

*Université Libre de Bruxelles
Laboratoire de l'Institut de Phonétique*

L'ÉTUDE DE LA PERCEPTION DES VOYELLES
PAR LA MÉTHODE D'ÉCHELLE MULTIDIMENSIONNELLE
«INDSCAL»

[THE "INDSCAL" MULTIDIMENSIONAL SCALING ANALYSIS IN VOWEL PERCEPTION]

RENAUD BEECKMANS

A recent model of multidimensional scaling, INDSCAL, was used in an attempt to investigate some aspects of the perception of vowels. Two major features make the INDSCAL method original: first, the external structure it provides has to be interpreted without any rotation of the axes; second, configurations can be obtained, that are specific to each subject or to each value of an experimental variable, depending on the input matrices. — In a first experiment, seven French vowels [a], [ɛ], [e], [i], [ɔ], [o], and [u] synthesised at seven different values of fundamental frequency mixed with the same vowels uttered by male and female, were presented for identification to eight subjects. INDSCAL analysis yielded a two-dimensional configuration that was found to be very similar to the physical plane F_1-F_2 . Furthermore, the two perceptual axes match perfectly well with the two first formants. — A second experiment, this time with natural speech, was designed to test the relevance of these results using other experimental variables. One period samples of ten vowels [a], [ɔ], [o], [u], [y], [i], [ɛ], [e], [ø] and [œ] pronounced by ten native French speakers were presented to ten listeners. From the several analyses of these data, it was concluded that F_1 and F_2 remain the perceptually relevant dimensions throughout the whole identification process. Comparison between individual speakers' and listeners' differences revealed normalisation processes which are specific to each perceptual dimension. An explanation in terms of reciprocal interactions perception-production is proposed. — This particular approach of INDSCAL is most attractive in that the dimensions emerge as being meaningful for one particular stage of the process and therefore leads to a functional description of it.

INTRODUCTION

Dans de nombreux domaines, la nécessité de traiter des données expérimentales très nombreuses et de nature fort variée, qualitatives à la limite, a provoqué ces dernières années le développement d'une abondance de méthodes basées sur les échelles multidimensionnelles.

Ces techniques ont comme but — soit de mettre en évidence la véritable structure présente dans une matrice de données, — soit d'obtenir une représentation condensée de cette structure qui soit accessible à l'œil, et ce, par la construction d'un espace où les relations géométriques entre objets sont fonction de la structure. Cette deuxième exigence n'est compatible qu'avec un nombre relativement restreint de dimensions (idéalement inférieur ou égal à 3) et explique

l'abandon des méthodes traditionnelles d'analyses factorielles ou en composantes principales qui aboutissent le plus souvent à un nombre élevé de facteurs ; ce fait est dû, en partie, à la rigidité des hypothèses de linéarité introduites dans ces méthodes. Le souci majeur des analyses multidimensionnelles s'inscrit dès lors dans la recherche d'hypothèses moins contraignantes.

Une des approches les plus fécondes en ce sens, a été introduite par Shepard (1962), qui substitue aux contraintes linéaires des relations monotones. Par leur souplesse même, les méthodes basées sur ce principe sont très générales : elles sont à même de dégager une représentation métrique à partir de données non métriques, ordinales par exemple. Enfin, la formalisation de ce modèle (Kruskal, 1964) qui fournit une estimation rigoureuse de la qualité de l'ajustement, ainsi que la possibilité de traitement sur ordinateur, permet d'aboutir à une représentation optimale des données par itérations successives. Pour l'ensemble de ces méthodes, une difficulté cependant subsiste, identique à celle que posaient les analyses classiques, quant à l'interprétation de la structure obtenue. En effet, cette structure peut fournir par elle-même, des renseignements concernant les interrelations entre les données. Ce qui permettra, par contre, une interprétation globale du phénomène étudié n'est pas la configuration elle-même, mais bien les dimensions qui la sous-tendent. Or, l'orientation de ces dimensions est complètement arbitraire par rapport à la structure ; ceci provient de ce que les distances inter-points sont fonction des données de départ mais restent strictement invariantes par rapport à une rotation rigide de l'ensemble de la configuration. Le choix d'une rotation "convenable" est tautologique dans la mesure où il repose sur des critères étrangers aux données elles-mêmes : le plus souvent, on essaie de retrouver des dimensions connues ou pertinentes par rapport aux dimensions physiques des objets.

Le problème est particulièrement important en psychologie, notamment dans le domaine de la perception. Dans ce cas, en effet, les stimuli employés n'interviennent qu'en tant que moyen d'étudier les dimensions fonctionnelles du système perceptif et dès lors, les relations entre les différents stimuli n'ont pas de valeur intrinsèque. Par contre, une définition des différents axes qui sous-tendent l'espace perceptif, fournit une première description des différents processus sous-jacents et ces axes, ainsi que leur orientation, sont nécessaires, en ce sens, à l'élaboration d'une théorie prédictive.

Pour les raisons que nous venons d'invoquer, l'apport des méthodes d'échelles multidimensionnelles classiques à l'étude de la perception des voyelles s'est attaché principalement à dénombrer les dimensions perceptives pertinentes, et à déterminer quelles sont les variables physiques du signal qui corrélerent ces dimensions. Dans la plupart des études (Hanson, 1967 ; Pols, 1969), trois dimensions sont suffisantes pour rendre compte des relations existant au sein des systèmes vocaux considérés. Parmi ces trois dimensions généralement obtenues, deux peuvent être mises en rapport avec les valeurs des deux premiers

formants des voyelles. Rien cependant à partir de ces résultats ne permet d'affirmer l'existence d'une analyse perceptive qui dégagerait les deux valeurs formantiques F_1 et F_2 comme paramètres de décision, puisque la même configuration globale serait obtenue pour n'importe quel processus perceptif utilisant de quelconques combinaisons linéaires de F_1 et F_2 . Or, la spécification des voyelles à l'aide des valeurs formantiques ne constitue qu'un moyen économique de décrire les systèmes vocaliques et qui ne recouvre à priori qu'une réalité acoustique. L'étude de la production orale montre qu'il n'y a pas de correspondance univoque entre les paramètres articulatoires et les valeurs formantiques. Par contre, certains paramètres constituent des combinaisons linéaires des deux premiers formants. Le Graph. 1 emprunté

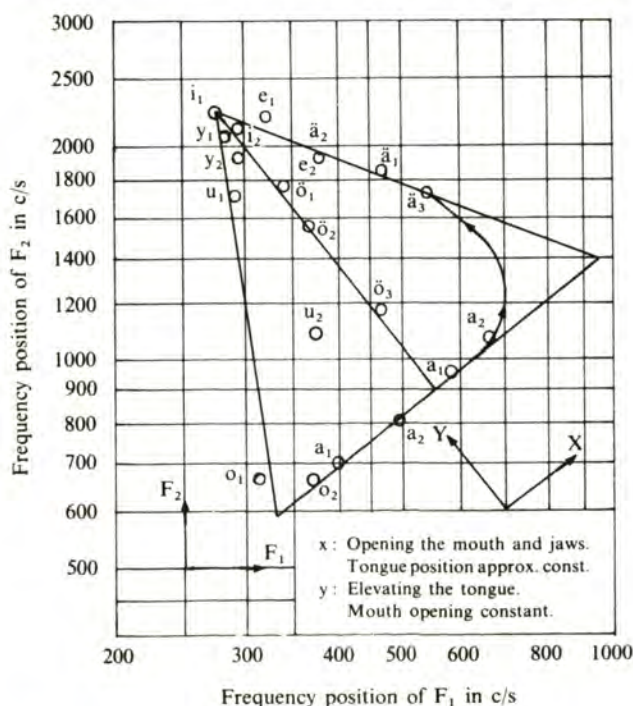


FIG. 1. REPRÉSENTATION DES VOWELLES SUÉDOISES DANS LE PLAN F_1 - F_2 — FIRST FORMANT VERSUS SECOND FORMANT REPRESENTATION OF SWEDISH VOWELS

à Fant (1973) montre que la configuration du système vocalique suédois en fonction de F_1 et F_2 d'une part, de deux paramètres articulatoires d'autre part (ouverture de la bouche et des mâchoires-élévation de la langue) est identique. Mais ce fait n'implique aucunement une correspondance univoque entre les deux types de paramètres comme l'indique clairement l'orientation différente des deux systèmes d'axes.

La relation entre les paramètres aux niveaux acoustique et articulatoire est réalisable du fait que les paramètres sont connus a priori. Par contre, les paramètres intervenant au niveau perceptif sont difficiles à postuler : les oppositions phonologiques classiques sont compatibles avec les structures obtenues, mais la diversité des systèmes phonologiques ne permet pas de préciser les orientations des axes perceptifs en fonction de F_1 et F_2 .

Une possibilité de résoudre ce problème de façon fort directe nous a été offerte par une méthode récente d'analyse multidimensionnelle, INDSCAL, (Carroll et Chang, 1970) qui a le mérite de déterminer les axes sous-tendant la structure de données, d'une manière univoque : la signification de ces dimensions "définie", par rapport au processus étudié, est liée au principe de la méthode.

