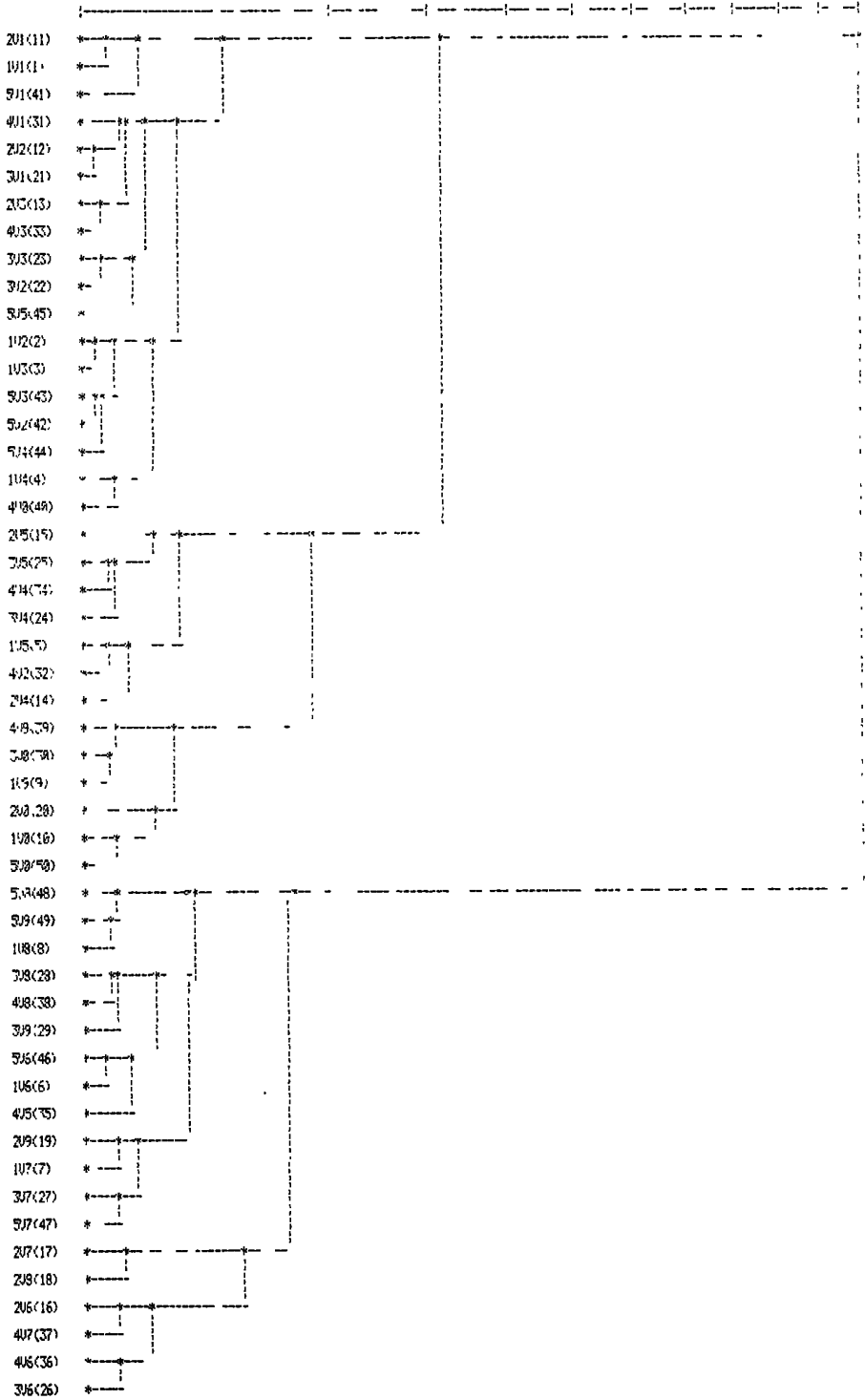
3.1 Le modèle PARAFAC CANDECOMP : En bref, pour les nombres x_{ijs} formant un tableau ternaire I x J x S, on cherche une représentation de la forme

$$x_{ijs} \approx \sum \{a_{ir} b_{jr} c_{sr} \mid r \in R\}$$

R est ce que nous appellerions l'ensemble des facteurs retenus : ici $R = \{1, 2\}$; on ne calcule que deux facteurs pour ces données phonétiques. $\{a_{ir} \mid i \in I\}$ est l'ensemble des valeurs du r-ème facteur sur l'ensemble I ; et de même $\{b_{jr} \mid j \in J\}$ pour l'ensemble J et

