

# THESE

présentée à

L'UNIVERSITE SCIENTIFIQUE ET MEDICALE DE  
G R E N O B L E

pour obtenir

LE GRADE DE DOCTEUR ÈS - SCIENCES PHYSIQUES

par

**René CARRÉ**

-

CONTRIBUTION AUX ETUDES SUR

**L'ANALYSE ET LA SYNTHÈSE DE LA PAROLE**

Rôle et importance des formants

-

Thèse soutenue le 30 Juin 1971, devant la Commission d'Examen

J U R Y

MM. **J. BENOIT** Président ,

**R. LANCIA**  
**R. LEHMANN**  
**M. WAJSKOP** } Examineurs



UNIVERSITE SCIENTIFIQUE  
ET MEDICALE DE GRENOBLE

---

LISTE DES  
PROFESSEURS

---

PRESIDENT

Monsieur Michel *S O U T I F*

VICE-PRESIDENT

Monsieur Gabriel *C A U* .

PROFESSEURS TITULAIRES

<b>MM.</b>	<i>ANGLES D'AURIAC</i>	Paul	Mécanique des Fluides
	<i>ARNAUD</i>	Georges	Clinique des Maladies Infectieuses
	<i>ARNAUD</i>	Paul	Chimie
	<i>AYANT</i>	Yves	Physique approfondie
	<i>BARBIER</i>	Jean-Claude	Physique expérimentale
	<i>BARBIER</i>	Reynold	Géologie appliquée
	<i>BARJON</i>	Robert	Physique nucléaire
	<i>BARNOUD</i>	Fernand	Biosynthèse de la Cellulose
	<i>BARRIE</i>	Joseph	Clinique chirurgicale
	<i>BENOIT</i>	Jean	Radioélectricité
	<i>BESSON</i>	Jean	Electrochimie
	<i>BEZES</i>	Henri	Chirurgie générale
	<i>BLAMBERT</i>	Maurice	Mathématiques pures
	<i>BOLLIET</i>	Louis	Informatique (IUT B)
	<i>BONNET</i>	Georges	Electrotechnique
	<i>BONNET</i>	Jean-Louis	Clinique ophtalmologique
	<i>BONNET-EYMARD</i>	Joseph	Pathologie médicale

MM.	BONNIER	Etienne	Electrochimie - Electrometallurgie
	BOUCHERLE	André	Chimie et Toxicologie
	BOUCHEZ	Robert	Physique nucléaire
	BRAVARD	Yves	Géographie
	BRISSENEAU	Pierre	Physique du Solide
	BUYLE-BODIN	Maurice	Electronique
	CABANAC	Jean	Pathologie chirurgicale
	CABANEL	Guy	Clinique rhumatologique et hydrologique
	CALAS	François	Anatomie
	CARRAZ	Gilbert	Biologie animale et Pharmacodynamie
	CAU	Gabriel	Médecine légale et Toxicologie
	CAUQUIS	Georges	Chimie organique
	CHABAUTY	Claude	Mathématiques pures
	CHATEAU	Robert	Thérapeutique
	CHENE	Marcel	Chimie papetière
	COEUR	André	Pharmacie chimique
	CONTAMIN	Robert	Clinique gynécologique
	COUDERC	Pierre	Anatomie pathologique
	CRAYA	Antoine	Mécanique
Mme	DEBELMAS	Anne-Marie	Matière médicale
MM.	DEBELMAS	Jacques	Géologie générale
	DEGRANGE	Charles	Zoologie
	DESSEAUX	Georges	Physiologie animale
	DODU	Jacques	Mécanique appliquée
	DREYFUS	Bernard	Thermodynamique
	DUCROS	Pierre	Cristallographie
	DUGOIS	Pierre	Clinique de Dermatologie et Syphiligraphie
	FAU	René	Clinique neuro-psychiatrique
	FELICI	Noël	Electrostatique
	GAGNAIRE	Didier	Chimie physique
	GALLISSOT	François	Mathématiques pures
	GALVANI	Octave	Mathématiques pures
	GASTINEL	Noël	Analyse numérique
	GERBER	Robert	Mathématiques pures
	GIRAUD	Pierre	Géologie
	KLEIN	Joseph	Mathématiques pures
Mme	KOFLER	Lucie	Botanique et Physiologie végétale
MM.	KOSZUL	Jean-Louis	Mathématiques pures
	KRAVTCHENKO	Julien	Mécanique
	KUNTZMANN	Jean	Mathématiques appliquées
	LACAZE	Albert	Thermodynamique
	LACHARME	Jean	Biologie végétale
	LATURAZE	Jean	Biochimie pharmaceutique
	LEDRU	Jean	Clinique médicale B
	LLIBOUTRY	Louis	Géophysique
	LOUP	Jean	Géographie
Mle	LUTZ	Elisabeth	Mathématiques pures
MM.	MALGRANGE	Bernard	Mathématiques pures
	MALINAS	Yves	Clinique obstétricale
	MARTIN-NOEL	Pierre	Séméiologie médicale
	MASSEPORT	Jean	Géographie
	MAZARE	Yves	Clinique médicale A
	MICHEL	Robert	Minéralogie et Pétrographie

MM.	<i>BOUVARD</i>	Maurice	Mécanique des Fluides
	<i>BRIERE</i>	Georges	Physique expérimentale
	<i>BRODEAU</i>	François	Mathématiques (IUT B)
	<i>BRUGEL</i>	Lucien	Energétique
	<i>BUISSON</i>	Roger	Physique
	<i>BUTEL</i>	Jean	Orthopédie
	<i>CHAMBAZ</i>	Edmond	Biochimie médicale
	<i>CHAMPETIER</i>	Jean	Anatomie et Organogénèse
	<i>CHARACHON</i>	Robert	Oto-rhyno-laryngologie
	<i>CHIAVERINA</i>	Jean	Biologie appliquée (EFP)
	<i>CHIBON</i>	Pierre	Biologie animale
	<i>COHEN-ADDAD</i>	Jean-Pierre	Spectrométrie physique
	<i>COLOMB</i>	Maurice	Biochimie médicale
	<i>CONTE</i>	René	Physique
	<i>CROUZET</i>	Guy	Radiologie
	<i>DURAND</i>	Francis	Métallurgie
	<i>DUSSAUD</i>	René	Mathématiques (CUS)
Mme	<i>ETERRADOSSI</i>	Jacqueline	Physiologie
MM.	<i>FAURE</i>	Jacques	Médecine légale
	<i>GAVEND</i>	Michel	Pharmacologie
	<i>GENSAC</i>	Pierre	Botanique
	<i>GERMAIN</i>	Jean-Pierre	Mécanique
	<i>GIDON</i>	Maurice	Géologie
	<i>GRIFFITHS</i>	Michaël	Mathématiques appliquées
	<i>GROULADE</i>	Joseph	Biochimie médicale
	<i>HOLLARD</i>	Daniel	Hématologie
	<i>HUGONOT</i>	Robert	Hygiène et Médecine préventive
	<i>IDELMAN</i>	Simon	Physiologie animale
	<i>IVANES</i>	Marcel	Electricité
	<i>JALBERT</i>	Pierre	Histologie
	<i>JOLY</i>	Jean-René	Mathématiques pures
	<i>JOUBERT</i>	Jean-Claude	Physique du Solide
	<i>JULLIEN</i>	Pierre	Mathématiques pures
	<i>KAHANE</i>	André	Physique générale
	<i>KUHN</i>	Gérard	Physique
Mme	<i>LAJZEROWICZ</i>	Jeannine	Physique
MM.	<i>LAJZEROWICZ</i>	Joseph	Physique
	<i>LANCIA</i>	Roland	Physique (Automatique)
	<i>LE JUNTER</i>	Noël	Electronique
	<i>LEROY</i>	Philippe	Mathématiques
	<i>LOISEAUX</i>	Jean-Marie	Physique nucléaire
	<i>LONGEQUEUE</i>	Jean-Pierre	Physique nucléaire
	<i>LUU DUC</i>	Cuong	Chimie organique
	<i>MACHE</i>	Régis	Physiologie végétale
	<i>MAGNIN</i>	Robert	Hygiène et Médecine préventive
	<i>MARECHAL</i>	Jean	Mécanique
	<i>MARTIN-BOUYER</i>	Michel	Chimie (CUS)
	<i>MAYNARD</i>	Roger	Physique du Solide
	<i>MICOUD</i>	Max	Maladies infectieuses
	<i>MOREAU</i>	René	Hydraulique (INP)
	<i>NEGRE</i>	Robert	Mécanique
	<i>PARAMELLE</i>	Bernard	Pneumologie

MM.	<i>MOURIQUAND</i>	Claude	Histologie
	<i>MOUSSA</i>	André	Chimie nucléaire
	<i>NEEL</i>	Louis	Physique du Solide
	<i>OZENDA</i>	Paul	Botanique
	<i>PAUTHENET</i>	René	Electrotechnique
	<i>PAYAN</i>	Jean-Jacques	Mathématiques pures
	<i>PEBAY-PEROULA</i>	Jean-Claude	Physique
	<i>PERRET</i>	René	Servomécanismes
	<i>PILLET</i>	Emile	Physique industrielle
	<i>RASSAT</i>	André	Chimie systématique
	<i>RENARD</i>	Michel	Thermodynamique
	<i>REULOS</i>	René	Physique industrielle
	<i>RINALDI</i>	Renaud	Physique
	<i>ROGET</i>	Jean	Clinique de Pédiatrie et de Puériculture
	<i>SANTON</i>	Lucien	Mécanique
	<i>SEIGNEURIN</i>	Raymond	Microbiologie et Hygiène
	<i>SENGEL</i>	Philippe	Zoologie
	<i>SILBERT</i>	Robert	Mécanique des Fluides
	<i>SOUTIF</i>	Michel	Physique générale
	<i>TANCHE</i>	Maurice	Physiologie
	<i>TREYNARD</i>	Philippe	Chimie générale
	<i>VAILLAND</i>	François	Zoologie
	<i>VAUQUOIS</i>	Bernard	Calcul électronique
Mme	<i>VERAIN</i>	Alice	Pharmacie galénique
M.	<i>VERAIN</i>	André	Physique
Mme	<i>VEYRET</i>	Germaine	Géographie
MM.	<i>VEYRET</i>	Paul	Géographie
	<i>VIGNAIS</i>	Pierre	Biochimie médicale
	<i>YOCOZ</i>	Jean	Physique nucléaire théorique