L'idée originale des auteurs est de postuler l'existence de dimensions communes à tous les sujets. Seule l'importance relative des différentes dimensions varie de sujet à sujet. Si ces contraintes sont respectées, la méthode fournit d'une part un espace général (espace stimuli) sous-tendu par un système d'axes qui représentent les dimensions "définies" communes à l'ensemble des sujets, d'autre part, une matrice de pondérations (espace sujets) qui permet de fixer l'importance relative des différentes dimensions pour chacun des sujets. La possibilité de poids nuls permet de rendre compte des cas où seules certaines dimensions sont communes et à la limite, d'un "super espace" où aucune dimension n'est commune. Si x_{jt} et x_{kt} sont les coordonnées des points j et k pour la dimension t de l'espace stimuli à n dimensions, la distance entre ces 2 points pour le sujet i sera :

$$d_{jk}^{(i)} = \left[\sum_{t=1}^r \omega_{it} (x_{jt} - x_{kt})^2 \right]^{1/2}$$

où ω_{it} est le poids sur la dimension t pour le sujet i . Cette distance peut être considérée comme une distance euclidienne normale dans l'espace spécifique au sujet i , où les coordonnées des points j et k sont alors :

$$y_{jt}^{(i)} = (\omega_{it})^{1/2} x_{jt} \quad \text{et} \quad y_{kt}^{(i)} = (\omega_{it})^{1/2} x_{kt}$$

Les espaces spécifiques à chacun des sujets diffèrent donc de l'espace stimuli uniquement par le fait qu'ils sont différentiellement contractés ou dilatés dans les directions correspondant aux axes.

Lorsque les dimensions de l'espace stimuli sont normalisées de manière appropriée ($N(0, 1)$ pour chaque dimension) et les poids réajustés en conséquence, le carré du poids d'un sujet pour une dimension particulière est inférieur ou égal à la proportion de variance pour cette dimension suivant que les dimensions sont corrélées ou non. Une méthode itérative d'ajustement au sens des moindres carrés

permet de déterminer les coordonnées des stimuli et les poids par dimension et par sujet qui rendent compte du maximum de variance totale; cette variance expliquée est calculable pour chaque sujet et aussi pour l'ensemble des sujets.

	subject 1				subject 2				subject 3			
	A	B	C	D	A	B	C	D	A	B	C	D
A	0	1	$\sqrt{2}$	1	0	1/2	$\sqrt{5}/2$	2	0	2	$\sqrt{5}/2$	1/2
B	1	0	1	$\sqrt{2}$	1/2	0	2	$\sqrt{5}/2$	2	0	1/2	$\sqrt{5}/2$
C	$\sqrt{2}$	1	0	1	$\sqrt{5}/2$	2	0	1/2	$\sqrt{5}/2$	1/2	0	2
D	1	$\sqrt{2}$	1	0	2	$\sqrt{5}/2$	1/2	0	1/2	$\sqrt{5}/2$	2	0

TAB. 1. DISTANCES INTER-STIMULI POUR TROIS SUJETS THÉORIQUES — DISTANCES BETWEEN STIMULI FOR THREE THEORETICAL SUBJECTS

Un exemple théorique permettra de mieux saisir le principe général de la méthode. Supposons connues les distances séparant quatre stimuli A, B, C, D, pour chacun des trois sujets I, II, III, (Tab. 1). L'ensemble des structures à deux dimensions (Fig. 2a) peut être réduite à une configuration globale et une matrice de poids par dimension et par sujet, cette matrice étant représentable graphiquement. L'application des poids pour un sujet transforme automatiquement l'espace-stimuli en l'espace spécifique à ce sujet (Fig. 2b).

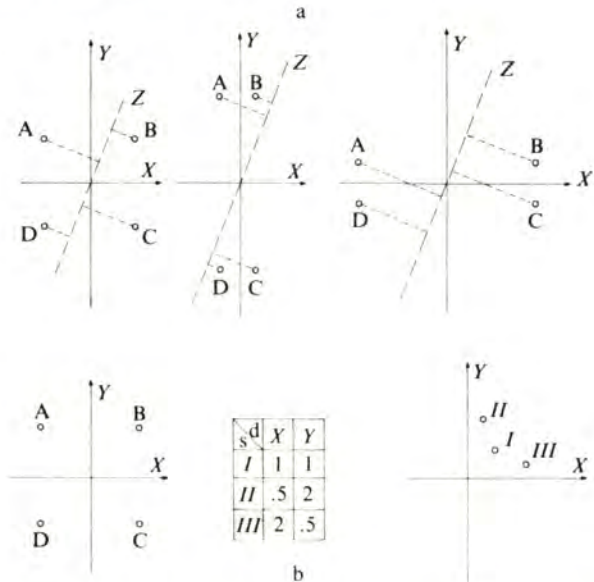


FIG. 2. ESPACES PERCEPTIFS POUR TROIS SUJETS THÉORIQUES (a) ET LA REPRÉSENTATION "INDSCAL" CORRESPONDANTE (b) — PERCEPTUAL CONFIGURATIONS OF THREE THEORETICAL SUBJECTS (a) AND THE CORRESPONDING "INDSCAL" REPRESENTATION (b)

On peut également remarquer comment s'opère l'unicité des axes, étant donné la contrainte des poids communs pour tous les points. En effet, si on considère un axe quelconque z , on voit que les coordonnées des points sur cet axe ne sont pas en relation linéaire comme c'est le cas pour x et y (dans l'exemple choisi, l'ordre même n'est pas conservé).

Ce modèle est également applicable lorsque l'on décompose les données en une partition autre que les sujets. Il est possible, par exemple de présenter un même matériel sous différentes conditions expérimentales. Dans l'hypothèse où l'influence de ces différentes conditions sur l'espace perceptif répond aux contraintes du modèle (influence limitée aux contributions relatives de chaque dimension), l'application d'INDSCAL permet de préciser le rôle de ces variables par rapport au traitement perceptif. Cette extension pose cependant un problème : dans la mesure où un même ensemble de données peut être décomposé en différentes partitions, il faut se demander si l'analyse fournira pour chaque cas des axes identiques ou non. Cette question sera discutée lors de la seconde expérience.

EXPÉRIENCE I

Nous avons à notre disposition une série de voyelles synthétiques produites à l'aide de l'analogie électrique du conduit vocal de l'E.N.S.E.R. de Grenoble. Ce matériel convenait parfaitement à un premier essai d'application d'INDSCAL du fait que la connaissance précise des valeurs formantiques éliminait les problèmes que pose l'analyse du signal. En outre, les voyelles avaient été produites à enveloppe spectrale constante pour différentes valeurs de fréquence fondamentale. C'était donc là l'occasion de tester les possibilités du modèle appliqué non aux différences individuelles mais bien aux différences de fondamentale.

Pour que les contraintes du modèle fussent respectées, il fallait postuler une normalisation perceptive linéaire par rapport à F_0 . Les études concernant la normalisation des voyelles dans l'espace F_1 , F_2 en fonction de la longueur du conduit vocal (Fant, 1973) montrent qu'une simple relation linéaire fournit une bonne correspondance entre les valeurs formantiques de voyelles homologues pour les voix d'hommes, de femmes et d'enfants. Dès lors, il n'est pas impossible qu'une relation analogue intervienne au niveau perceptif.

MÉTHODE EXPÉRIMENTALE

Stimuli : sept voyelles synthétiques ont été retenues : [i], [e], [ɛ], [a], [u], [o] et [ɔ]. Chacune de ces voyelles a été produite à sept fréquences fondamentales différentes : 80 - 100 - 120 - 150 - 200 - 250 - 350 Hz. L'enveloppe spectrale définie par les valeurs de F_1 et F_2 (Fig. 6) restait identique pour toutes les valeurs de F_0 . Pour éviter les phénomènes de réduction temporelle, la durée de chaque voyelle a été normalisée à 500 ms avec des fronts d'ouverture et de fermeture exponentiels.

A ces 49 (7×7) stimuli artificiels ainsi obtenus, furent ajoutées 14 voyelles naturelles prononcées par une voix d'homme ($F_0 = 100$ Hz) et une voix de femme ($F_0 = 250$ Hz). Les deux premiers formants de chaque voyelle synthétique étaient ceux de la voyelle correspondante émise par la voix d'homme. L'ensemble des stimuli a été copié en ordre aléatoire sur bande magnétique; l'intervalle séparant chaque voyelle était de 8 secondes. Cinq bandes de 63 (49 + 14) stimuli furent ainsi préparées.

Sujets : huit sujets, âgés en moyenne de 21 ans, ont identifié les voyelles qui leur furent présentées binauralement (écouteur Sharp HA 10). Ils entendirent une fois quatre premières bandes choisies de manière et en ordre aléatoires, et deux fois une cinquième, la séquence étant 5 - 1 - 2 - 3 - 4 - 5. Ils furent ainsi soumis à l'audition de 378 (63×6) stimuli dont 6 pour chaque voyelle et chaque fondamentale. Les réponses écrites, obligatoires pour chaque stimulus, étaient ventilées sur les 7 voyelles. Aucun retour en arrière n'était autorisé. Avant l'expérience, les sujets avaient subi un entraînement leur permettant de maîtriser la transcription des voyelles en système I.P.A.