$\{c_{sr} \mid s \in S\}$ pour l'ensemble S. Comparant l'analyse du tableau ternaire I x J x S, à celle d'un tableau binaire I x J (calculé comme une moyenne sur l'ensemble des sujets) Harshman et coll. disent en termes fort suggestifs que le coefficient c_{sr} représente l'importance relative du facteur r dans le mode articulatoire propre au sujet s ("represents the relative emphasis of the [r-th] feature by person [s]").

| NOSE | (NO) | (NO) | (NO) | (NO) | REPERTOIRE DES SERIES DE LA SUPERFONCTION |
|------|------|------|------|------|---|
| 99 | 10 | 90 | 88 | 100 | 211 104 71 411 212 211 213 415 217 216 215 117 117 512 512 514 114 418 218 218 218 214 115 412 214 |
| 98 | 121 | 97 | 94 | 171 | 211 118 511 411 212 211 213 415 217 216 215 117 117 512 512 514 114 418 218 218 218 214 115 412 214 |
| 97 | 8 | 89 | 91 | 117 | 211 218 218 417 418 218 |
| 96 | 7 | 95 | 97 | 119 | 211 218 218 417 418 218 |
| 95 | 4 | 83 | 76 | 116 | 211 218 218 417 418 218 |
| 94 | 3 | 89 | 87 | 118 | 211 218 218 417 418 218 |
| 93 | 12 | 88 | 92 | 117 | 211 218 218 417 418 218 |
| 92 | 12 | 87 | 85 | 119 | 211 218 218 417 418 218 |
| 91 | 2 | 79 | 89 | 7 | 211 218 218 417 418 218 |
| 90 | 2 | 85 | 84 | 115 | 211 218 218 417 418 218 |
| 89 | 1 | 85 | 74 | | 211 218 218 417 418 218 |
| 88 | 1 | 71 | 215 | 4 | 211 218 218 417 418 218 |
| 87 | 1 | 77 | 69 | 6 | 211 218 218 417 418 218 |
| 86 | 1 | 72 | 210 | 3 | 211 218 218 417 418 218 |
| 85 | 1 | 73 | 67 | 7 | 211 218 218 417 418 218 |
| 84 | 1 | 81 | 73 | 8 | 211 218 218 417 418 218 |
| 83 | 1 | 66 | 78 | 4 | 211 218 218 417 418 218 |
| 82 | 1 | 81 | 61 | 3 | 211 218 218 417 418 218 |
| 81 | 10 | 815 | 88 | 3 | 211 218 218 417 418 218 |
| 80 | 10 | 68 | 64 | 4 | 211 218 218 417 418 218 |
| 79 | 10 | 214 | 67 | 3 | 211 218 218 417 418 218 |
| 78 | 10 | 56 | 75 | 5 | 211 218 218 417 418 218 |
| 77 | 10 | 415 | 54 | 3 | 211 218 218 417 418 218 |
| 76 | 10 | 218 | 717 | 12 | 211 218 218 417 418 218 |
| 75 | 10 | 53 | 411 | 3 | 211 218 218 417 418 218 |
| 74 | 10 | 68 | 419 | 7 | 211 218 218 417 418 218 |
| 73 | 10 | 418 | 114 | 2 | 211 218 218 417 418 218 |
| 72 | 10 | 810 | 110 | 2 | 211 218 218 417 418 218 |
| 71 | 10 | 314 | 183 | 7 | 211 218 218 417 418 218 |
| 70 | 10 | 417 | 215 | 12 | 211 218 218 417 418 218 |
| 69 | 10 | 319 | 59 | 3 | 211 218 218 417 418 218 |
| 68 | 10 | 317 | 317 | 2 | 211 218 218 417 418 218 |
| 67 | 10 | 57 | 51 | 5 | 211 218 218 417 418 218 |
| 66 | 10 | 316 | 418 | 12 | 211 218 218 417 418 218 |
| 65 | 10 | 58 | 518 | 3 | 211 218 218 417 418 218 |
| 64 | 10 | 117 | 219 | 12 | 211 218 218 417 418 218 |
| 63 | 10 | 414 | 315 | 12 | 211 218 218 417 418 218 |
| 62 | 10 | 412 | 115 | 12 | 211 218 218 417 418 218 |
| 61 | 10 | 111 | 211 | 12 | 211 218 218 417 418 218 |
| 60 | 10 | 119 | 318 | 12 | 211 218 218 417 418 218 |
| 59 | 10 | 418 | 318 | 12 | 211 218 218 417 418 218 |
| 58 | 10 | 118 | 919 | 12 | 211 218 218 417 418 218 |
| 57 | 10 | 514 | 52 | 7 | 211 218 218 417 418 218 |
| 56 | 10 | 413 | 213 | 12 | 211 218 218 417 418 218 |
| 55 | 10 | 317 | 315 | 12 | 211 218 218 417 418 218 |
| 54 | 10 | 115 | 316 | 12 | 211 218 218 417 418 218 |
| 53 | 10 | 311 | 212 | 12 | 211 218 218 417 418 218 |
| 52 | 10 | 512 | 913 | 12 | 211 218 218 417 418 218 |
| 51 | 10 | 113 | 112 | 12 | 211 218 218 417 418 218 |