#### PROFESSEURS ASSOCIES

MM.	<i>BULLEMER</i>	Bernhard	Physique
	<i>RADHAKRISHNA</i>	Pidatala	Thermodynamique

#### PROFESSEURS SANS CHAIRE

M.	<i>AUBERT</i>	Guy	Physique
Mme	<i>BARBIER</i>	Marie-Jeanne	Electrochimie
MM.	<i>BARRA</i>	Jean	Mathématiques appliquées
	<i>BEAUDOING</i>	André	Pédiatrie
	<i>BERTRANDIAS</i>	Jean-Paul	Mathématiques appliquées
	<i>BIAREZ</i>	Jean-Pierre	Mécanique

M.	<i>BONNETAIN</i>	Lucien	Chimie minérale
Mme	<i>BONNIER</i>	Jane	Chimie générale
MM.	<i>CARLIER</i>	Georges	Biologie végétale
	<i>COHEN</i>	Joseph	Electrotechnique
	<i>COUMES</i>	André	Radioélectricité
	<i>DEPASSEL</i>	Roger	Mécanique des Fluides
	<i>DEPORTES</i>	Charles	Chimie minérale
	<i>DESRE</i>	Pierre	Métallurgie
	<i>DOLIQUE</i>	Jean-Michel	Physique des Plasmas
	<i>GAUTHIER</i>	Yves	Sciences biologiques
	<i>GEINDRE</i>	Michel	Electroradiologie
	<i>GIDON</i>	Paul	Géologie et Minéralogie
	<i>GLENAT</i>	René	Chimie organique
	<i>HACQUES</i>	Gérard	Calcul numérique
	<i>JANIN</i>	Bernard	Géographie
Mme	<i>KAHANE</i>	Josette	Physique
MM.	<i>LATREILLE</i>	René	Chirurgie générale
	<i>LAURENT</i>	Pierre	Mathématiques appliquées
	<i>MULLER</i>	Jean-Michel	Thérapeutique
	<i>PERRIAUX</i>	Jean-Jacques	Géologie et Minéralogie
	<i>POLOUJADOFF</i>	Michel	Electrotechnique
	<i>REBECQ</i>	Jacques	Biologie (CUS)
	<i>REVOL</i>	Michel	Urologie
	<i>REYMOND</i>	Jean-Charles	Chirurgie générale
	<i>ROBERT</i>	André	Chimie papetière
	<i>SARRAZIN</i>	Roger	Anatomie et Chirurgie
	<i>SARROT-REYNAULD</i>	Jean	Géologie
	<i>SIBILLE</i>	Robert	Construction mécanique
	<i>SIROT</i>	Louis	Chirurgie générale
Mme	<i>SOUTIF</i>	Jeanne	Physique générale
M.	<i>VALENTIN</i>	Jacques	Physique nucléaire

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

Mles	<i>AGNUS-DELORD</i>	Claudine	Physique pharmaceutique
	<i>ALARY</i>	Josette	Chimie analytique
MM.	<i>AMBLARD</i>	Pierre	Dermatologie
	<i>AMBROISE-THOMAS</i>	Pierre	Parasitologie
	<i>ARMAND</i>	Yves	Chimie
	<i>BEGUIN</i>	Claude	Chimie organique
	<i>BELORIZKY</i>	Elie	Physique
	<i>BENZAKEN</i>	Claude	Mathématiques appliquées
Mme	<i>BERTRANDIAS</i>	Françoise	Mathématiques pures
MM.	<i>BLIMAN</i>	Samuel	Electronique (ENSERG)
	<i>BLOCH</i>	Daniel	Electrotechnique
Mme	<i>BOUCHE</i>	Liane	Mathématiques (CUS)
MM.	<i>BOUCHET</i>	Yves	Anatomie
	<i>BOUSSARD</i>	Jean-Claude	Mathématiques appliquées

MM.	PECCOUD	François	Analyse (IUT B)
	PEFFEN	René	Métallurgie
	PELMONT	Jean	Physiologie animale
	PERRET	Jean	Neurologie
	PERRIN	Louis	Pathologie expérimentale
	PFISTER	Jean-Claude	Physique du Solide
	PHELIP	Xavier	Rhumatologie
Mle	PIERY	Yvette	Biologie animale
MM.	RACHAIL	Michel	Médecine interne
	RACINET	Claude	Gynécologie et Obstétrique
	RICHARD	Lucien	Botanique
Mme	RINAUDO	Marguerite	Chimie macromoléculaire
MM.	ROMIER	Guy	Mathématiques (IUT B)
	ROUGEMONT (de)	Jacques	Neuro-chirurgie
	STIEGLITZ	Paul	Anesthésiologie
	STOEBNER	Pierre	Anatomie pathologique
	VAN CUTSEM	Bernard	Mathématiques appliquées
	VEILLON	Gérard	Mathématiques appliquées (INP)
	VIALON	Pierre	Géologie
	VOOG	Robert	Médecine interne
	VROUSSOS	Constantin	Radiologie
	ZADWORNY	François	Electronique

MAITRES DE CONFERENCES ASSOCIES

---

MM.	BOUDOURIS	Georges	Radioélectricité
	CHEEKE	John	Thermodynamique
	GOLDSCHMIDT	Hubert	Mathématiques
	YACOUD	Mahmoud	Médecine légale

CHARGES DE FONCTIONS DE MAITRES DE CONFERENCES

---

Mmes	BERIEL	Hélène	Physiologie
	RENAUDET	Jacqueline	Microbiologie

# TABLE DES MATIERES

	<i>Pages :</i>
INTRODUCTION .. .. .	1
CHAPITRE 1 - ETUDE STATISTIQUE SUR LES FORMANTS VOCAUX .. .. .	7
1.1. Introduction	7
1.2. Plages des formants	11
1.3. Distances entre formants voisins	14
1.4. Vitesses de variation des formants	15
1.5. Relations entre le troisième et les deux premiers formants	16
1.6. Conclusions	21
CHAPITRE 2 - ETUDE DES FORMANTS DANS LE CAS DE LA PRODUCTION DES VOYELLES PAR DES LOCUTEURS HOMMES ET FEMMES .. .. .	23
2.1. Introduction	23
2.2. Etude de la production des voyelles à l'aide d'un simulateur du conduit vocal	27
2.2.1. Simulateur vocal	27
2.2.2. Expériences de simulation du conduit vocal	28
2.2.3. Résultats	28
2.3. Analyse spectrale de voyelles naturelles	35
2.4. Conclusions	38
CHAPITRE 3 - LES FORMANTS, ELEMENTS D'IDENTIFICATION DES LOCUTEURS .. .. .	39
3.1. Introduction	39
3.2. Comparaison des fréquences des paramètres dans les plans $F_1-F_2$ , $F_2-F_3$ et $F_0-F_3$	42
3.2.1. Comparaisons dans le plan $F_1-F_2$	43
3.2.2. Comparaisons dans le plan $F_2-F_3$	46
3.2.3. Comparaisons dans le plan $F_0-F_3$	46
3.2.4. Discussion	50
3.3. Comparaison des fréquences des paramètres dans les plans répartition - fréquence et probabilité - fréquence	53
3.3.1. Ecart quadratique dans le cas du premier formant	53
3.3.2. Ecart quadratique dans le cas du deuxième formant	53
3.3.3. Ecart quadratique dans le cas du troisième formant	56
3.3.4. Ecart quadratique dans le cas du fondamental	56
3.3.5. Discussion	56
3.4. Conclusions	61
CHAPITRE 4 - ETUDE SUR LES TRANSITIONS D'AMPLITUDE .. .. .	63
4.1. Introduction	63
4.2. Mode opératoire	64
4.3. Résultats	66
4.4. Conclusions	69
CONCLUSION .. .. .	71
ANNEXE 1 .. .. .	73
ANNEXE 2 .. .. .	79
BIBLIOGRAPHIE .. .. .	81



*Le travail présenté dans ce mémoire a été effectué dans les locaux de l'Ecole Nationale Supérieure d'Electronique et de Radioélectricité de GRENOBLE. J'exprime ma reconnaissance à son Directeur, Monsieur le Professeur J. BENOIT, dont l'accueil bienveillant et les nombreux encouragements m'ont permis d'effectuer ce travail dans de bonnes conditions. Je le remercie vivement d'avoir bien voulu me faire l'honneur d'accepter la présidence du Jury.*

*Je tiens à exprimer ma profonde gratitude à Monsieur le Professeur R. LANCIA, Directeur du Laboratoire de la Communication parlée et de l'Instrumentation de Mesures, qui n'a cessé de me prodiguer encouragements et conseils. J'ai été particulièrement touché par son humanisme, sa bienveillante compréhension et la confiance qu'il m'a toujours témoignée.*

*Je remercie Monsieur le Professeur R. LEHMANN, Président du Centre Universitaire du MANS, qui a bien voulu faire partie du Jury et qui s'est toujours intéressé à mes travaux.*

*De nombreux échanges avec Monsieur le Professeur M. WAJSKOP, Directeur de l'Institut de Phonétique de l'Université libre de BRUXELLES, ont favorisé le développement de ce travail. Je l'en remercie vivement et lui suis reconnaissant d'avoir accepté de participer à ce Jury.*

*Mes recherches ont été facilitées par une excellente ambiance de travail et de coopération. J'en sais gré à tous mes camarades du "Groupe de Recherches sur l'Analyse et la Synthèse de la Parole".*

*Je n'oublierai pas d'adresser également mes remerciements à tout le personnel de l'Ecole auprès duquel j'ai toujours trouvé toute facilité pour la réalisation de mon travail. Je remercie, en particulier, Mademoiselle MERKLEN, qui a assuré la réalisation matérielle de cet ouvrage, pour son efficacité et son dynamisme.*



## I N T R O D U C T I O N

Les expériences effectuées à l'aide des synthétiseurs de parole nous ont montré que le riche spectre de la parole pouvait être, sans qu'il en découle une dégradation de l'intelligibilité du message parlé, réduit à un petit nombre de composantes.

La parole est très redondante, aussi bien sur le plan du signal lui-même que sur le plan linguistique.

Nous savons que 40 000 unités d'information par seconde environ sont transmises par une ligne téléphonique classique au cours d'une conversation alors que notre cerveau, selon certaines hypothèses, ne pourrait décoder, en fin de traitement, qu'un maximum de 50 unités d'information par seconde, ces 50 unités d'information étant porteuses de données sur le message ainsi que sur la qualité et sur les caractéristiques individuelles de la voix.