RÉSULTATS

Résultats globaux : le Tab. 2 montre les confusions, pour chaque fréquence fondamentale et pour les deux voix naturelles, cumulées sur l'ensemble des sujets. Le Tab. 3 montre pour ces mêmes conditions, le pourcentage d'identifications correctes spécifique à chacune des 7 voyelles. En première analyse, le pourcentage brut de réponses correctes pour l'ensemble des stimuli décroît avec l'augmentation de la fréquence fondamentale. L'hypothèse d'une relation simple, fondamentale-enveloppe spectrale, explique ce résultat : la variation d'un terme seulement (fréquence fondamentale) dégrade nécessairement l'intelligibilité. Cette relation semble en outre relativement stricte. Si on divise le groupe des fréquences en deux sous-groupes correspondant respectivement aux zones des fréquences fondamentales d'homme et de femme, on remarque que le pourcentage de réponses correctes pour les fréquences 80 - 100 et 120 Hz est du même ordre que celui pour la voix d'homme (74,3% > < 77,7%) alors que le pourcentage de réponses correctes pour les fréquences 200, 250 et 350 Hz est très nettement inférieur à celui pour la voix de femme (46,0% > < 67,9%). Du fait que le taux de reconnaissance correcte pour la voix d'homme était relativement faible, il nous a semblé prudent de confirmer le résultat global en tenant compte de l'orientation des erreurs; nous avons donc calculé à partir des matrices de confusion pour chaque condition, le coefficient de corrélation entre les stimuli présentés aux sujets et les réponses correspondantes. Stimuli et réponses étant des variables qualitatives, il fut nécessaire d'utiliser une méthode d'estimation basée sur l'information transmise (Pearson, 1966) et d'appliquer aux résultats les tests de signification

F = 80 Hz.								F = 100 Hz.								F = 120 Hz.							
	a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u
a	44	2	0	0	5	0	1	a	46	0	0	0	3	0	1	a	44	2	2	0	1	0	
ε	1	35	1	1	0	0	0	ε	1	38	3	0	0	0	0	ε	1	36	2	1	0	0	
e	0	0	33	4	0	2	2	e	0	2	32	9	1	5	3	e	0	5	38	6	0	5	
i	0	0	1	26	0	0	3	i	0	0	3	32	0	0	9	i	0	0	3	33	0	1	2
ɔ	2	9	7	4	40	2	0	ɔ	0	3	4	2	39	4	3	ɔ	3	1	1	1	43	6	
o	1	1	4	2	3	42	6	o	0	2	5	0	5	34	6	o	0	3	2	0	4	26	
u	0	1	2	11	0	2	36	u	0	2	1	5	0	5	25	u	0	1	0	7	0	10	2
F = 150 Hz.								F = 200 Hz.								F = 250 Hz.							
	a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u
a	39	1	0	0	1	0	0	a	34	2	0	0	1	1	0	a	42	3	0	1	0	0	
ε	5	30	6	1	7	1	0	ε	1	36	1	3	11	3	0	ε	3	12	0	0	4	2	
e	1	8	29	6	0	2	3	e	6	5	44	5	9	11	9	e	1	21	10	1	10	6	
i	0	0	9	40	0	3	7	i	0	0	1	36	2	0	24	i	0	1	34	40	4	0	
ɔ	1	6	2	0	33	4	1	ɔ	4	3	0	0	15	4	1	ɔ	0	1	0	0	4	2	
o	2	3	0	0	7	14	7	o	2	2	2	2	10	26	3	o	1	6	0	0	23	4	
u	0	0	2	1	0	24	30	u	1	0	0	2	0	3	11	u	1	4	6	6	2	34	3
F = 350 Hz.								male								female							
	a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u		a	ε	e	i	ɔ	o	u
a	31	5	0	0	3	1	3	a	48	0	0	0	4	1	0	a	47	2	0	0	34	2	
ε	4	8	0	0	8	1	0	ε	0	40	3	0	1	0	2	ε	0	27	4	2	2	0	
e	1	20	3	8	8	3	3	e	0	2	42	11	0	4	0	e	1	6	36	2	3	6	
i	4	7	37	40	15	6	19	i	0	0	0	33	0	0	0	i	0	1	0	44	1	0	
ɔ	3	0	0	0	1	0	0	ɔ	0	6	1	0	35	11	5	ɔ	0	8	0	0	4	2	
o	5	1	1	0	4	4	0	o	0	0	2	0	8	32	10	o	0	4	8	0	4	38	1
u	0	7	7	0	9	33	19	u	0	0	0	4	0	0	31	u	0	0	0	0	0	0	3

TAB. 2. CONFUSIONS INTER-VOYELLES À CHAQUE FRÉQUENCE FONDAMENTALE ET POUR LES DEUX VOIX NATURELLES — CONFUSIONS BETWEEN VOWELS CORRESPONDING TO EACH FUNDAMENTAL FREQUENCY AND TO THE NATURAL VOICES

	80 Hz.	100 Hz.	120 Hz.	150 Hz.	200 Hz.	250 Hz.	350 Hz.	male	female
a	92	98	92	81	71	87	65	100	98
ε	73	81	75	63	75	25	16,5	83	56
e	69	67	79	60	92	21	6,2	87	75
i	54	67	69	83	75	83	83	69	92
ɔ	83	81	90	69	31	8,5	21	73	8,3
o	87	71	54	29	54	8,3	9,3	67	78
u	75	53	52	63	23	77	40	65	67

TAB. 3. POURCENTAGE DE RÉPONSES CORRECTES POUR CHACUNE DES 7 VOYELLES — PERCENTAGE OF CORRECT REPONSES CORRESPONDING TO EACH OF THE SEVEN VOWELS

proposés par le même auteur. On s'est servi, comme référence, des résultats obtenus lors de la présentation des voyelles naturelles. Le Graph. 3a montre que l'évolution du coefficient de corrélation en fonction de la fréquence fondamentale, est de la même forme que le pourcentage de réponses correctes. Il y a donc clairement, au niveau perceptif, une normalisation des valeurs formantiques en fonction de la fréquence fondamentale.

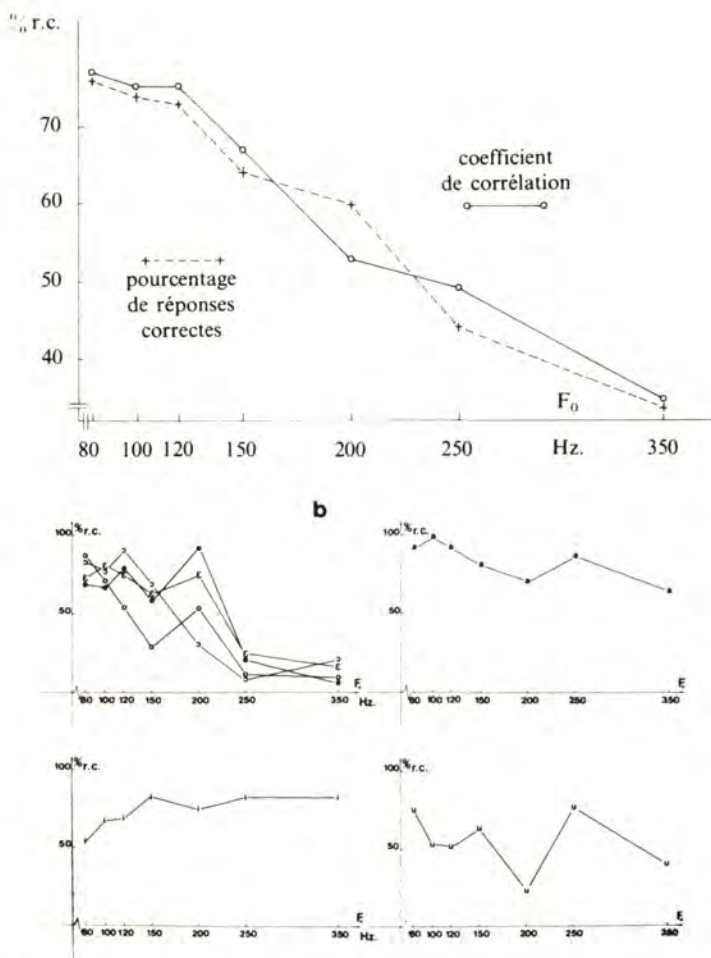


FIG. 3. POURCENTAGE D'IDENTIFICATION EN FONCTION DE LA FRÉQUENCE FONDAMENTALE POUR CHAQUE VOYELLE (b) ET COMPARAISON AVEC LE COEFFICIENT DE CORRÉLATION POUR L'ENSEMBLE DES VOYELLES (a) — IDENTIFICATION SCORES VERSUS FUNDAMENTAL FREQUENCY FOR EACH VOWEL (b) AND THE COMPARISON WITH THE CORRELATION COEFFICIENT (a)

Évolution spécifique des voyelles : si la détérioration de l'intelligibilité en moyenne, ne fait aucun doute, par contre l'évolution du taux de reconnaissance en fonction des différences de fondamentale n'est pas identique pour chaque voyelle (Fig. 3b). Le phénomène global ne se vérifie que pour le groupe {[e], [ɛ], [o], [ɔ]} avec cependant une valeur élevée à 200 Hz; pour le [a], le taux de reconnaissance tend à rester constant; pour le [i], le pourcentage croît avec la fondamentale, ce qui est à l'opposé de la tendance moyenne; pour le [u], enfin, aucune tendance nette apparaît. Il semble donc que l'influence de la fréquence fondamentale, pour cette expérience au moins, soit plus complexe que la simple normalisation linéaire que nous avons postulée. Nous espérons par conséquent, que la méthode INDSCAL nous fournirait, outre les dimensions perceptives véritables, une possibilité d'interpréter de façon plus détaillée le rôle de la fréquence fondamentale.

TRAITEMENT PAR "INDSCAL"

Méthode : l'application du modèle nécessite une série de transformations préalables des résultats bruts (matrices de confusion) pour aboutir aux produits scalaires des vecteurs reliant l'origine aux points-stimuli : confusions (c_{jk}) distances relatives (δ_{jk}) distances absolues (d_{jk}) distances absolues symétriques (D_{jk}) produits scalaires (p_{jk}) produits scalaires normalisés (p_{jk}).