Il faut signaler que les données ont été centrées en ce sens que :

$$\forall i, s : \sum \{x_{ijs} \mid j \in J\} = 0 ;$$

ce qui, dans les notations de l'analyse des correspondances, peut se traduire :

$$x_{ijs} = d^{ijs} - d^{is} ; \text{ où } ; d^{ijs} = f_{ijs} / (f_{i.} f_{.j} f_{.s}) ; d^{is} = f_{is} / (f_{i.} f_{.s}).$$

D'autres versions du centrage sont possibles. L'auteur (J.B.K.) ne peut s'étendre sur la méthode choisie il dit en bref " which combination to use with a given set of data... should be decided partly by intuition... and partly by trying different methods and comparing the results..." De même c'est par tâtonnements que l'on s'est arrêté à retenir deux facteurs. Et le calcul de ceux-ci, reposant sur une procédure itérative, est beaucoup plus long qu'une analyse de

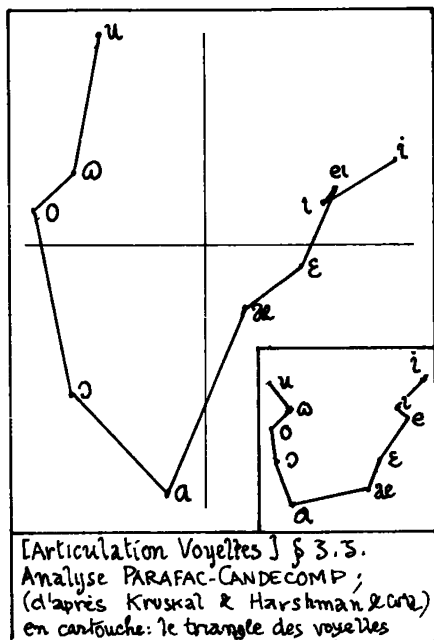
correspondances ; (longueur prohibitive, pour le traitement de données étendues) ; le résultat final dépendant même des valeurs de départ du calcul itératif.