Nous cherchons donc à détecter, parmi les 40 000 unités d'information, les 50 essentielles qui permettent la décision, en fin de traitement, au niveau du cerveau.

Préalablement, l'oreille -comme capteur- a déjà effectué un tri et transformé le support de l'information. Ensuite, d'autres types de transformation sont mis en oeuvre pour calibrer les données ; enfin, des règles linguistiques, fruits d'un apprentissage, permettent encore de simplifier les signaux.

Nous voyons donc qu'il existe différents stades de traitement. Nous nous limiterons, ici, à une étude du signal physique de la parole sans tenir compte des phénomènes linguistiques qui entrent en jeu au niveau de la perception. Nous chercherons à préciser le rôle que peuvent jouer certaines composantes que nous appellerons "*paramètres*" comme support direct ou indirect de l'intelligibilité du message proprement dit ou des caractéristiques individuelles de la voix.

A ce niveau de traitement, ces paramètres réagissent les uns sur les autres et un paramètre donné n'est pas spécifique, sur le plan de la perception, d'une caractéristique déterminée.

Il est aussi possible qu'en se limitant à un simple traitement du signal ne tenant pas compte des données linguistiques, on soit dans l'impossibilité, à un instant donné, de détecter sûrement des paramètres porteurs directs ou indirects de l'information, dans le cas d'un masquage par du bruit par exemple ou du fait d'une insuffisance du nombre de composantes du signal original.

Que le problème de la détection des paramètres soit traité globalement ou bien par différentes approches moins ambitieuses, les phénomènes abordés sont, de toute manière, très complexes.

Le rôle et l'importance de telle ou telle composante peuvent être étudiés au moyen de l'analyse et de la synthèse et la méthode employée dépend des objectifs à atteindre :

- . le signal peut être étudié à l'aide de tous les procédés physiques connus d'analyse dans le but d'en détecter les éléments caractéristiques ;
- . ou bien, on étudie le signal en tenant compte de son mode de production en vue, en particulier, de déterminer les caractéristiques de l'appareil vocal qui a pu produire un tel signal ;
- . ou encore, le signal est étudié dans le cadre des contraintes imposées par le récepteur - dans ce dernier cas, l'analyseur doit alors avoir les caractéristiques de notre organe de perception.

Ces trois démarches peuvent être fructueuses mais la dernière est difficile sinon impossible à mettre en oeuvre, car le mécanisme de notre organe de perception est mal connu.

Par contre, les phénomènes qui entrent en jeu lors de la production de la parole sont plus faciles à étudier.

A la suite de ces deux remarques, il apparaît que notre organe de perception pourra avantageusement être testé en étudiant le rôle et l'importance de paramètres ayant une signification sur le plan de la production de la parole. Nous supposons alors que notre mécanisme de perception est adapté à celui de notre appareil de production, et réciproquement.

Nous pouvons étudier le rôle des paramètres articulatoires mais ces paramètres sont difficiles à préciser et leur détection automatique en temps réel n'est pas simple.

C'est là une des raisons pour lesquelles nos travaux sur la parole ont porté principalement sur les formants, paramètres caractéristiques des résonances des cavités du conduit vocal.

Ces paramètres sont très intéressants car on a montré qu'il était possible, à partir de ceux-ci, d'effectuer des synthèses de qualité. Les synthétiseurs à formants sont, en effet, réputés pour la qualité de la parole qu'ils élaborent ainsi, d'ailleurs, que pour la très importante compression de bande qu'ils permettent d'escompter (cette dernière possibilité découle directement du caractère "pertinent" des formants).

D'autre part, le synthétiseur à formants est tout-à-fait adapté aux recherches fondamentales sur la parole.

Malheureusement, les formants sont difficiles à détecter automatiquement et en temps réel et c'est la raison pour laquelle les vocoders à formants n'ont pas encore pris le pas sur les vocoders classiques à canaux.

Les travaux sur les formants sont donc entrepris selon deux perspectives :

- Pour faciliter la détection des formants, on doit effectuer de nombreuses analyses pour en préciser toutes les caractéristiques. On interprètera les résultats obtenus en se référant aux phénomènes qui entrent en jeu lors de la production de la parole.
- En se situant au niveau du récepteur, on doit déterminer le rôle et l'importance de chacun des formants comme porteur d'informations.

La théorie du fonctionnement de l'appareil vocal, rappelée en Annexe I, montre que les formants sont bien définis si le nombre de composantes spectrales est important. Pour cela, la fréquence fondamentale doit être basse. Le spectre d'une voix de femme est pauvre et, par suite, les formants sont peu "visibles" et difficiles à détecter. En effet, la fréquence fondamentale d'une telle voix est, en moyenne, supérieure d'une octave à celle d'une voix d'homme.

Nous avons représenté figures 1 et 2 (page 4) les sonagrammes d'une même phrase prononcée respectivement par un homme et par une femme.

La bande passante du filtre d'analyse est de 300 Hz.

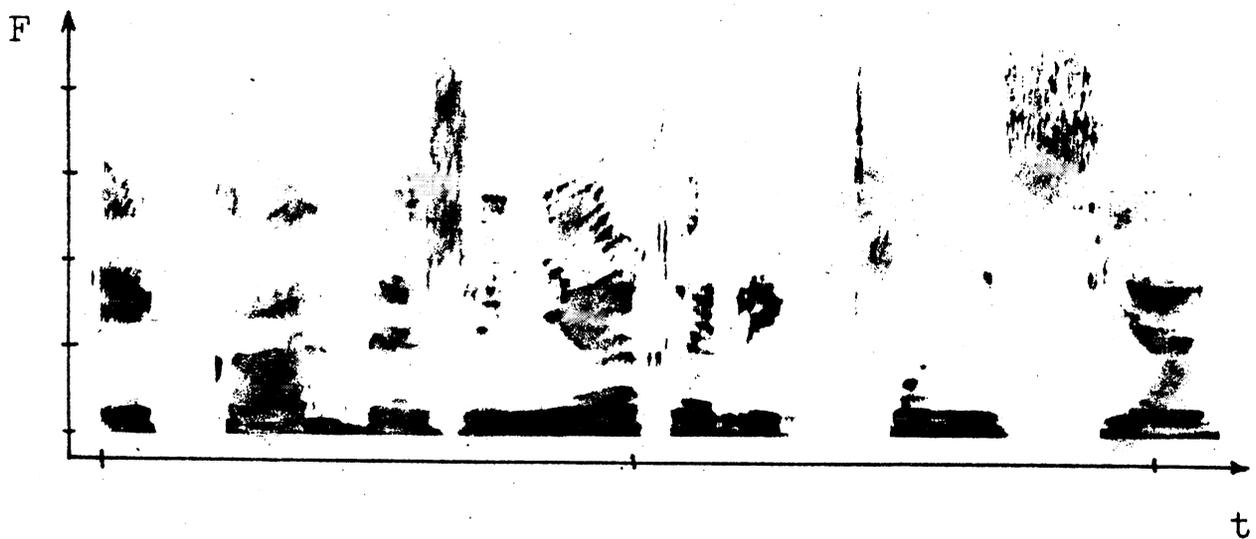
FIGURE 1

*Sonagramme d'une phrase prononcée par un homme.*



FIGURE 2

*Sonagramme d'une phrase prononcée par une femme.*



La détection des formants est également compliquée par le fait qu'il y a presque toujours une légère nasalisation lors des phrases voisées. La parole d'un locuteur souffrant d'une obstruction, même partielle, du conduit nasal est particulièrement marquée par l'absence de nasalisation. On dit, d'ailleurs à tort, de ce locuteur qu'il "parle du nez". La nasalisation complique le spectre du signal vocal par un apport de zéros et de pôles. Par ailleurs, le spectre de la source évolue.

Pour ces raisons, les maximums du spectre ne correspondent pas toujours aux formants. Nous avons eu la possibilité de prendre la mesure des difficultés que présente la détection des paramètres, à l'occasion d'une étude sur la détection automatique des formants (*R. CARRE - 1966*). Il s'agit d'un problème très important mais très difficile à résoudre auquel se sont intéressés divers auteurs (*J.L. FLANAGAN - 1956 ; A.P. PAUL, A.S. House, K.N. STEVENS - 1964 ; M.R. SCHROEDER, A.M. NOLL - 1965*).

L'objet de notre étude n'est pas de rechercher des méthodes sûres de détection de ces paramètres, mais d'approfondir les relations qui existent entre les formants et notre appareil vocal, d'effectuer quelques études statistiques sur ces paramètres et d'étudier leur rôle sur le plan de la perception.

Dans un premier chapitre, nous préciserons quelques données statistiques sur les formants.

Le deuxième chapitre sera plus particulièrement consacré à l'étude des relations existantes entre les formants et notre appareil vocal. Une comparaison entre les voix d'homme et de femme sera effectuée.

Dans le troisième chapitre, nous préciserons, à l'aide de résultats d'analyse, le rôle et l'importance des formants sur le plan de l'identification du locuteur et de la reconnaissance du message.

Enfin, le quatrième chapitre sera plus particulièrement consacré au rôle des transitions d'amplitude pour faciliter la reconnaissance de voyelles dont le spectre est "pauvre" en harmoniques.

Dans le texte, nous utilisons des symboles phonétiques extraits de l'Alphabet de l'Association Phonétique Internationale (I.P.A.). Un tableau de correspondance de ces symboles est rappelé à l'Annexe II.



## - CHAPITRE 1 -

## ETUDE STATISTIQUE SUR LES FORMANTS VOCAUX

## 1.1. INTRODUCTION

L'objet de cette étude statistique est de préciser diverses caractéristiques relatives aux trois premiers formants vocaux, la connaissance de celles-ci pouvant, en particulier, faciliter l'analyse en temps réel du signal de parole et la reconnaissance de la parole.

La détection automatique des formants à partir du spectre de la parole n'est pas simple. Or, pour commander un synthétiseur à formants, pour étudier les problèmes de synthèse par règles, pour effectuer des études de reconnaissance de parole à partir des formants, il est nécessaire de pouvoir disposer rapidement, autant que possible en temps réel, de données sur les fréquences des formants.