Les nombres c_{jk} figurant dans les matrices de confusion représentent le nombre de fois que le stimulus j a été identifié comme le stimulus k . Plus le nombre de confusions est élevé, plus proches sont les deux stimuli. Nous avons donc considéré comme distances relatives :

$$\delta_{jk} = -c_{jk}$$

La conversion des distances relatives en distances absolues pose le problème de la constante additive :

$$d_{jk} = \delta_{jk} + a$$

Les auteurs proposent l'estimation de cette constante par la méthode de Torgerson (1958); dans notre expérience, une valeur de a était fournie très directement par le nombre de présentations (changé de signe) qui était commun pour tous les stimuli et toutes les conditions ($a = -48$). De cette façon, le nombre maximal possible de confusions correspondait à une distance nulle. La transformation était donc :

$$d_{jk} = \delta_{jk} - 48$$

Les distances absolues ainsi obtenues furent symétrisées par moyen-nisation arithmétique :

$$D_{jk} = D_{kj} = \frac{d_{jk} + d_{kj}}{2}$$

Tenant compte de la nature des stimuli (cf. expérience II), cette méthode de transformation, relativement rudimentaire, nous est apparue suffisante (ce que les résultats confirment) pour dégager les axes principaux de l'espace perceptif. Il faut souligner cependant que l'obtention de résultats plus fins (entre autres le nombre exact de dimensions) nécessiterait une procédure plus sophistiquée. A partir de ces distances, les produits scalaires peuvent être calculés par la formule de Torgerson (1968) :

$$p = \frac{1}{2} \left[\sum_j D_{jk} + \sum_k D_{jk}^2 - \sum_j \sum_k D_{jk}^2 - D_{jk}^2 \right] \quad <1>$$

Enfin, pour éviter que les variances propres à chaque condition (ici fréquence) ne conditionnent leur contribution à l'espace général, les matrices des produits scalaires ont été normalisées (0, 1); leur moyenne étant nulle de par l'application de la formule <1>, les produits scalaires ont été divisés par leur variance sur une condition, ce qui donne

$$b_{jk} = \frac{p_{jk}}{\sigma_p}$$

Les auteurs ont montré qu'en adoptant la formulation de distance spécifique à chaque sujet, on a la relation :

$$b_{jk}^{(i)} \simeq w_{it} x_{jt} x_{kt} \quad <2>$$

où le signe $\simeq$ implique une solution au sens des moindres carrés pour les paramètres du membre de droite, qui peut être déterminée par une méthode itérative.

La détermination du score de variance expliquée que réalise le modèle est possible par normalisation de la solution obtenue : l'équation <2> montre qu'il est possible de multiplier les x_t par une constante arbitraire non nulle et de compenser en multipliant les w_t par l'inverse du carré de cette constante; en choisissant cette constante de façon à normaliser (0, 1) les projections des points-stimuli sur les différents axes, le carré de la distance euclidienne d'un point condition à l'origine, peut être interprété comme la variance totale expliquée des produits scalaires pour cette condition.

L'ensemble de ces traitements a été programmé en langage Fortran sur le CDC 6400 du Centre de Calcul de l'U.L.B.

Résultats : les données préalablement transformées en produits scalaires normalisés furent traités de deux manières : - 7 conditions : les stimuli synthétiques uniquement. - 9 conditions : les stimuli synthétiques plus les stimuli naturels. Pour chacun de ces traitements, le nombre de dimensions testées fut 1, 2, 3, 4, 5 (5 étant le maximum pour 7 conditions car le modèle ne détermine univoquement les axes que pour

au plus $n-2$ dimensions, n étant le nombre de conditions) (Kruskal, 1976). Le Graph. 4 montre pour 7 et 9 conditions, le pourcentage de variance globale expliquée en fonction du nombre de dimensions.

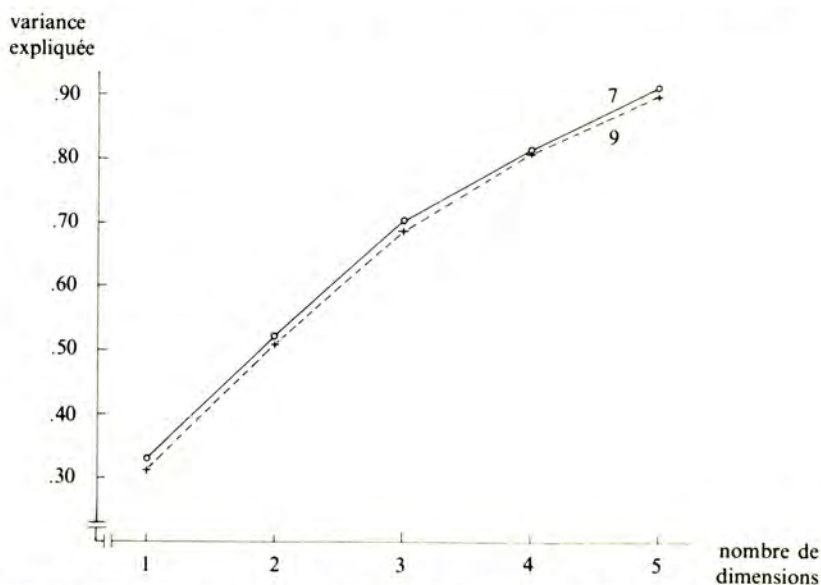


FIG. 4. AJUSTEMENT OBTENU POUR LES STIMULI SYNTHÉTIQUES (7) ET POUR L'ENSEMBLE DES STIMULI (9) — THE FIT CURVE OBTAINED WITH SYNTHETIC STIMULI (7) AND WITH BOTH SYNTHETIC AND NATURAL STIMULI (9)

Les deux faits importants sont : – la faible différence entre les variances expliquées pour 7 et 9 conditions montre que l'espace psychologique des stimuli naturels est compatible (au sens du modèle) avec l'espace psychologique des stimuli synthétiques; – le coude de la courbe, quoique léger, montre que les trois premières dimensions rendent compte d'une grande part de l'information contenue dans les matrices de confusions.

DISCUSSION

Les Graph. 5a et 5b montrent l'espace stimuli et l'espace des poids spécifiques à chaque dimension et à chaque valeur de fréquence fondamentale. La structure obtenue pour les deux premières dimensions est fort semblable à celles mises en évidence par d'autres travaux sur les espaces multidimensionnels (Hanson, 1967; Pols, 1969). Ces auteurs ont montré que cette structure perceptive était en accord avec l'espace physique déterminé par les valeurs des deux premiers formants.

Notre souci majeur était de vérifier si ces valeurs des deux premiers formants étaient directement utilisées comme deux critères indépendants

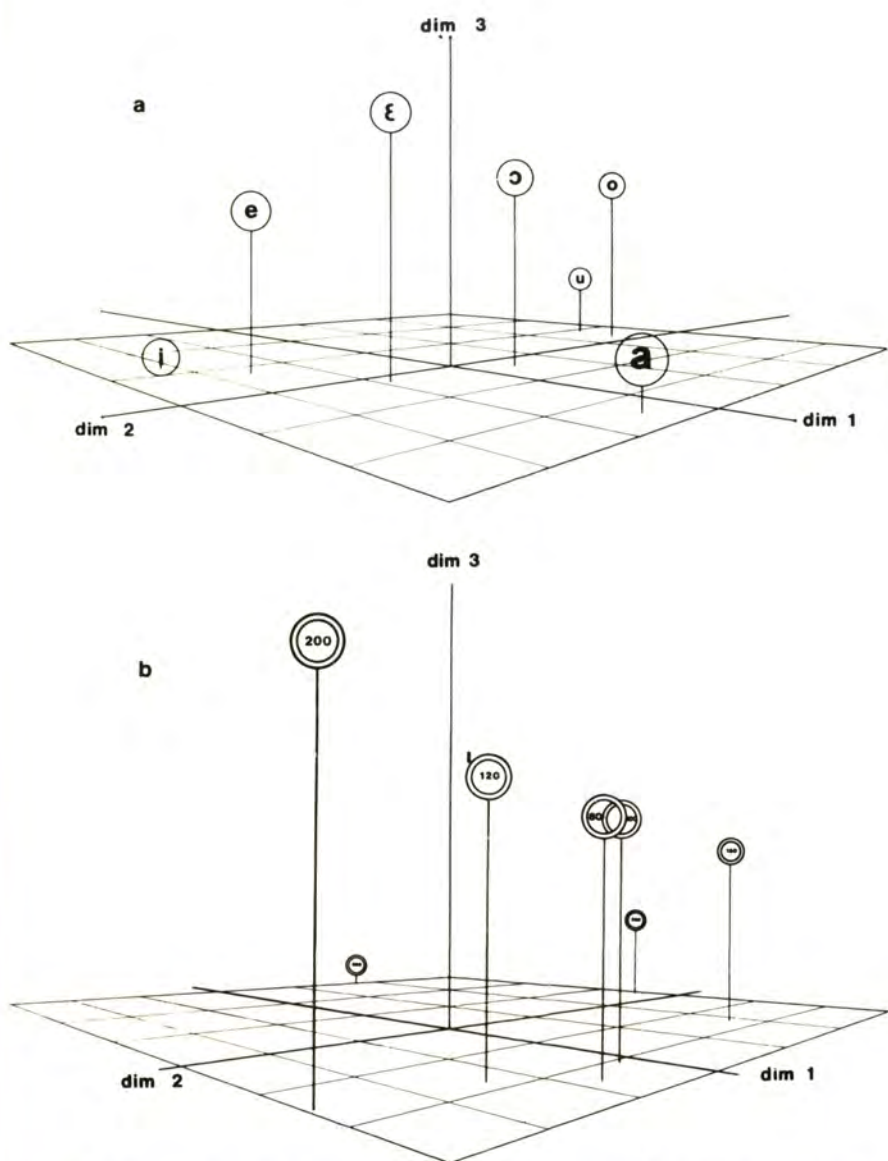


FIG. 5. ESPACES PERCEPTIFS TRIDIMENSIONNELS POUR LES VOYELLES (a) ET LES DIFFÉRENTES FRÉQUENCES FONDAMENTALES (b) — THREE-DIMENSIONAL CONFIGURATIONS FOR THE PERCEPTUAL VOWEL SPACE (a) AND FOR THE DIFFERENT FUNDAMENTAL FREQUENCIES (b)

par le système perceptif. Dans cette hypothèse, les directions des axes DIM 1 et DIM 2 fournis par l'analyse devaient correspondre aux directions des axes physiques F_1 et F_2 . Par contre, des directions différentes entre dimensions acoustiques et perceptives devraient apparaître dans l'hypothèse où les véritables critères perceptifs seraient des combinaisons linéaires de F_1 et F_2 . Nous avons donc testé la correspondance entre les deux espaces par un ajustement au sens des moindres carrés avec comme seule contrainte DIM 1 perpendiculaire à DIM 2. La Fig. 6