3.2 Critique mathématique : Il faut répéter sans se lasser qu'on ne connaît pas de décomposition *canonique* d'un élément de produit tensoriel de plus de deux espaces euclidiens : à la différence de l'analyse de correspondance binaire, l'analyse des correspondances multiples (ternaires, quaternaires...) ne peut donc fournir une suite bien déterminée de facteurs épuisant de façon ordonnée, toute l'information contenue dans le tableau initial. Cependant, d'une part il est possible d'explorer cette information de façon satisfaisante, en analysant divers tableaux binaires ; d'autre part divers auteurs, dont Harshman, J.B. Kruskal, D. Carroll, etc., ont proposé des méthodes spécifiques pour les tableaux ternaires, quaternaires. Selon nous ces méthodes doivent être jugées en les comparant au modèle de décomposition des interactions en termes successifs (binaires, ternaires etc.) que rappelle par exemple A. Bener (in C.A. D. Vol VII n° 1 pp 25-32 ; 1982). Ce modèle sans prétendre distinguer une base canonique de facteurs fait bien voir ce qu'il y a dans un tableau ternaire, par exemple, en plus de ce que donnent les trois tableaux de marge binaires ($I \times J$, $J \times S$, $S \times I$; dans les notations du présent exemple). *A priori*, il ne semble pas possible de mettre sous la forme d'une somme simple telle que celle proposée par R.H. & J.B.K., à la fois les termes ternaires (de forme quelconque) et les termes binaires : toutefois en jouant sur le centrage on peut dans les termes ternaires incorporer ce que nous appelons des termes binaires, c'est-à-dire des produits de deux fonctions de moyenne nulle sur deux des ensembles, par une fonction constante sur le troisième ensemble. Le modèle PARAFAC et les modèles de cette sorte peuvent pour certains tableaux, donner une représentation suggestive, d'un réel intérêt ; mais d'une part le modèle n'a pas de valeur universelle ; d'autre part même s'il réussit, l'écart entre les données initiales et la représentation qu'on en élabore ne peut être apprécié avec finesse, mais seulement globalement : par exemple si l'on s'arrête à deux facteurs, on n'a pas de troisième facteur ; alors qu'en analyse de correspondance, la représentation du plan 1×2 peut toujours être complétée par des nuances venant du facteur 3. Ces considérations générales peuvent être illustrées sur le présent exemple, dans la mesure où les auteurs ont publié leurs résultats ; et même indépendamment des résultats fournis par PARAFAC, s'il apparaît que les résultats de l'analyse des correspondances ont une structure dont la complexité dépasse le modèle PARAFAC.

3.3 Les résultats publiés par R. Harshman & coll. et J.B. Kruskal :

Ayant décidé de faire un ajustement à deux facteurs, les auteurs obtiennent pour chacun des trois ensembles I, J, S, une représentation plane. Selon J.B. K. les résultats les plus intéressants concernent l'ensemble I des voyelles, et l'ensemble J des positions le long de la langue.

Comme nous, J.B. K. est heureux de voir apparaître le triangle des voyelles.



Sans surestimer l'intérêt de ce résultat ("At first, this may give the impression that the method has merely reproduced the obvious", dit-il) l'auteur remarque qu'il est particulièrement intéressant de voir sortir ce triangle, non de données phonétiques proprement dites (description des sons), mais de données articulatoires (description de l'appareil buccal).

Quoi qu'il en soit de l'appréciation des résultats, nous soulignerons que l'analyse des tableaux ternaires n'est pas véritablement en jeu ici : la figure publiée par J.B. K. ne diffère guère de celle obtenue par nous en analysant le tableau binaire $I \times J$: c'est même d'après cette ressemblance très précise (avec rebroussement entre i et e) que nous avons pu identifier les voyelles de la figure avec les numéros de (1 à 10) qui figurent seuls sur le tableau de données publié avec l'article de J.B. K. (Ultérieurement, cette identification a été confirmée en consultant l'article de Harshman & coll.).

En fait l'analyse ternaire proprement dite commence seulement ici quand on présente et interprète les différences, relativement mineures, existant entre les articulations d'un même son par plusieurs sujets. De ces différences, J.B. K. ne dit rien, mais Harshman & coll. s'y intéressent en phonéticiens avertis.

3.4 Différences interindividuelles selon l'analyse des correspondances

Que le tableau principal soit $I \times J$ (voyelles \times positions) ou $(I \times S) \times J$ ((voyelles \times sujets) \times positions), les axes 1 et 2 sont les mêmes ; et l'axe 3 ne varie guère. Sur le plan 1×2 on voit non seulement un triangle des voyelles moyen I, mais aussi pour chaque sujet s ($s \in S$) son propre triangle ou ensemble $\{(i, s) | i \in I\}$.