Plusieurs procédés de détection automatique de ces paramètres ont été expérimentés par divers chercheurs :

calcul des moments spectraux (*J. SUZUKI, Y. KADOKAWA, K. NAKATA - 1963*),  
détection des passages par zéro du signal de parole (*S.H. CHANG - 1956*),  
détection des maximums du spectre de la parole (*J.L.FLANAGAN-1956;R.CARRÉ-1966*)  
étude du spectre (*M.R. SCHROEDER, A.M. NOLL - 1965*).

Aucun de ces procédés n'est parfait. Les principales causes de mauvais fonctionnement sont les suivantes :

- la structure formantique n'est pas toujours bien définie ;
- dans les sons nasalisés, des formants particuliers apparaissent et s'intercalent entre les formants vocaux ;
- les plages d'évolution des formants se chevauchent.

Aussi, pour concevoir un détecteur automatique de formants, est-il indispensable d'étudier préalablement les évolutions de ceux-ci dans le but de déterminer les règles qui régiront le fonctionnement du détecteur.

Nous avons étudié, à partir de séquences de parole et pour plusieurs locuteurs, les limites des plages des formants, les distances entre formants et les vitesses d'évolution de ceux-ci (R. CARRE - 1968).

Enfin, nous avons examiné s'il n'existait pas, globalement, certains liens entre le formant  $F_3$  et les formants  $F_1$  et  $F_2$ .

Les études des évolutions des formants ont été faites à partir d'analyses spectrographiques effectuées à l'aide d'un "Sonagraphe". Toutes les précautions ont été prises pour obtenir des résultats exploitables (étalonnages fréquents, exécutions de sections dans tous les cas d'interprétation délicate...) car il est bien connu que la précision de telles analyses est très variable et dépend, en particulier, de la fréquence fondamentale.

Nous avons concrétisé nos résultats par des courbes correspondant à des fréquences multiples de 200 Hz.

Par ailleurs, afin d'obtenir des résultats significatifs, nous avons déterminé expérimentalement la longueur minimale de l'échantillon à analyser, longueur à partir de laquelle les résultats obtenus restent sensiblement invariants.

La figure 3 (page 9) représente la répartition R des fréquences du deuxième formant, pour un locuteur donné, obtenue à partir des dépouillements de 5, 10, 15, 20 et 25 sonagrammes correspondant respectivement à des séquences analysées de durées 6,85 , 19,35 , 26,30 et 33,60 secondes.

Nous pouvons constater que 20 à 25 sonagrammes sont suffisants.

Il y a lieu de noter que les résultats obtenus dépendent du contenu du texte prononcé comme le révèle, en particulier, la figure 4 (page 10). Cette figure représente deux courbes de répartition de la fréquence du deuxième formant obtenues pour un locuteur donné prononçant deux textes différents. Les différences observées ne sont pas surprenantes car on peut, en effet, supposer que la loi d'apparition des unités phonétiques dépend des mots utilisés et, par suite, du sujet traité.

Malgré ces différences, la plage de variation du deuxième formant est la même dans les deux cas.

FIGURE 3

Répartition des fréquences du deuxième formant à partir de 5, 10, 15, 20 et 25 sonagrammes.

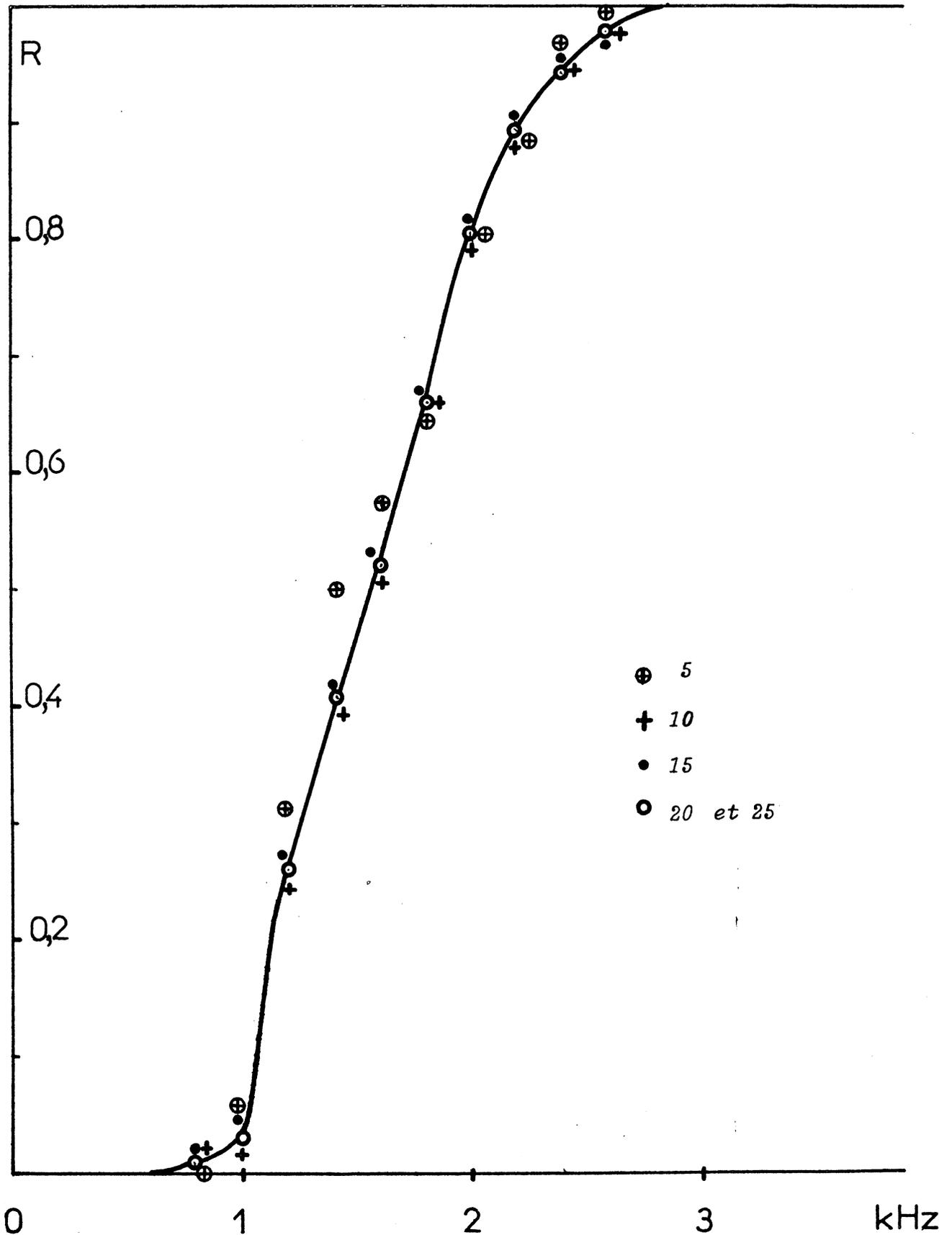
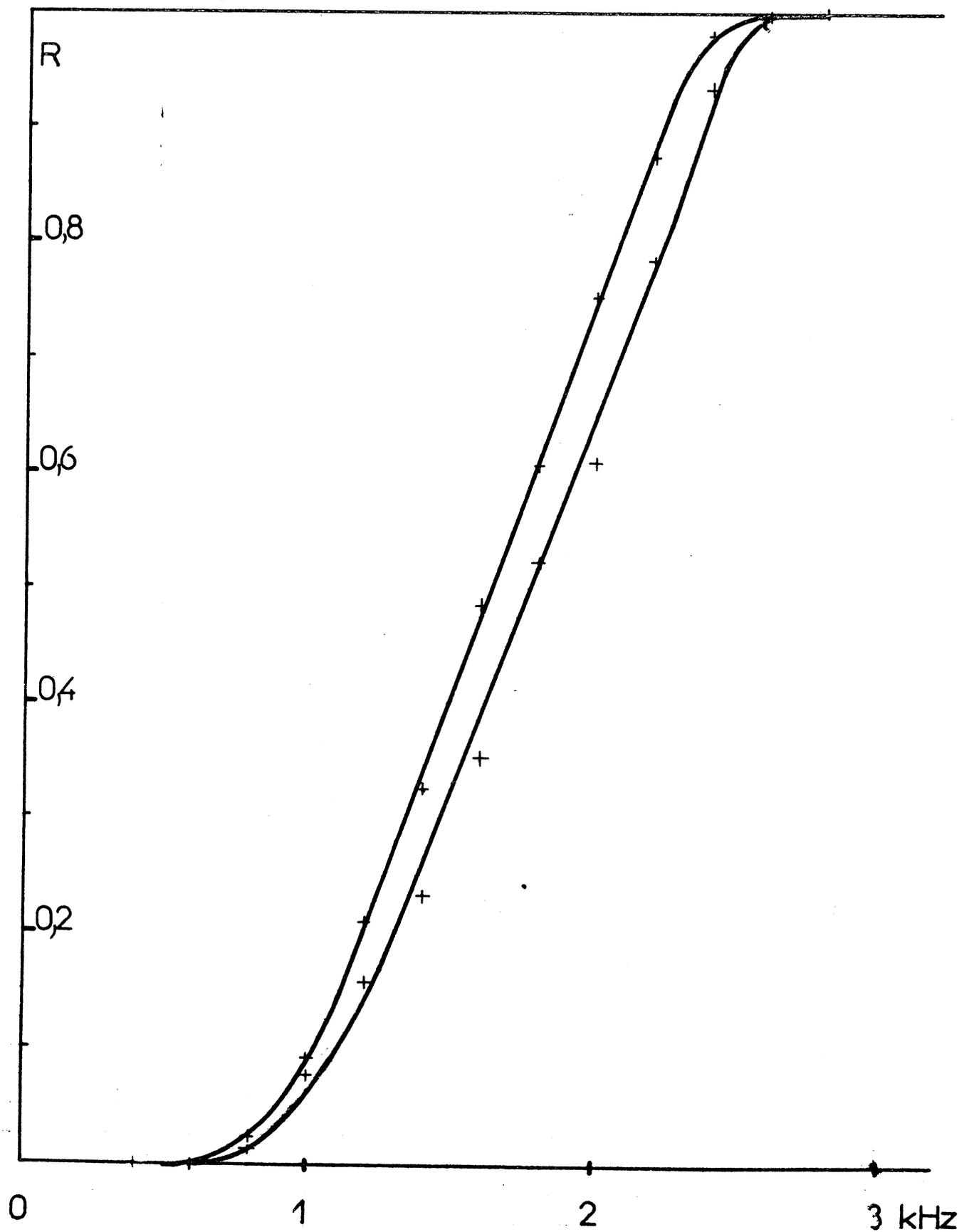


FIGURE 4

Répartition des fréquences du deuxième formant dans le cas de deux textes différents prononcés par un même locuteur.