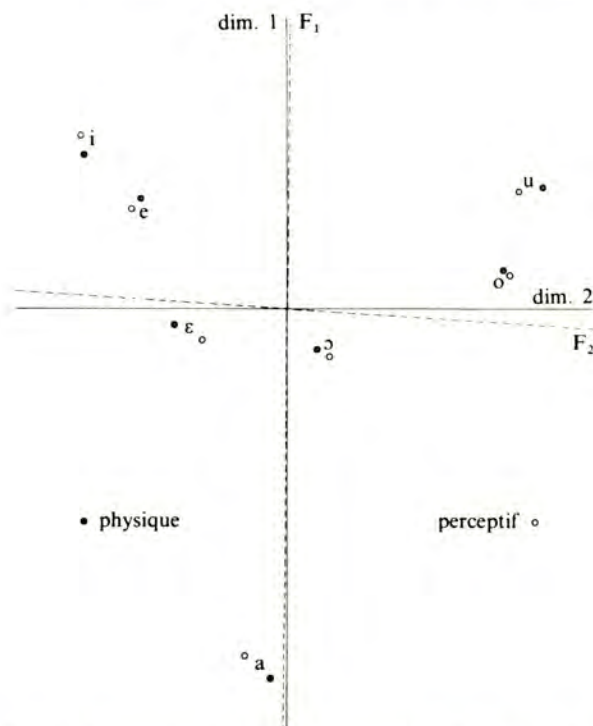


FIG. 6. MISE EN CORRESPONDANCE DES ESPACES PERCEPTIF ET PHYSIQUE F_1 - F_2 — RESULT OF MATCHING THE PERCEPTUAL SPACE TO THE PHYSICAL F_1 - F_2 PLANE

montre que l'ajustement obtenu entre les deux espaces est excellent et ce, pour toutes les voyelles. Le résultat essentiel est que l'ajustement optimal se réalise lorsque les dimensions physiques et perceptives sont pratiquement confondues. On peut en conclure que, pour cette expérience au moins, les deux formants F_1 et F_2 sont effectivement localisés par le système perceptif et utilisés comme deux critères indépendants pour l'identification des voyelles.

Les deux premières dimensions rendent parfaitement compte de l'ensemble de l'information contenue dans les stimuli puisque seuls

F_1 et F_2 différenciaient les voyelles présentées; il est donc étonnant de trouver une troisième dimension qui explique un pourcentage de variance aussi élevé (18%). A notre connaissance, aucune hypothèse valable n'a été avancée jusqu'à présent, pour rendre compte de cette troisième dimension qui apparaît dans la plupart des études portant sur les espaces perceptifs des voyelles. Ceci est dû vraisemblablement au fait qu'on a toujours tenté d'expliquer cette dimension supplémentaire en termes de variables physiques; dans notre cas, cette possibilité est exclue et la seule explication possible doit être trouvée au niveau de la méthode de traitement des données: la transformation des données expérimentales (confusion inter-stimuli) en données utilisables par le modèle (distances inter-stimuli) introduit une distorsion de l'information dont la troisième dimension n'est que la conséquence. La symétrisation des données de confusions supprime une partie de l'information; ensuite la transformation des confusions symétrisées en distances par une fonction simple¹ introduit un biais entre stimuli "marginaux", et stimuli "centraux". Supposons quatre stimuli équidistants dans un espace perceptif unidimensionnel: les risques de confusions des deux stimuli "centraux" seront supérieurs à ceux des stimuli "marginaux", ce dont la conversion employée ne tient pas compte (Fig. 7). Il en

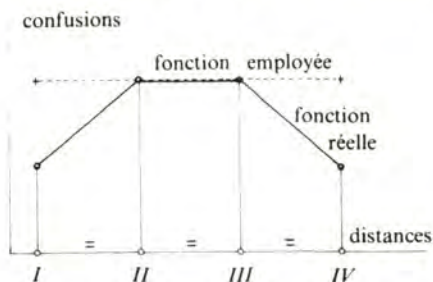


FIG. 7. RELATION ENTRE CONFUSIONS ET DISTANCES PAR RAPPORT À LA POSITION (CENTRALE OU NON) DU STIMULUS — THE RELATION BETWEEN CONFUSIONS AND DISTANCES WITH RESPECT TO THE POSITION (CENTRAL OR NOT) OF THE STIMULUS

résulte une augmentation des distances des "marginaux". L'examen de la figure 5a, montre que l'apport de la troisième dimension est justement d'augmenter les distances des trois voyelles marginales [i], [u], [a]; on peut en conclure que seules les deux premières dimensions sont pertinentes pour l'identification des voyelles. Cette explication peut être étendue aux autres résultats d'analyses multidimensionnelles, notamment par le fait que la voyelle [ø], en tant que voyelle fort "centrale", est souvent responsable de la troisième dimension artéfactuelle.

¹ Il est difficile de proposer un autre type de fonction qui n'introduise pas de biais dans la transformation des données, puisqu'il faudrait quantifier le caractère "isolé" de chaque stimulus ce qui n'apparaît qu'avec la configuration finale.

En ce qui concerne le rôle de la fondamentale mis en évidence par la décroissance de l'intelligibilité globale, la qualité de l'ajustement du modèle aux données implique une normalisation linéaire des deux dimensions F_1 et F_2 en fonction de F_0 . Ce type de normalisation est compatible avec les résultats obtenus pour d'autres langues, notamment en japonais (Fujisaki et Kawashima, 1968). Il reste à expliquer les différences observées entre les voyelles. Ces différences se traduisent à l'analyse par des fluctuations des pondérations, différentes pour chaque dimension. La Fig. 8 montre l'évolution des poids relatifs

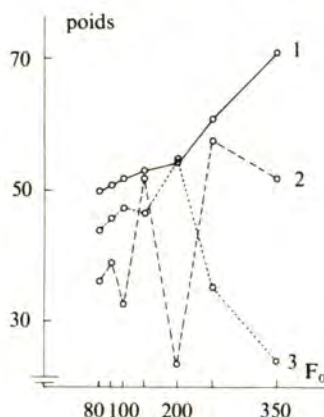


FIG. 8. IMPORTANCES RELATIVES DE CHAQUE DIMENSION PAR RAPPORT À LA FRÉQUENCE FONDAMENTALE — THE SALIENCES CORRESPONDING TO EACH DIMENSION VERSUS THE FUNDAMENTAL FREQUENCY

à chaque dimension en fonction de la fréquence fondamentale. Si l'importance de la première dimension F_1 croît régulièrement avec F_0 , par contre F_2 ne montre pas d'évolution constante. Ces deux courbes s'expliquent parfaitement si l'on tient compte du facteur de définition spectrale : la proximité entre un formant et l'harmonique la plus proche affecte l'échelle perceptive correspondant à ce formant. Les risques de confusions sont donc plus grands quand la localisation du formant est moins précise.

Ce facteur joue principalement pour des formants fort bas, en particulier les premiers formants de [u] et [o]. Les valeurs de F_0 pour lesquelles le poids relatif de la dimension F_2 est élevé, correspondent aux cas où la définition spectrale de F_1 est médiocre ; à l'inverse, les poids faibles pour F_2 correspondent à une excellente définition de F_1 (voir Fig. 9). L'effet réciproque des deux dimensions est dû à la normalisation introduite par le modèle ; l'égalisation des variances opérée induit à partir des fluctuations de poids d'une dimension, un effet inverse sur le poids de l'autre dimension.