Les 5 triangles afférents aux 5 sujets ressemblent au triangle moyen ; toutefois il y a des différences visibles : sans insister sur le rebroussement (i,ei) qui manque aux sujets 2 et 1, il est maximum pour le sujet 4, on notera le décalage global des triangles les uns par rapport aux autres et aussi une inégale ouverture transversale (dans la direction de l'axe 2). Le décalage des triangles peut être mesuré par la position des centres de gravité, c'est-à-dire des points supplémentaires afférents à chaque sujet ; l'ouverture transversale relève plutôt de ce que nous appelons proprement "interaction ternaire" ; en ce sens que pour un sujet tel que 2 est augmentée l'opposition existant entre les voyelles 6 (o) et 10 (u) quant à leurs associations avec les diverses positions le long de la langue. Comme les coordonnées sur les axes factoriels 1 et 2 décrivent presque exactement les profils, on peut chercher des formules avec un coefficient telles que :

$$F_2(is) = F_2(i) + F_2(s) + \text{coeff } F_2(i) \times F_2(s)$$

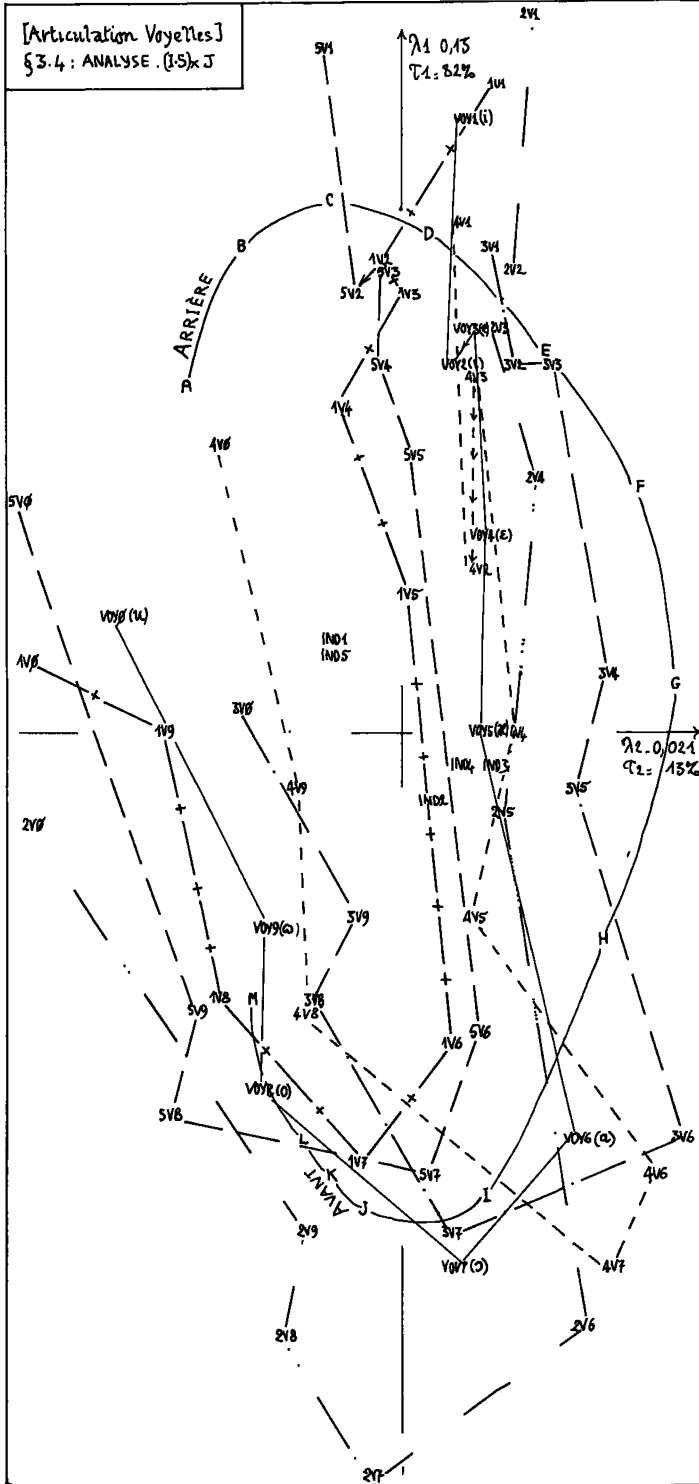
en se souvenant que celles-ci donnent (par la reconstitution des données en fonction des facteurs) des termes en "coeff $F_2(i) F_2(s)$

$G_2(j)$ " pour $k(i,j,s)$. (On trouvera au § 4.2 une expression exacte de $F(is)$ en fonction des interactions binaires et ternaires). Ces formules n'ont pas d'intérêt en elles-mêmes : car sur 5 sujets seulement aucune typologie ne peut être assise ; mais elles montrent combien la finesse des différences interindividuelles rend difficile l'application systématique du modèle PARAFAC.