## 1.2. PLAGES DES FORMANTS

Nous savons que les fréquences des trois premiers formants vocaux évoluent à l'intérieur de plages de fréquences bien déterminées.

FANT (1960) retient les valeurs limites suivantes pour les voix masculines  
 $F_1 : 150 - 850 \text{ Hz}$  ;  $F_2 : 500 - 2\,500 \text{ Hz}$  ;  $F_3 : 1\,500 - 3\,500 \text{ Hz}$  .  
 Il admet que ces valeurs doivent être majorées de 17 % environ pour les voix féminines.

Comme on peut le remarquer, des chevauchements de plages importants existent d'où peuvent résulter des anomalies de fonctionnement d'un détecteur automatique de formants. On admettra, sans difficulté, que la fréquence de ces anomalies est d'autant plus faible que ces chevauchements sont plus réduits. C'est pour cette raison, en particulier, que nous avons envisagé de préciser, pour un locuteur donné : - l'importance des chevauchements des plages de formants ;

- l'erreur découlant d'une réduction des plages.

Les répartitions des fréquences des formants ont été déterminées, pour quatre locuteurs (deux hommes "a" et "b", deux femmes "c" et "d"), à partir du dépouillement de 25 sonagrammes par locuteur.

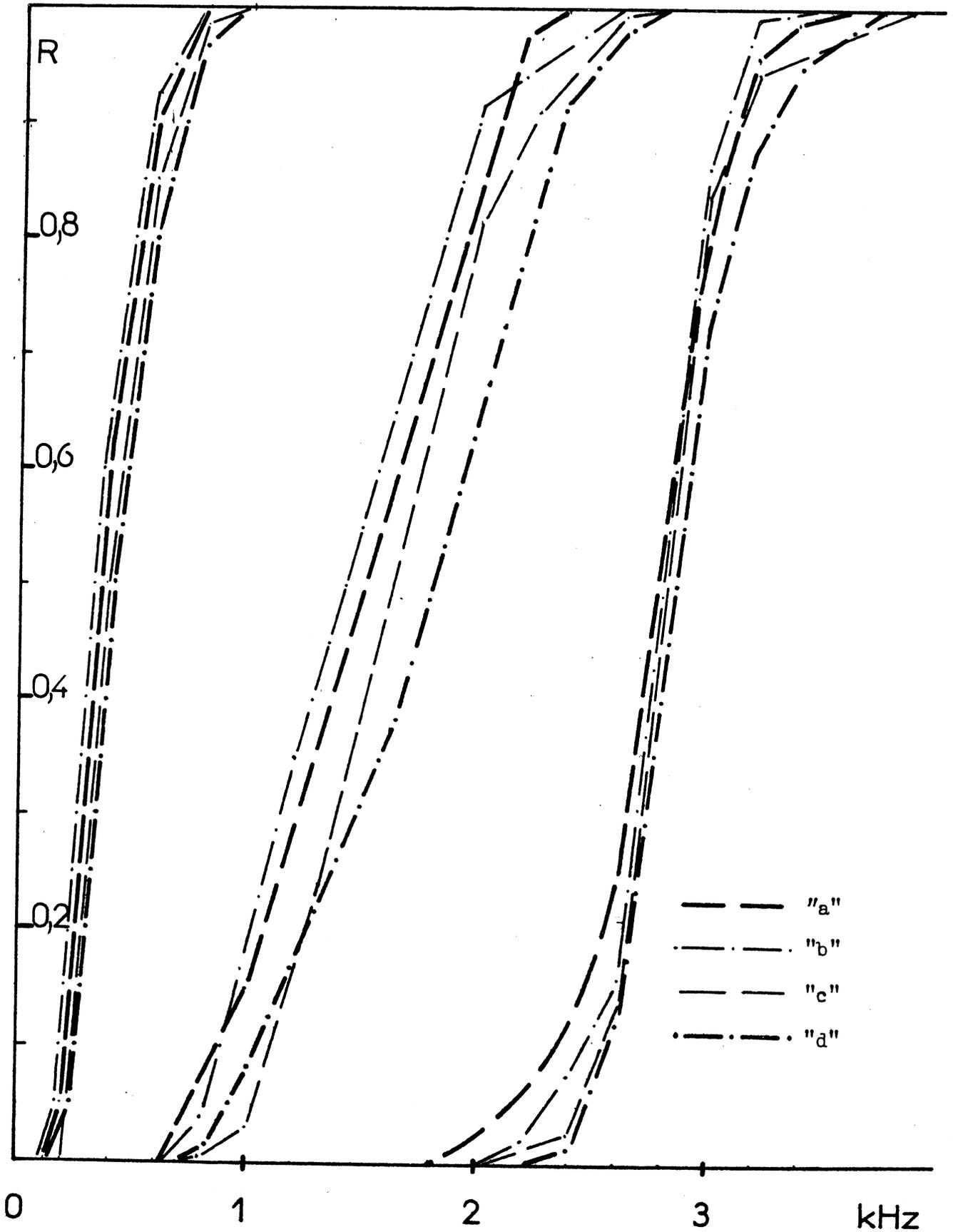
Nos résultats sont représentés figure 5 (page 12), laquelle doit être interprétée comme suit : la fréquence des formants est notée en abscisse et la durée totale, réduite à l'unité, des séquences voisées analysées pour l'ensemble des 25 sonagrammes, est portée en ordonnée. Tel point d'une courbe relative à un formant donné, d'abscisse  $F$ , a pour ordonnée  $R < 1$ , représentant alors la fraction de la durée totale durant laquelle le formant considéré possède une fréquence inférieure à  $F$ . Donnons un exemple : pour le locuteur b, le deuxième formant est, en moyenne, inférieur à 1 400 Hz pendant 50 % de la durée de la séquence analysée.

Les limites de plages, pour les divers locuteurs, sont, comme le montre la figure 5, les suivantes :

• locuteur "a" $F_1 : \sim 150 - 800 \text{ Hz}$	• locuteur "c" $F_1 : 200 - 1\,000 \text{ Hz}$
$F_2 : 600 - 2\,400 \text{ Hz}$	$F_2 : 600 - 2\,500 \text{ Hz}$
$F_3 : 1\,800 - 3\,600 \text{ Hz}$	$F_3 : 1\,800 - 4\,000 \text{ Hz}$
• locuteur "b" $F_1 : \sim 150 - 800 \text{ Hz}$	• locuteur "d" $F_1 : \sim 150 - 1\,000 \text{ Hz}$
$F_2 : 600 - 2\,600 \text{ Hz}$	$F_2 : 600 - 2\,800 \text{ Hz}$
$F_3 : 2\,000 - 3\,400 \text{ Hz}$	$F_3 : 2\,200 - 3\,800 \text{ Hz}$ .

FIGURE 5

Répartition des fréquences des trois premiers formants dans le cas de quatre locuteurs (deux hommes et deux femmes).



Nous constatons que les plages sont moins étendues pour un quelconque de nos locuteurs que celles, moyennes, indiquées par *FANT*.

Par ailleurs, à l'aide de ces courbes, nous pouvons estimer l'erreur découlant d'une réduction des plages.

Cette réduction est souhaitable car, si elle pouvait effectivement être mise en oeuvre, elle permettrait, grâce à la diminution du chevauchement des plages qui en résulterait, une simplification d'un système de détection automatique des formants.

A titre indicatif, exprimons l'erreur découlant de la suppression du chevauchement des formants  $F_1$  et  $F_2$  pour le locuteur "c".

Nous pouvons, par exemple, adopter comme frontière commune de ces deux formants, la fréquence 800 Hz. De ce choix, il résulte, comme le montrent nos courbes, que, pendant 1 % environ de la durée d'une séquence, les fréquences de  $F_1$  et  $F_2$  seront définies avec une erreur maximale de l'ordre de 200 Hz.

Encore faut-il préciser à quoi correspond, sur le plan de la perception, une telle erreur !

Enfin nous pouvons également remarquer, sur la figure 5, que les courbes de répartition correspondant aux voix féminines sont plus élevées dans l'échelle des fréquences que celles relatives aux voix masculines.

Nous avons calculé, à partir des fréquences moyennes des formants homologues d'un locuteur féminin et d'un locuteur masculin ( $\overline{F}_f$  et  $\overline{F}_m$ ), le coefficient  $k$ , écart relatif entre fréquences moyennes :

$$k = \left( 1 - \frac{\overline{F}_f}{\overline{F}_m} \right) .$$

Rappelons que les fréquences moyennes  $\overline{F}_f$  et  $\overline{F}_m$  peuvent être calculées à partir de la fonction de répartition  $R$  :

$$\overline{F} = \int_0^1 f \, dR .$$

Pour les locuteurs "a" et "d", nous avons trouvé :

pour $F_1$ :	$k_1 = 13,5 \%$
pour $F_2$ :	$k_2 = 17,5 \%$
pour $F_3$ :	$k_3 = 5,0 \%$

Il est intéressant de rapprocher ces résultats de ceux obtenus par *FANT* (1959)