En conclusion, l'application du modèle INDSCAL aux résultats de

cette première expérience s'est révélée extrêmement positive. La qualité de l'ajustement entre les plans physique et perceptif est excellente sinon supérieure à ceux obtenus généralement (Hanson, 1967; Pols, 1969). Ceci montre que les deux principes directeurs de la méthode INDSCAL : – dimensions identiques sous différentes conditions; – importances relatives de ces dimensions, propres à chaque condition rendent parfaitement compte du processus perceptif. Et en effet, chacun de

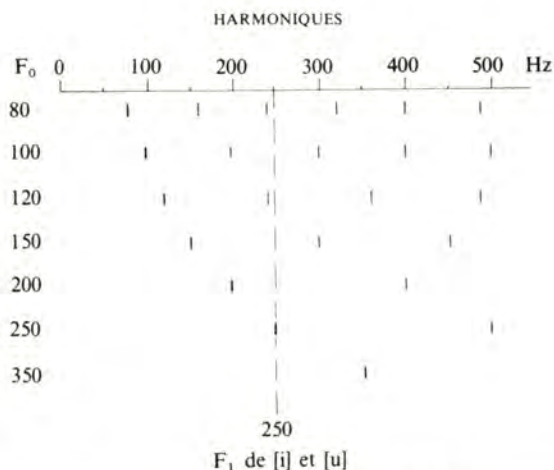


FIG. 9. LOCALISATION DU PREMIER FORMANT DE [i] ET [u] PAR RAPPORT À LA POSITION DES HARMONIQUES DE F_0 — THE LOCALISATION OF THE FIRST FORMANT FOR [i] AND [u] WITH RESPECT TO THE POSITION OF THE HARMONICS OF F_0

ces deux principes s'est concrétisé par des résultats nets et parfaitement interprétables. D'une part, les deux premières dimensions obtenues correspondent très clairement aux deux premiers formants. D'autre part, l'évolution des pondérations de ces deux dimensions met en évidence un double rôle de la fréquence fondamentale : – une normalisation globale de type linéaire; – une influence du facteur de définition spectrale sur chaque formant.

EXPÉRIENCE II

L'intérêt d'INDSCAL pour la compréhension du mécanisme d'identification de voyelles synthétiques nous a conduit à étendre ce type d'analyse à une expérience plus vaste portant sur des stimuli naturels. Deux aspects au moins des résultats obtenus pouvaient en effet être attribués au caractère artificiel des stimuli synthétiques employés :

1. Les seules variables physiques différenciant les 7 voyelles étaient F_1 et F_2 . Dès lors, qu'elles apparaissent comme dimensions perceptives

ne prouve pas qu'il en aille nécessairement de même lorsque d'autres paramètres physiques interviennent avec leurs fluctuations propres.

2. Le facteur de définition spectrale et donc son influence sur les importances relatives des deux dimensions était dû vraisemblablement à l'absence de micromélogie.

En outre, la première expérience nous aura permis de mieux cerner la signification de "dimensions perceptives". L'obtention de dimensions pertinentes est conditionnée par la variable utilisée. Il faut que cette variable influence chacune des dimensions de manière spécifique. C'est le cas, nous l'avons vu, pour le facteur de définition spectrale qui affecte différemment F_1 et F_2 . A l'opposé, la normalisation globale en fonction de la fondamentale a pour effet principal un glissement des valeurs formantiques. Ce type de normalisation est compatible avec les dimensions mais n'aurait pas permis de les mettre en évidence.

Il faut, par ailleurs, garder en mémoire que les échelles multidimensionnelles ne sont qu'une manière de visualiser un processus. Le processus qui nous occupe, l'identification, consiste en une série de transformations de l'information pour aboutir à une décision finale. Définir des dimensions à un niveau global risque de donner une image figée qui ne recouvre pas la réalité dynamique du phénomène. Par contre, la notion de dimensions d'un espace psychologique est tout à fait pertinente si on se place à un niveau bien précis dans la chaîne de traitement. C'est exactement ce que fait le modèle : les configurations sont structurées par rapport à la variable considérée et les dimensions sont alors représentatives du niveau auquel cette variable intervient. L'influence de la fréquence fondamentale à cet égard se produit vraisemblablement à un stade de traitement périphérique et dès lors, il serait prématuré de conclure que les deux formants constituent des dimensions effectives pour l'ensemble du traitement.

Il nous est donc apparu indispensable de reposer le problème des dimensions perceptives des voyelles à partir de plusieurs variables susceptibles d'intervenir à divers niveaux du traitement.

MÉTHODE EXPÉRIMENTALE

Stimuli : les dix voyelles suivantes [a], [ɔ], [o], [u], [y], [i], [ɛ], [e], [ə], [œ] ont été produites par dix locuteurs naïfs, cinq hommes et cinq femmes, tous francophones, habitant la région bruxelloise. Le locuteur était installé dans une chambre sourde face à un micro Neumann U 67 relié à une chaîne d'enregistrement Ampex 300. Il avait pour consigne de prononcer les dix voyelles dans l'ordre en conservant, dans la mesure du possible, une hauteur de fondamentale constante. Tous les locuteurs furent à même de remplir cette exigence après un bref entraînement. Trois séquences de dix voyelles furent enregistrées pour chaque locuteur. La série la plus homogène du point de vue de la fondamentale fut recopiée en éliminant les différences d'intensité inter-voyelles. Le matériel ainsi obtenu, soit 10 × 10 voyelles, fut ensuite

digitalisé sur P.D.P. 12, à un échantillonnage de 40 μ s grâce à un programme réalisé par P. Jospa (à paraître). Une période fut prélevée dans la partie stable de chaque voyelle de manière à obtenir des échantillons de longueur égale pour toutes les voyelles d'un même locuteur. Les fréquences correspondantes s'évaluaient de 80 à 235 Hz.

Sujets : dix sujets choisis selon les mêmes critères que les locuteurs ont participé à l'expérience.

Procédure : dix séries expérimentales de 200 stimuli ont été constituées, chaque série correspondant à un locuteur. Chacune des 10 voyelles apparaissait 20 fois dans un ordre pseudo-aléatoire de sorte que chaque stimulus était suivi deux fois par chacun des autres et deux fois par lui-même. Les sujets ont été soumis à l'écoute des dix séries, dans un ordre aléatoire identique mais débutant par un locuteur différent. Chaque séance durait une heure environ et comportait deux séries séparées par une pause. L'ensemble des cinq séances s'est étalé sur deux semaines. Le sujet était placé dans une cabine insonorisée et prévenu de l'arrivée d'un stimulus par l'affichage d'un chiffre sur un compteur lumineux. L'affichage des chiffres et l'envoi des stimuli suivait un rythme régulier contrôlé par ordinateur. Le sujet disposait de quatre secondes pour choisir entre les dix réponses possibles et l'inscrire sur le protocole. A cette fin, nous lui avons fourni l'éventail des voyelles transcrites d'une manière naturelle :

a ɔ o ou u i è é eu e

correspondant à

[a] [ɔ] [o] [u] [y] [i] [ɛ] [e] [ø] [œ]

RÉSULTATS

L'examen des 10×10 matrices de confusions pour chaque sujet et pour chaque locuteur montre que les voyelles cardinales sont les mieux perçues à l'exception du [i]. Dans ce cas précis, la segmentation a vraisemblablement un effet perturbateur plus important du fait que les hautes fréquences dominent (Serniclaes, 1973). Les différences inter-voyelles sont sensiblement les mêmes pour tous les sujets; par contre, le taux global de reconnaissance varie de 10 à 80%. Les variations correspondant aux locuteurs sont nettement plus faibles (40 à 60%). L'examen de ces différences montre que la longueur du segment vocalique (différente pour chaque locuteur) n'a pas d'effet systématique sur le pourcentage de reconnaissance. Ceci est vraisemblablement dû au fait que les phénomènes d'amortissement et de traînage engendrés par le haut-parleur nivellent les différences de longueur des stimuli sans toutefois fournir d'information supplémentaire concernant la fréquence fondamentale.

Les variables retenues, susceptibles d'intervenir dans le processus d'identification étaient 1. différents locuteurs; 2. différents sujets; 3. différents contextes objectifs (voyelles précédentes); 4. différents contextes subjectifs (réponses précédentes); 5. entraînement (ordre de passation). Pour chacune de ces cinq conditions, les résultats étaient sommés sans tenir compte des autres variables. La matrice globale de confusions inter-voyelles fut donc décomposée de cinq manières différentes en 10 matrices correspondant à chaque locuteur, chaque sujet, etc... La méthode INDSCAL fut appliquée à chacune de ces cinq situations; la procédure utilisée pour l'ensemble des analyses était conforme en tout point à celle décrite pour la première expérience.