D'autre part ce modèle comporte un choix du nombre r des facteurs. Certes il est satisfaisant que les auteurs (peut-être guidés par le modèle classique du triangle des voyelles) aient retenu $r = 2$. Mais il importe de souligner que PARAFAC, comme les autres procédures spécifiques d'analyse ternaire que nous avons pu rencontrer, fournit des résultats, qui même pour les premiers facteurs dépendent du choix de r ; et de plus, ne sont même pas déterminés par ce seul choix ; car le calcul en est fait par une itération dont l'issue dépend des valeurs initiales adoptées. Au contraire en analyse des correspondances (ou dans toute autre analyse euclidienne d'un tableau binaire) les facteurs sont déterminés successivement, indépendamment du nombre de ceux qu'on pourra interpréter.

Ici le facteur 3 issu de l'analyse des correspondances n'est pas dénué d'intérêt : il est créé (cf. listage) par une association entre le sujet 2 et une forte ouverture au niveau M (pointe de la langue) ; association qui affecte le plus la voyelle 5 (ae). Un phonéticien pourrait s'interroger sur les particularités anatomiques ; les habitudes articulatoires ; l'influence réelle de l'ouverture à ce niveau pour le son produit etc. .

Harshman & coll., quant à eux ont vu sur leurs deux facteurs, l'originalité du sujet 2. Disposant de nombreuses mesures sur la conformation du tractus vocal ils ont découvert que (2) diffère des autres sujets à la fois quant au rapport (cavité pharyngale/cavité orale) (rapport plus faible chez (2)) ; et quant à la longueur (accrue chez (2)) de la cavité orale. Mais ils ne peuvent offrir de toutes leurs données avec leurs nuances une vue d'ensemble telle que la figure qui illustre le présent §.



4 Appendice : décomposition des interactions et modèles de correspondance ternaire

4.0 Rappel du modèle général : Nous rappelons ce modèle (cf. A. Bener, *op. laud.*) pour le confronter au modèle PARAFAC, et à l'analyse binaire de $I \times J$ avec $(I \times S) \times J$ en supplémentaire. On a le développement :

$$d^{ijs} = 1 + \sum_{\alpha\beta} d^{\alpha\beta 0} \varphi_{\alpha}^i \varphi_{\beta}^j + \sum_{\beta\gamma} d^{0\beta\gamma} \varphi_{\beta}^j \varphi_{\gamma}^s + \sum_{\alpha\gamma} d^{\alpha 0\gamma} \varphi_{\alpha}^i \varphi_{\gamma}^s \\ + \sum_{\alpha\beta\gamma} d^{\alpha\beta\gamma} \varphi_{\alpha}^i \varphi_{\beta}^j \varphi_{\gamma}^s ;$$

dans cette formule d^{IJS} est la fonction densité de la loi ternaire par rapport au produit des marges : $d^{ijs} = f_{ijs} / (f_i f_j f_s)$: on a muni I d'une base de fonctions orthonormées $\{\varphi_{\alpha}^I\}$; la fonction φ_0^I est la constante 1, et les autres fonctions ont moyenne nulle (pour la loi f_I) et sont telles que :

$$\forall \alpha' \alpha'' : \sum_1 f_i \varphi_{\alpha'}^i \varphi_{\alpha''}^i = \delta_{\alpha\alpha''} = \begin{cases} 1 & \text{si } (\alpha = \alpha') \text{ alors } 1, \text{ sinon } 0; \end{cases}$$

et de même sur J et S on a des bases $\{\varphi_{\beta}^J\}$ et $\{\varphi_{\gamma}^S\}$; les $d^{\alpha\beta\gamma}$ sont des coefficients ; dans toute sommation par rapport aux indices α, β, γ , ceux-ci sont supposés non-nuls (la fonction constante étant mise à part dans le développement).