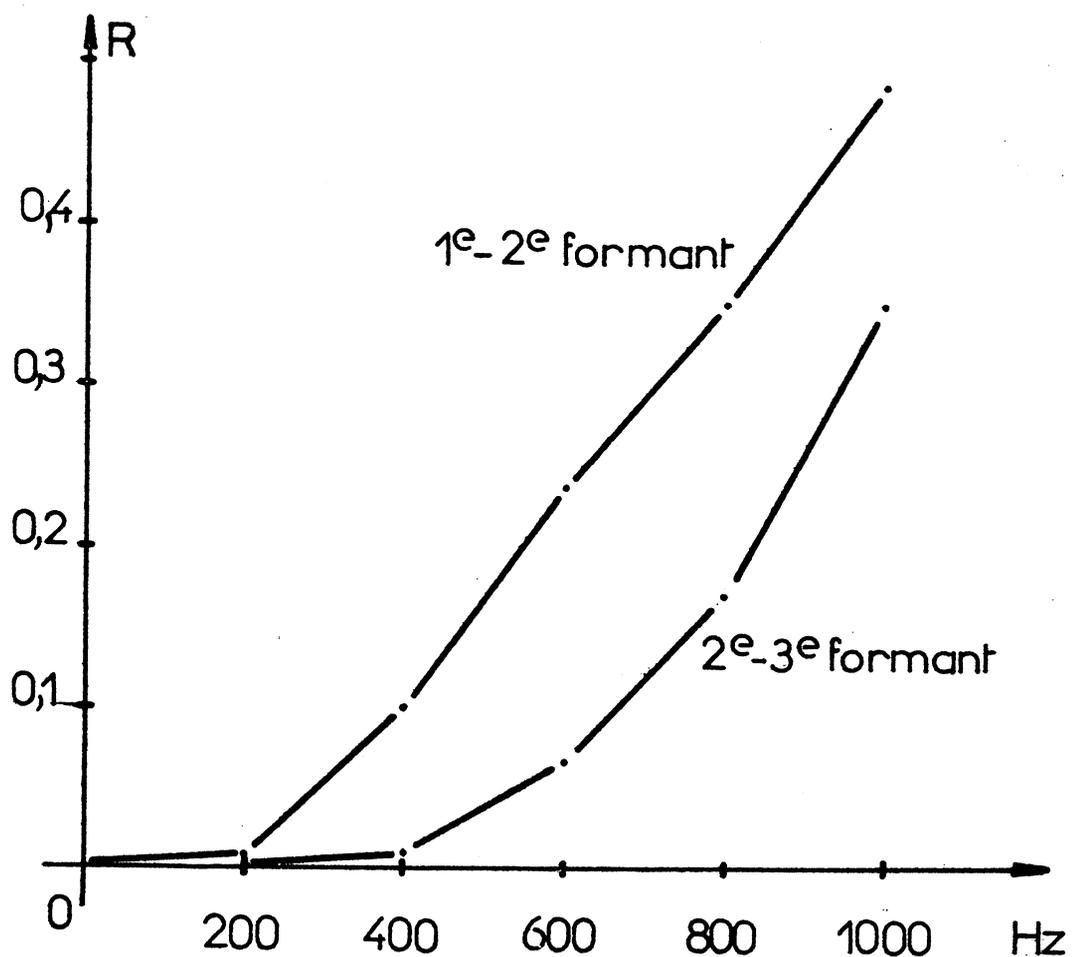
pour les voyelles prononcées par des Américains  $k_1 = 15,5\%$ ,  $k_2 = 17 \%$ ,  $k_3 = 17 \%$ ;  
 .par des Suédois :  $k_1 = 13,5\%$ ,  $k_2 = 13,5\%$ ,  $k_3 = 17,5\%$ .

## 1.3. DISTANCES ENTRE FORMANTS VOISINS

A partir des sonagrammes exploités pour l'étude des plages, nous avons déterminé (figure 6) la fonction de répartition des écarts en fréquence, à un instant donné, de deux formants voisins. Le dépouillement de nos résultats a montré que cette fonction ne différait pas sensiblement d'un locuteur à un autre.

FIGURE 6

Répartition des écarts entre deux formants voisins.



Nous voyons, par exemple, qu'un écart de 200 Hz est très rarement rencontré, qu'un écart compris entre 200 et 400 Hz entre les deux premiers formants est manifesté durant 9 % de la durée de la séquence de parole.

On peut donc, dans un détecteur automatique de formants, après la détection du premier formant, différer de 200 Hz la possibilité de détection du deuxième formant sans commettre d'erreur appréciable (alors que si l'on ne tient pas compte de cette règle, la composante voisine de l'harmonique représentatif du premier formant pourrait être détectée).

## 1.4. VITESSES DE VARIATION DES FORMANTS

Toujours à l'aide des sonagrammes précédents, nous avons étudié la vitesse d'évolution des formants. Nous savons que celle-ci est limitée.

Selon un procédé de détection automatique (*R. CARRE - 1966*), les formants sont extraits du spectre de la parole, lequel est obtenu au moyen d'un analyseur spectral séquentiel à fréquence de répétition 50 Hz.

Les informations concernant les formants résultent donc d'un échantillonnage de période 20 ms.

La vitesse de variation des formants étant limitée, nous poserons que, pendant l'intervalle de 20 ms, ces fréquences ne peuvent varier de plus de  $\Delta F$ . De ceci il résulte que, d'une séquence d'analyse à la suivante, la fréquence d'un formant demeure à l'intérieur d'une plage de fréquences s'étendant de  $F - \Delta F$  à  $F + \Delta F$ , dont la connaissance est d'un grand intérêt pour la conception d'un détecteur automatique de formants.

La figure 7 (page 16) concrétise, sous forme de courbes, les résultats obtenus :

en abscisse, la variation  $\Delta F$  de fréquence durant l'intervalle de 20 ms, en ordonnée, la répartition exprimée en fraction de la durée totale de la séquence analysée,

(dans le cas présent, les courbes obtenues pour les quatre locuteurs sont pratiquement confondues).

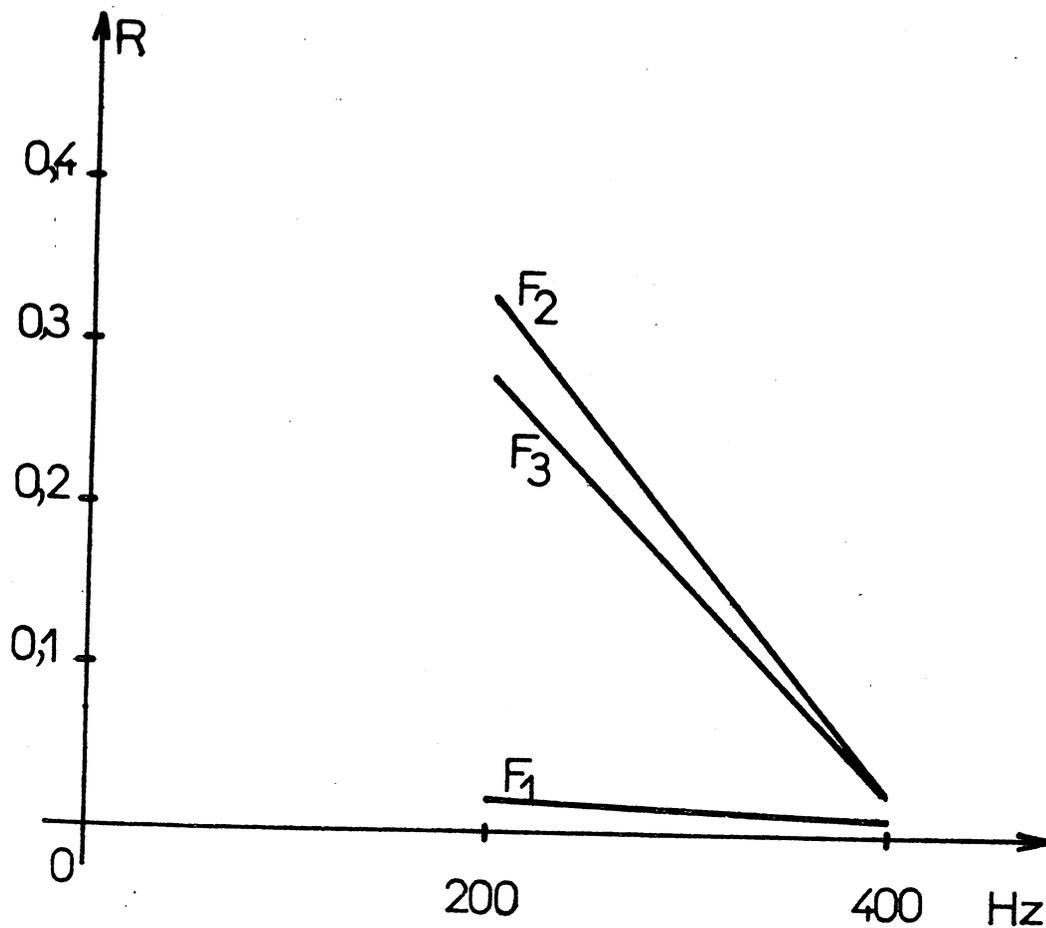
Ces courbes montrent, par exemple, que, pendant la fraction 0,99 de la durée totale (soit 99 %), les trois formants auront varié de moins de 400 Hz en 20 ms.

Les résultats que nous avons obtenus peuvent être rapprochés des résultats publiés par *T. GAY (1968)*.

Ce dernier note une variation maximale de 10 Hz/ms pour le deuxième formant dans le cas des diphtongues.

FIGURE 7

Répartition des vitesses de variation des formants.



### 1.5. RELATIONS ENTRE LE TROISIEME ET LES DEUX PREMIERS FORMANTS

Pour étudier les évolutions de  $F_3$  en fonction de  $F_1$  et  $F_2$ , nous avons représenté (figures 8.a, b, c, d ; pages 17 à 20), toujours à partir des sonagrammes précédents, des régions dans le plan  $F_1 - F_2$  correspondant à quatre plages de fréquences du formant  $F_3$ . Ces régions sont très difficiles à préciser car, dans la réalité, elles empiètent les unes sur les autres. Mais, malgré ces empiètements, nous pouvons noter des tendances.

On peut donc penser qu'il existe certains liens entre  $F_1$ ,  $F_2$  et  $F_3$ , et que ceux-ci dépendent, comme le laissent paraître les figures, du locuteur considéré.

FIGURE 8.a

Evolution de  $F_3$  en fonction de  $F_1$  et  $F_2$  - Locuteur "a".