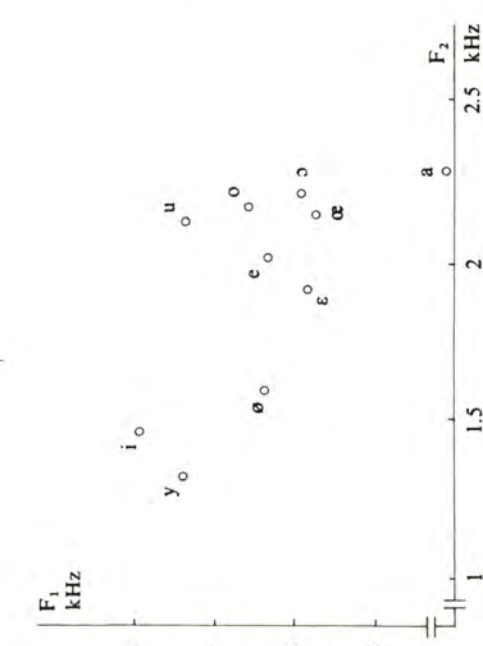
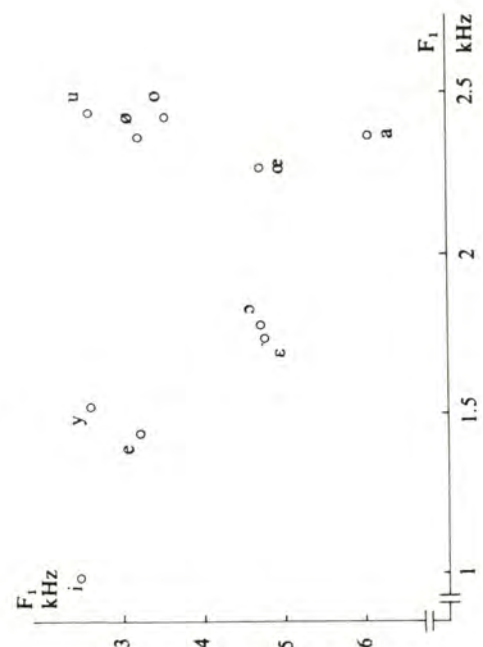
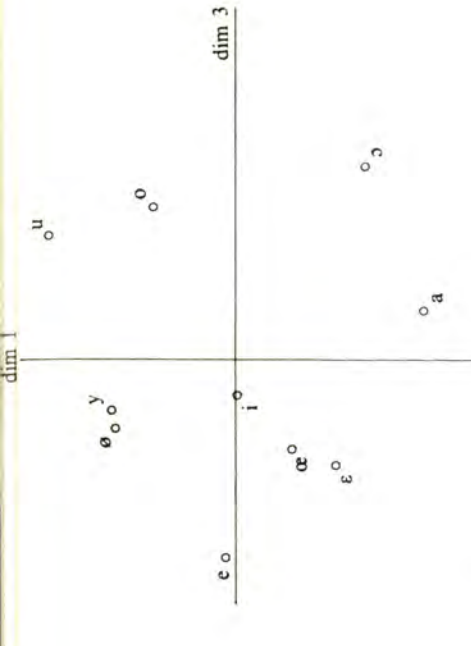
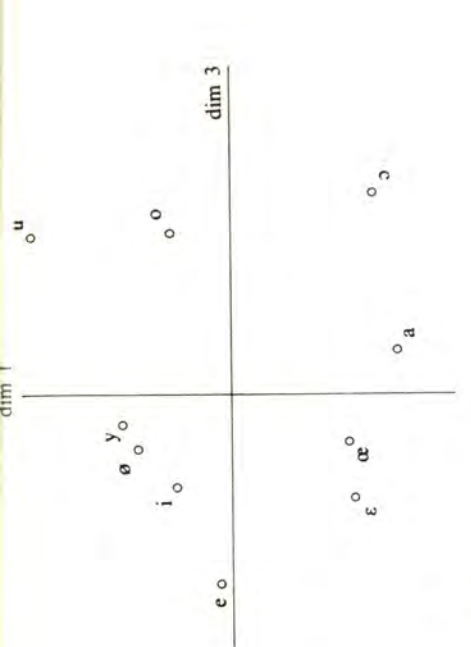
DISCUSSION

En ce qui concerne les variables 3, 4 et 5, les dimensions extraites par l'analyse sont aléatoires: lorsque le nombre de dimensions testées croît d'une unité, on ne trouve pas dans la nouvelle structure les dimensions extraites antérieurement. Ceci signifie soit que ces variables ne jouent aucun rôle dans le processus perceptif, soit que leur influence ne se manifeste pas de façon spécifique pour chaque dimension.

Cette deuxième hypothèse est plus plausible au moins en ce qui concerne le contexte objectif: des effets séquentiels importants ont été mis en évidence pour l'identification des voyelles. Thompson & Hollien (1970) ont montré que les déplacements obtenus étaient linéaires dans le plan F_1, F_2 . Ce type d'adaptation est tout-à-fait comparable à celle mise en évidence pour la fréquence fondamentale au cours de notre première expérience et explique parfaitement le caractère aléatoire des dimensions.

Par contre, en ce qui concerne les variables locuteurs et sujets, l'analyse fournit des dimensions pertinentes. Seules les trois premières dimensions ont été retenues dans chaque cas, les dimensions supplémentaires n'apportant qu'une faible augmentation de la variance expliquée. Les pourcentages de cette variance dont rendent compte les trois premières dimensions pour les conditions locuteur et sujet étaient respectivement 64 et 58%. Ce résultat est médiocre par rapport à la première expérience mais s'explique si l'on tient compte, d'abord du nombre plus élevé de stimuli, ensuite, de la difficulté de la tâche qui est responsable de bruit chez les sujets les moins performants, enfin de l'hétérogénéité des voyelles introduite par l'emploi de plusieurs locuteurs. Ce dernier facteur ne joue que dans la condition sujet, et en effet la variance expliquée est moindre dans ce cas.

FIG. 10. ESPACES PERCEPTIFS POUR LES CONDITIONS SUJET (a) ET LOCUTEUR (b), ET ESPACES PHYSIQUES F_1 - F_2 POUR LES VOIX D'HOMMES (c) ET DE FEMMES (d) — THE PERCEPTUAL SPACES CORRESPONDING TO THE "LISTENER CONDITION" (a) AND THE "SPEAKER CONDITION" (b); THE PHYSICAL SPACES F_1 - F_2 CORRESPONDING TO THE AVERAGE FEMALE VOWEL (d) AND MALE VOWEL (c) DATA



Les graphiques 10a et 10b montrent les espaces DIM 1, DIM 3 pour les deux conditions sujets et locuteurs². Les deux structures sont très semblables et correspondent grossièrement au plan formantique, exception faite du [i]; en outre, les deux dimensions correspondent respectivement à F_1 et F_2 dans les deux cas. Il semble donc que les deux premiers formants soient des dimensions pertinentes pour l'ensemble du processus d'identification et non pas des dimensions spécifiques du niveau acoustique auquel agit la fréquence fondamentale. Ce résultat, dans son ensemble, confirme une des alternatives que nous avons posées au départ de cette expérience.

On pouvait s'attendre à ce que les différences individuelles aillent dans ce sens: les premier et second formants pour l'ensemble des voyelles se localisent dans deux régions fréquentielles différentes. Or, le degré de précision dans la localisation des concentrations d'énergie est fonction de la fréquence. Dès lors, la simple hypothèse de variations individuelles pour cette fonction implique des différences spécifiques pour F_1 et F_2 . Par contre, que F_1 et F_2 constituent les dimensions perceptives pour la condition locuteur, est surprenant.

Ce résultat implique en effet que les sujets sont capables de s'adapter à un locuteur en ce sens qu'ils attribuent des poids différents à chacune des deux informations concernant F_1 et F_2 . Ceci est d'autant plus étonnant que dans le cadre de l'expérience, les sujets ne pouvaient tenir compte des variances de F_1 et F_2 propres à chaque locuteur, puisqu'ils n'avaient à leur disposition qu'un exemplaire de chaque voyelle. L'explication doit donc être trouvée au niveau d'un processus de normalisation basé sur les données spectrales de l'ensemble des voyelles. Une étude de Haggard (1973) suggère que ce type de normalisation au locuteur serait plus effectif que la normalisation basée sur F_0 : cette dernière en effet, n'était employée par un certain nombre de sujets que lorsque les données spectrales étaient insuffisantes. Dans une autre étude concernant l'adaptation au locuteur, Sheldon (1973) a montré qu'une simple normalisation linéaire de F_1 et F_2 échouait à rendre compte des résultats expérimentaux. La comparaison des résultats obtenus pour la condition locuteur et pour la fondamentale (première expérience) permet de différencier ces deux types d'adaptation.

L'adaptation en fonction de la fondamentale se présente comme un simple déplacement des valeurs formantiques, identique pour les deux formants. Ce type de normalisation est sans doute important pour de grandes différences de fondamentale (notamment les différences hommes-femmes-enfants) mais insuffisant pour expliquer une adaptation au locuteur. En outre, le fait que les deux informations concernant F_1 et F_2 sont extraites d'un même continuum (spectral), suggère que F_0 affecte ce continuum avant l'émergence des deux dimensions psychologiques.

² Comme dans le cas de la première expérience, la seconde dimension sépare les stimuli "extrêmes" des stimuli "moyens".

Par contre, en ce qui concerne la normalisation en fonction des données spectrales, tout se passe comme si les deux échelles psychologiques correspondant respectivement à F_1 et F_2 étaient calibrées de façon appropriée à chaque locuteur. Pour que cette "équation personnelle", que le système perceptif semble à même d'établir à partir des données spectrales pour un locuteur, constitue une normalisation efficace, il faut logiquement trouver une loi identique à la production.

Cette loi pourrait se traduire soit, par des déformations systématiques du triangle vocalique suivant l'un ou l'autre formant, soit, par des différences systématiques des variances propres à chaque formant. Le schéma 11 illustre ces deux types d'hypothèses : dans les

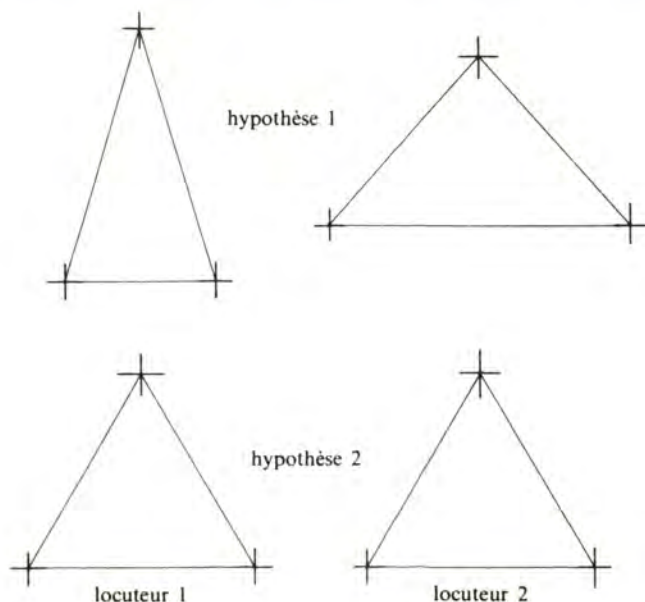


FIG. 11. DEUX HYPOTHÈSES POSSIBLES POUR RENDRE COMPTE DES DIFFÉRENCES D'IMPORTANCES RELATIVES DE F_1 ET F_2 SPÉCIFIQUES À UN LOCUTEUR DONNÉ — TWO POSSIBLE HYPOTHESES TO EXPLAIN THE RELATIVE SALIENCES ACCOUNTED FOR EACH F_1 AND F_2 , RELATIVE TO ONE SPEAKER

deux cas, l'attribution d'importances relatives de F_1 et F_2 adaptées à chaque locuteur représenterait un gain certain dans la tâche d'identification.