De ce développement ternaire, on peut extraire des développements binaires tels que :

$$d^{ij} = (f_{ij} / (f_i f_j)) = 1 + \sum_{\alpha\beta} d^{\alpha\beta 0} \varphi_{\alpha}^i \varphi_{\beta}^j.$$

Il est permis de supposer que sur deux ensembles, les bases de fonctions sont choisies canoniquement en analysant une correspondance binaire par exemple f_{IJ} . En ce cas les indices α et β se correspondent et on a :

$$d^{ij} = 1 + \sum_{\alpha} \lambda_{\alpha}^{1/2} \varphi_{\alpha}^i \varphi_{\alpha}^j ; \quad d^{\alpha\beta 0} = \lambda_{\alpha}^{1/2} \delta_{\alpha\beta}.$$

Il faut prendre garde que si, e.g., $\text{Card } I > \text{Card } J$, la suite des $(\text{Card } J - 1)$ facteurs extraits de f_{IJ} n'est pas une base complète sur I ; car il y manque $(\text{Card } I - \text{Card } J)$ fonctions : toutefois cette précision a peu d'intérêt dans le présent appendice, où on n'a en vue que des approximations.

4.1 Essai d'ajustement entre PARAFAC et le modèle général : Le modèle PARAFAC peut s'écrire, compte tenu du centrage :

$$d^{ijs} = d^{is} + \sum_r \{a_{ir} b_{jr} c_{sr}\} ;$$

en identifiant avec le modèle général, il vient (compte tenu de l'expression de d^{is}) :

$$\sum \{a_{ir} b_{jr} c_{sr}\} = \\ \sum_{\alpha\beta} d^{\alpha\beta 0} \varphi_{\alpha}^i \varphi_{\beta}^j + \sum_{\beta\gamma} d^{0\beta\gamma} \varphi_{\beta}^j \varphi_{\gamma}^s + \sum_{\alpha\beta\gamma} d^{\alpha\beta\gamma} \varphi_{\alpha}^i \varphi_{\beta}^j \varphi_{\gamma}^s$$

En supposant que les φ_{α}^I et φ_{β}^J sont choisis de façon canonique que par analyse de f_{IJ} , le second membre devient :

$$\sum_{\beta} \lambda_{\beta}^{1/2} \varphi_{\beta}^i \varphi_{\beta}^j + \sum_{\beta\gamma} d^{\alpha\beta\gamma} \varphi_{\beta}^i \varphi_{\gamma}^s + \sum_{\alpha\beta\gamma} d^{\alpha\beta\gamma} \varphi_{\alpha}^i \varphi_{\beta}^j \varphi_{\gamma}^s .$$

On peut tenter d'identifier les facteurs de l'analyse des correspondances (indice β) avec ceux de l'analyse PARAFAC (indice r) en posant :

$$a_{i\beta} = A^{\beta} \varphi_{\beta}^i + a^{\beta} ; b_{j\beta} = B^{\beta} \varphi_{\beta}^j + b^{\beta} ; \\ c_{s\beta} = c^{\beta} + \sum_{\gamma} C^{\beta\gamma} \varphi_{\gamma}^s .$$

L'absence de terme en is dans le second membre oblige à poser $b^{\beta} = 0$ (i.e. les facteurs b de PARAFAC sont centrés) ; l'identification des termes en ij donne $B^{\beta} C^{\beta} A^{\beta} = \lambda_{\beta}^{1/2}$; enfin $a^{\beta} c^{\beta} = 0$ parce que le second membre ne contient pas de termes ne dépendant que de j (tous les termes dépendent de 2 ou 3 variables). On peut donc poser $A^{\beta} = B^{\beta} = 1$; $c^{\beta} = \lambda_{\beta}^{1/2}$ et l'identification possible est :

$$a_{i\beta} = \varphi_{\beta}^i ; b_{j\beta} = \varphi_{\beta}^j ; c_{s\beta} = \lambda_{\beta}^{1/2} + \sum_{\gamma} d^{\beta\beta\gamma} \varphi_{\gamma}^s$$

Avec ce choix des facteurs de PARAFAC, on voit que le 1-er membre (formule approchée de PARAFAC) diffère du 2-ème membre (modèle général exact) en ce que manquent à PARAFAC, les termes en $d^{\alpha\beta\gamma}$ (interactions binaires en J et S) et les termes en $d^{\alpha\beta\gamma}$, pour $\alpha \neq \beta$ (termes d'interactions ternaires dont la partie en IJ , n'est pas un produit de deux facteurs de même rang issus de la correspondance $I \times J$).