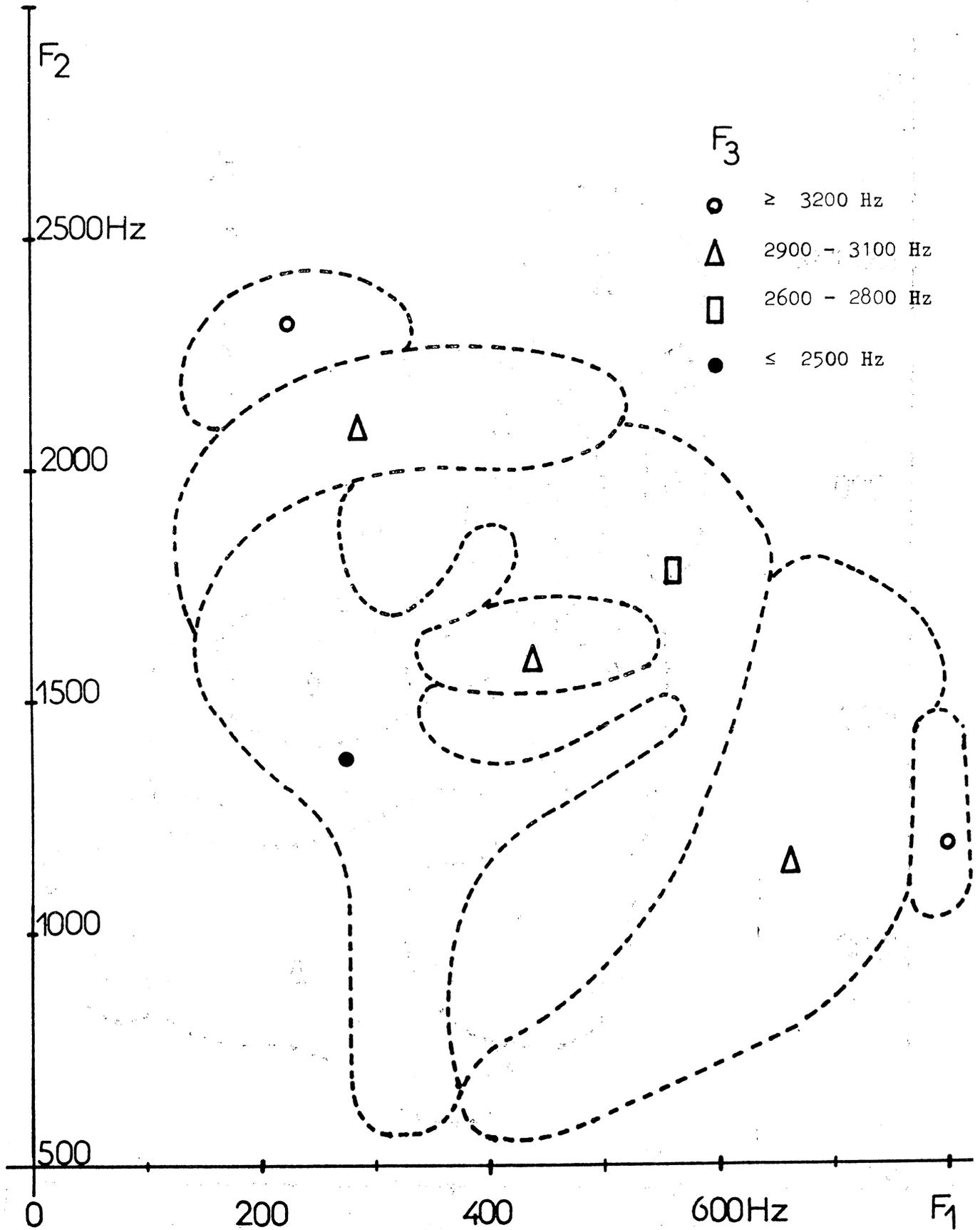


FIGURE 8.b

Evolution de  $F_3$  en fonction de  $F_1$  et  $F_2$  - Locuteur "b".

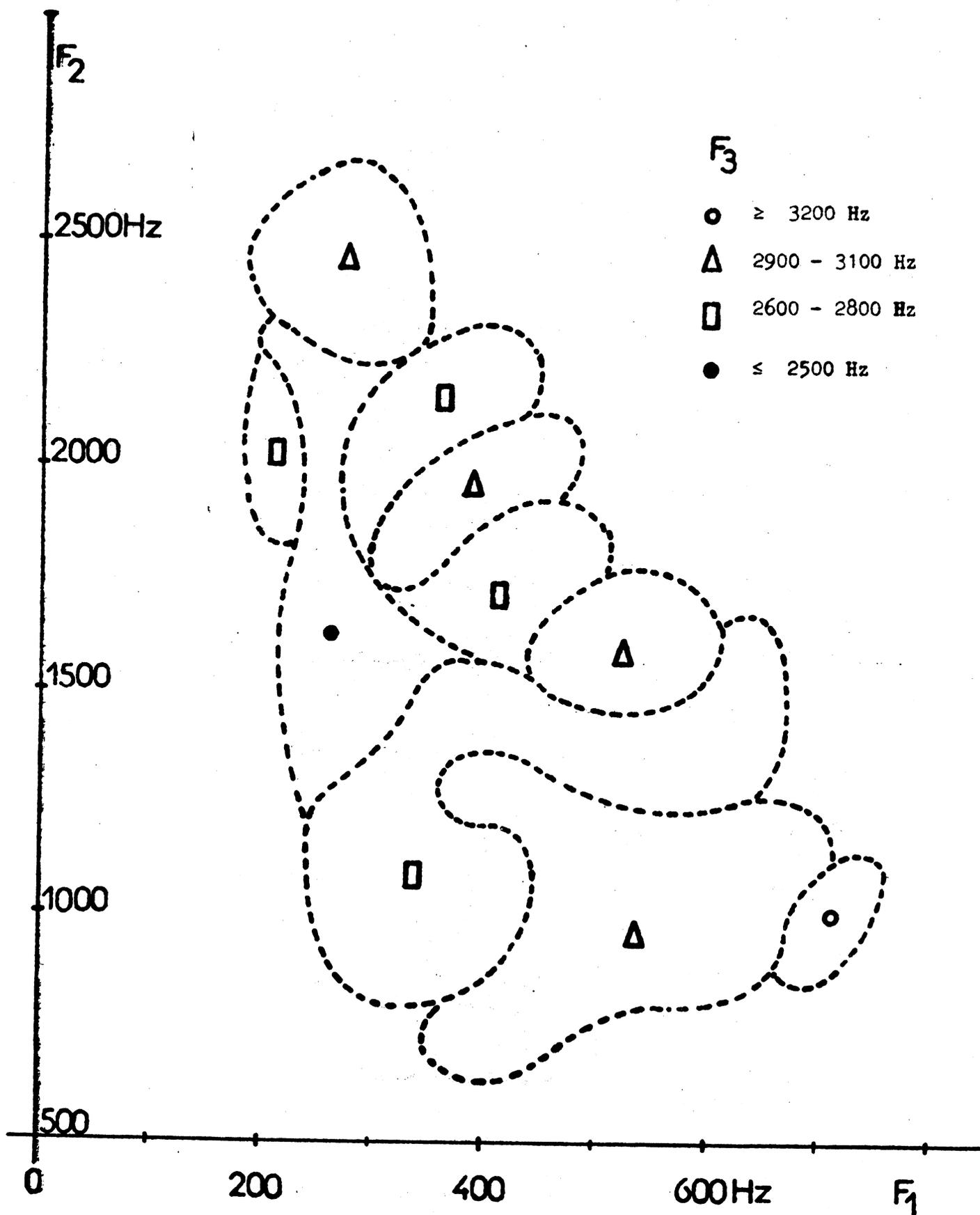


FIGURE 8.c

Evolution de  $F_3$  en fonction de  $F_1$  et  $F_2$  - Locuteur "c".