A notre connaissance, aucune étude à la production ne permet pour l'instant de confirmer l'une ou l'autre de ces hypothèses. La seconde hypothèse est néanmoins peu probable : les sujets ne pouvaient estimer les variances de F_1 et F_2 propres à chaque locuteur, puisqu'ils n'avaient à leur disposition qu'un seul exemplaire de chaque voyelle.

A l'appui de la première hypothèse – déformations du triangle vocalique suivant F_1 ou F_2 – on ne peut évoquer d'argument d'ordre articulatoire dans la mesure où les deux formants ne correspondent pas à deux paramètres spécifiques et indépendants. Ce serait une erreur, cependant, que de limiter la description de la production de la parole à de seules considérations articulatoires. A l'intérieur des contraintes imposées par l'appareil vocal, le locuteur ajuste son système vocalique à partir des informations que son système perceptif extrait de l'entourage linguistique. Dès lors, il n'est pas exclu qu'une composante importante des différences individuelles entre locuteurs dérive directement de caractéristiques au niveau perceptif.

Dans cette optique, la différenciation perceptive des deux formants aboutirait à un phénomène analogue à la production. Ceci expliquerait la grande similitude entre les espaces obtenus pour les deux conditions. En outre, cette hypothèse rend compte d'une certaine cohérence interne au système de communication parlée : l'individu étant à la fois auditeur et locuteur, les propriétés perceptives responsables de différences individuelles à la production seraient elles-mêmes à la base du mécanisme d'adaptation au locuteur.

Une possibilité de tester directement les hypothèses proposées consistait à vérifier la différenciation de F_1 et F_2 à partir du matériel employé pour l'expérience. Nous avons dans ce but, extrait à partir des valeurs formantiques fournies par l'analyse des 10×10 voyelles, les espaces spécifiques à chaque locuteur. Nous espérons que l'application d'INDSCAL aux matrices de distances dans le plan physique F_1 , F_2 fournirait des résultats analogues à ceux obtenus avec les espaces perceptifs.

Cette tentative n'a pu aboutir à cause de la difficulté d'extraire des valeurs "correctes" de F_1 et F_2 à partir des données spectrales. De nombreux auteurs (Fant, 1960; Van Nierop, 1973) ont signalé que dans 30% des cas environ, la détermination des formants à partir du spectre n'était possible qu'à condition de tenir compte de l'identité de la voyelle. Dans notre cas, le nombre d'ambiguïtés était encore plus élevé, vraisemblablement par l'emploi de locuteurs non entraînés. Plusieurs programmes d'analyses réalisés par Carayannis utilisant la méthode de prédiction linéaire (Makhoul, 1975; Carayannis, 1973) et la méthode spectre (Oppenheim, 1969) furent appliqués soit directement au signal envoyé par STIM, soit à la sortie du haut-parleur.

Les divers traitements aboutissaient à des spectres fort semblables et, en aucun cas, ne permettaient de lever les ambiguïtés quant au choix de F_1 et F_2 . Le critère objectif (sans tenir compte de l'identité de la voyelle) qui s'est avéré le plus efficace consistait à choisir la fréquence correspondant au maximum d'énergie dans deux régions du spectre spécifiques à chaque formant (F_1 entre 0 et 100 Hz.; F_2 entre 700 et 2700 Hz.).

Partant des valeurs ainsi obtenues, les graphiques 10c et 10d illustrent la répartition des voyelles dans le plan F_1 , F_2 respectivement pour les locuteurs hommes et femmes. La comparaison entre ces espaces phy-

siques et l'espace perceptif (Graph. 10b) montre clairement que les sujets extraient les deux formants de manière bien supérieure à ce que permet l'analyse spectrale. On pouvait donc s'attendre à ce qu'INDSCAL échoue à dégager F_1 et F_2 comme dimensions à partir d'espaces où figure un pourcentage important de valeurs aberrantes.

La question qui se pose immédiatement est de savoir comment les sujets procèdent pour déterminer les formants dans les cas ambigus. Nous n'avons aucune hypothèse à proposer en ce qui concerne cette question; elle nous semble fondamentale dans la mesure où le perfectionnement des méthodes d'analyse spectrale n'apporte pas de solution au problème du passage du spectre aux formants, ce qui reste la difficulté majeure en analyse automatique aussi bien que pour la compréhension du mécanisme perceptif.

La seule indication que fournissent nos résultats est que la procédure employée dans les situations ambiguës doit se baser sur des éléments différents suivant le cas, puisque seules les deux dimensions F_1 et F_2 apparaissent.

CONCLUSION

Pour chacune de ces deux applications, la méthode INDSCAL s'est avérée un outil extrêmement efficace non seulement pour la description d'un processus tel que l'identification des voyelles mais aussi pour la mise en évidence de ses caractéristiques fonctionnelles. C'est principalement la possibilité d'extraire des dimensions pertinentes en fonction d'une variable donnée, qui rend cette méthode plus fructueuse que les analyses multidimensionnelles classiques.

En effet, l'application d'INDSCAL à l'espace perceptif des voyelles a montré clairement que les deux premiers formants constituent deux dimensions perceptives qui se différencient très tôt dans le processus d'identification et évoluent indépendamment tout au long du traitement pour constituer à son terme, les deux critères majeurs de décision.

Outre ce résultat, l'emploi de différentes variables, a permis de mettre en évidence deux types distincts de normalisation: d'une part, la normalisation en fonction de la fréquence fondamentale qui consiste en un déplacement identique pour chaque dimension, d'autre part, l'adaptation au locuteur qui se traduit par un rapport d'importance spécifique, relativement aux différentes dimensions.

RÉFÉRENCES

- CARAYANNIS, G. *Analyse de la parole par identification récurrente d'un modèle de phonation*. Thèse de Docteur Ingénieur non publiée, Université de Paris 7, 1973.
- CARROL, J.D., & CHANG, J.J. Analysis of individual differences in multidimensional scaling via an N-way generalisation of "Eckart-Young" decomposition. *Psychometrika*, 1970, 35, 283-319.
- FANT, G. *Acoustic theory of speech production*. 's-Gravenhage: Mouton, 1960.

- FANT, G. *Speech sounds and features*. Cambridge-London, 1973.
- FUJISAKI, H., & KAWASHIMA, T. The roles of pitch and higher formants in the perception of vowels. *I.E.E.E. (The Institute of Electrical and Electronics Engineers) Audio Electroacoustics*, AU-16, 73-77.
- HANSON, G. Dimensions in speech sound perception: An experimental study of vowel perception. *Ericsson Technics*, 1967, 23, 3-175.
- KRUSKAL, J.B. Multidimensional scaling: A numerical method. *Psychometrika*, 1964 (a), 29, 1-27.
- KRUSKAL, J.B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 1964(b), 29, 115-129.
- KRUSKAL, J.B. More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika*, 1976, 41, 281-293.
- MAKHOUL, J. Linear prediction: A tutorial review. *Proceedings of the I.E.E.E. (The Institute of Electrical and Electronics Engineers)*, 1975, 63, 561-580.
- OPPENHEIM, A.V. Speech analysis-synthesis system based on homomorphic filtering. *Journal of the Acoustical Society of America*, 1969, 45, 458-465.
- PEARSON, W.H. Estimation of a correlation coefficient from an uncertainty measure. *Psychometrika*, 1966, 31, 421-433.
- POLS, L.C.W., VANDER KAMP, L.J.TH., & PLOMB, R. Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, 1969, 46, 458-467.
- SERNICLAES, W. L'ajustement du cadre vocalique interne en fonction de la voix et des voyelles. *Rapport d'Activités de l'Institut de Phonétique (U.L.B.)*, 1973, 7/1, 51-74.
- SHELDON, D.R. The identification of synthetic vowels in real vowel frames from different speakers. *Journal of Phonetics*, 1963, 1, 149-163.
- SHEPARD, R.N. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 1962, 27, 125-140; 219-246.
- SUMMERFIELD, A.Q., & HAGGARD, M.P. Vocal tract normalisation as demonstrated by reaction times. *Speech Perception (Queen's University of Belfast)*, Prog. Rep., series 2, n° 2, 1-8.
- THOMPSON, C.L., & HOLLIN, H. Some contextual effects on the perception of synthetic vowels. *Language and Speech*, 1970, 13, 1-13.
- TORGERSON, W.S. *Theory and methods of scaling*. New York: Wiley, 1958.
- VAN NIEROP, D.J.P.J., POLS, L.C.W., & PLOMB, R. Frequency analysis of Dutch vowels from 25 female speakers. *Acustica*, 1973, 29, 110-118.

Institut de Phonétique
Université Libre de Bruxelles
1050 Bruxelles

Reçu mars 1977