Assurément, la recherche itérative des facteurs de PARAFAC peut aboutir à des solutions meilleures que celle suggérée par une identification superficielle ; mais celle-ci nous paraît donner une vue suggestive des limites du modèle PARAFAC quant à l'expression des interactions binaires et ternaires.

4.2 Modèle général ternaire et analyse binaire : Partons de l'analyse du tableau $I \times J$ (en supposant pour simplifier les notations que $\text{Card } J \leq \text{Card } I$; en sorte que les facteurs sur J issus de cette analyse forment un système orthonormé complet). Au tableau $I \times J$ on peut adjoindre les tableaux supplémentaires $S \times J$ et $(I \times S) \times J$; en sorte que pour S et $(I \times S)$ sont définis des facteurs, dont les indices sont ceux des facteurs de base sur J ; nous noterons :

$$F_{\beta}(s) = \lambda_{\beta}^{1/2} \psi_{\beta}^s ; F_{\beta}(is) = \lambda_{\beta}^{1/2} \psi_{\beta}^{is}$$

(cette notation aidant à combiner aux fonctions φ les coordonnées des éléments supplémentaires).

La formule de reconstitution, appliquée aux éléments supplémentaires, donne pour les d^{js} et les d^{ijs} des expressions exactes

qu'on peut identifier à celles du modèle général. Cette identification nous montrera où apparaissent dans les $F_\beta(is)$ les interactions ternaires.

Commençons par la densité $d^{js} = f_{js}/(f_j f_s)$: on a :

$$d^{js} = 1 + \sum_\beta \lambda_\beta^{1/2} \varphi_\beta^j \psi_\beta^s = 1 + \sum_{\beta\gamma} d^{\alpha\beta\gamma} \varphi_\beta^j \varphi_\gamma^s ; \text{ d'où :}$$

$$\psi_\beta^s = \lambda_\beta^{-1/2} \sum_\gamma d^{\alpha\beta\gamma} \varphi_\gamma^s$$

Pour la densité d^{ijs} on a les deux expressions :

$$d^{ijs} = d^{is} (1 + \sum_\beta \lambda_\beta^{1/2} \psi_\beta^{is} \varphi_\beta^j) =$$

$$1 + \sum_\beta \lambda_\beta^{1/2} \varphi_\beta^i \varphi_\beta^j + \sum_{\alpha\gamma} d^{\alpha\beta\gamma} \varphi_\alpha^i \varphi_\gamma^s + \sum_{\beta\gamma} d^{\alpha\beta\gamma} \varphi_\beta^j \varphi_\gamma^s + \sum_{\alpha\beta\gamma} d^{\alpha\beta\gamma} \varphi_\alpha^i \varphi_\beta^j \varphi_\gamma^s .$$

Dans la 2-ème expression (issue du modèle général) on reconnaît d^{is} et d^{js} , d'où la nouvelle identité :

$$d^{is} \sum_\beta \lambda_\beta^{1/2} \psi_\beta^{is} \varphi_\beta^j =$$

$$\sum_\beta \lambda_\beta^{1/2} \varphi_\beta^i \varphi_\beta^j + \sum_\beta \lambda_\beta^{1/2} \varphi_\beta^j \psi_\beta^s + \sum_{\alpha\beta\gamma} d^{\alpha\beta\gamma} \varphi_\alpha^i \varphi_\beta^j \varphi_\gamma^s \lambda_\beta^{-1/2}$$

En faisant le produit scalaire des deux membres par une fonction de base φ_β^j , on voit qu'on a, pour tout β , l'identité entre fonctions de i et s :

$$d^{is} \psi_\beta^{is} = \varphi_\beta^i + \psi_\beta^s + \sum_{\alpha\gamma} d^{\alpha\beta\gamma} \varphi_\alpha^i \varphi_\gamma^s$$

On a donc en première approximation :

$$\psi_\beta^{is} \approx \varphi_\beta^i + \psi_\beta^s ; \text{ (i.e. : } F_\beta(is) = F_\beta(i) + F_\beta(s) \text{) ;}$$

la différence entre les deux membres provenant d'une part des interactions binaires entre i et s ($d^{is} \neq 1$) et d'autre part des interactions ternaires entre i , j et s (termes en $d^{\alpha\beta\gamma}$).