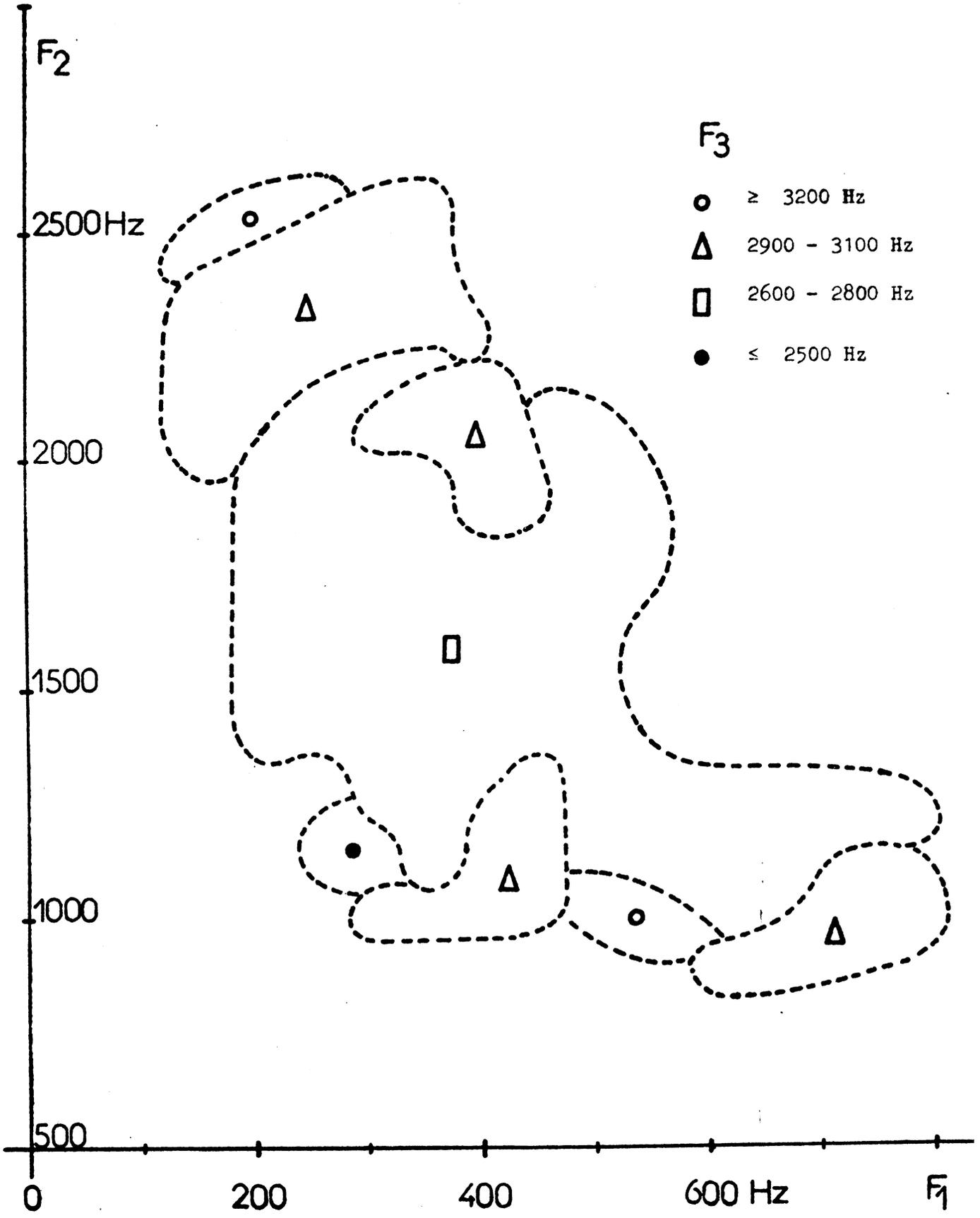
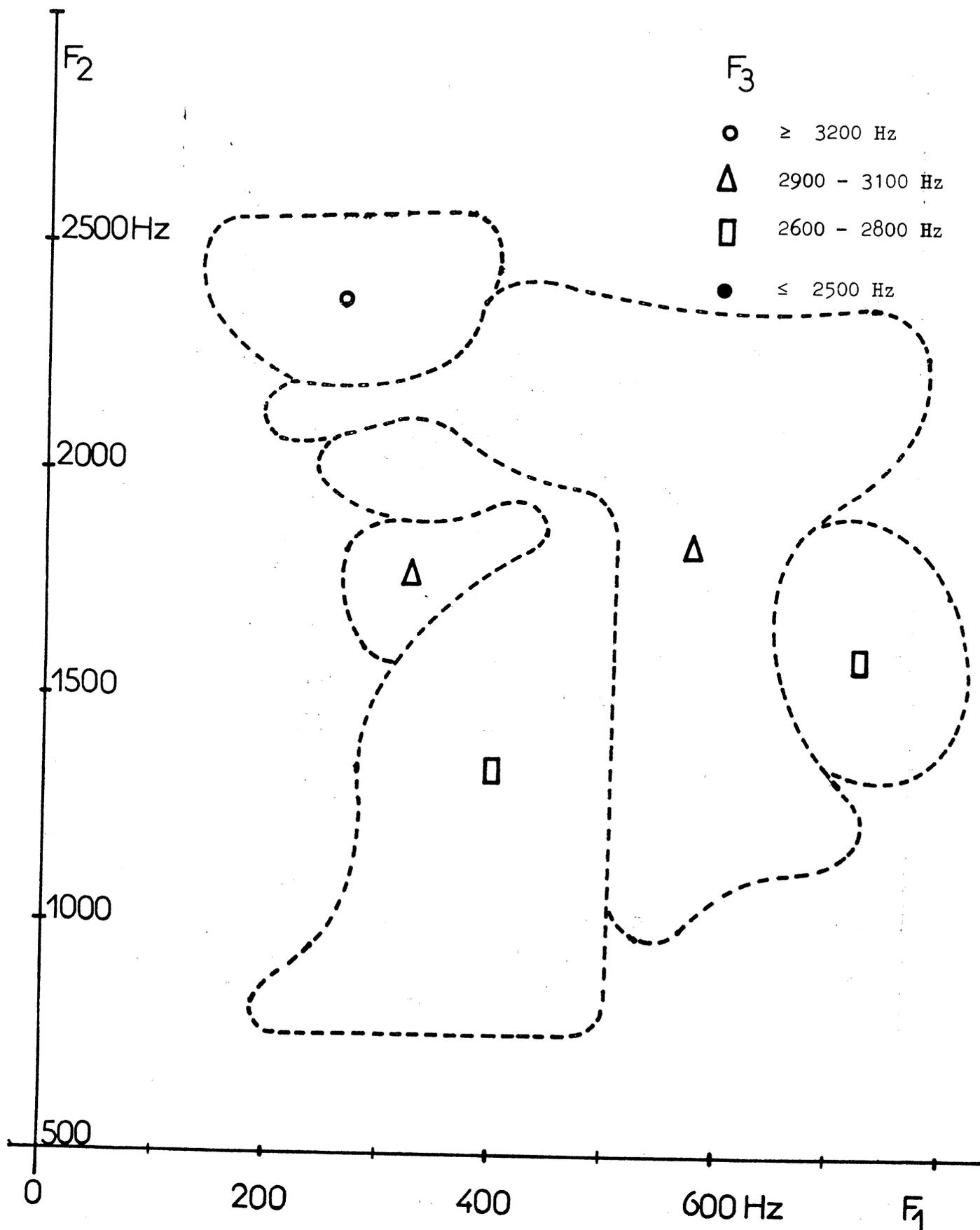


FIGURE 8.d

Evolution de  $F_3$  en fonction de  $F_1$  et  $F_2$  - Locuteur "d".



Il est certain qu'une étude plus approfondie devrait être entreprise en vue de préciser la valeur de  $F_3$  en fonction du sens des variations de  $F_1$  et  $F_2$ . Les régions obtenues seraient alors plus complexes mais plus significatives.

Les résultats obtenus pourraient être vérifiés par synthèse. On sait que le troisième formant n'est pas un facteur important sur le plan de l'intelligibilité mais qu'il contribue à la qualité de la parole de synthèse.

Il faudrait étudier si une synthèse effectuée, entre autres, à partir d'un formant  $F_3$  déterminé automatiquement à partir de  $F_1$  et  $F_2$ , est de meilleure qualité qu'une synthèse effectuée à partir de  $F_3$  maintenu constant et si, par ce type de commande, selon les liens entre  $F_3$  et  $F_1$ ,  $F_2$  adoptés, on peut modifier les "caractéristiques individuelles" de la parole synthétisée.

## 1.6. CONCLUSIONS

Les caractéristiques que nous avons précisées, sont d'un grand intérêt pour l'élaboration d'un détecteur automatique de formants devant présenter une bonne sécurité de fonctionnement.

Par ailleurs, l'existence de liens entre  $F_1$  et  $F_2$ , d'une part, et  $F_3$  d'autre part, nous conduit à penser que le troisième formant pourrait être calculé au moyen d'une expression mathématique, fonction du locuteur.

Nous étudierons, sous d'autres aspects (chapitre 4) les évolutions de  $F_3$  en vue de l'identification des locuteurs.

Donc, dans le cas d'un analyseur-synthétiseur à bande réduite, du type vocoder à formants et si nous désirons une parole de synthèse uniquement intelligible, le nombre de paramètres à transmettre pourra être réduit ( $F_3$  non transmis par exemple) ; si nous désirons une parole de synthèse naturelle et individualisée, un nombre de paramètres plus grand sera nécessaire. Mais, les caractéristiques individuelles étant réputées constantes pour un locuteur donné, on peut envisager de transmettre avant toute communication des données caractéristiques du locuteur telles que : limites de

plages, relation entre  $F_1$ ,  $F_2$  et  $F_3$ , caractéristiques de la source vocale, etc. Ensuite, seuls les paramètres d'intelligibilité seraient transmis.

On pourrait aussi envisager une adaptation automatique de l'analyseur au locuteur au cours des premières séquences de parole pendant lesquelles un nombre important de paramètres serait traité puis, après détermination automatique des caractéristiques individuelles, le nombre de paramètres serait réduit à ceux reconnus indispensables pour une bonne intelligibilité.

On pourrait encore imaginer des synthèses de parole caractéristiques d'un nombre limité de locuteurs. En télécommunications, chaque locuteur aurait un numéro de code choisi au mieux parmi ceux correspondant aux divers types de voix pouvant être synthétisés par l'appareil.

- CHAPITRE 2 -  
ETUDE DES FORMANTS  
DANS LE CAS DE LA PRODUCTION DES VOYELLES  
PAR DES LOCUTEURS HOMMES ET FEMMES

### 2.1. INTRODUCTION

L'étude statistique précédente nous a montré qu'il existait des différences sensibles au niveau des courbes de répartition des formants, ces différences étant importantes si nous comparons entre eux les résultats obtenus à partir des voix d'hommes et de femmes.

Il nous est apparu important d'essayer de préciser l'origine de ces différences en étudiant plus particulièrement les résultats d'analyses effectuées sur des voyelles d'une part, les phénomènes qui entrent en jeu lors de la production de celles-ci d'autre part.

Parmi les travaux déjà effectués dans ce domaine, notons, en particulier, les études sur les voyelles anglaises (*G.E. PETERSON* et *H.L. BARNEY - 1952*) et suédoises (*G. FANT - 1959*).

De ces études, on a pu tirer de nombreuses conclusions. C'est ainsi que, pour un même locuteur prononçant plusieurs fois une même voyelle, les fréquences des formants de cette voyelle se dispersent d'une manière sensible ; pour plusieurs locuteurs provenant d'une même région linguistique, la dispersion augmente. Plusieurs auditeurs peuvent percevoir différemment une même voyelle isolée prononcée par un même locuteur. Un même auditeur peut percevoir comme étant différents deux sons physiquement identiques. Le contexte joue, les origines du locuteur et de l'auditeur doivent être prises en considération et, au cours de tests de perception, des processus d'apprentissage interviennent.

Les résultats que l'on peut donner sur les fréquences des formants sont donc des moyennes avec des dispersions plus ou moins importantes.

Par ailleurs -fait important- des écarts notables apparaissent entre les fréquences des formants de voyelles émises par des hommes d'une part, par des femmes d'autre part (ou par des enfants), voyelles perçues comme étant de même voyelles. A ce sujet, rappelons les écarts relatifs, donnés au chapitre précédent, déterminés par *FANT* :

$$\text{pour le premier formant : } k_1 = \left( \frac{F_{1f}}{F_{1m}} - 1 \right) = 13,5 \% ,$$

$$\text{pour le deuxième formant : } k_2 = \left( \frac{F_{2f}}{F_{2m}} - 1 \right) = 13,5 \% ,$$

$$\text{pour le troisième formant : } k_3 = \left( \frac{F_{3f}}{F_{3m}} - 1 \right) = 17,5 \% .$$

$F_{1f}$ ,  $F_{2f}$ ,  $F_{3f}$ , représentent les fréquences moyennes des trois premiers formants pour les voyelles suédoises émises par des femmes,

$F_{1m}$ ,  $F_{2m}$ ,  $F_{3m}$ , les fréquences moyennes pour les mêmes voyelles émises par des hommes.

Ces écarts trouvent certainement leur origine dans les différences anatomiques existant entre le conduit vocal de l'homme et celui de la femme. La nature aurait pu compenser ces différences, ce qui aurait permis que, seuls, des sons physiquement identiques soient perçus comme identiques. Il n'en est donc pas ainsi.

Sur le plan dynamique, les caractéristiques physiques de la parole n'assurent qu'une approche du système de référence d'un locuteur inconnu par notre système de perception. L'aide de réflexes découlant de l'apprentissage des règles de grammaire, du vocabulaire, l'aide du sens général d'une phrase, permettront la compréhension du message parlé. Après adaptation de notre système de référence à celui du locuteur, des sons isolés pourront alors être identifiés.

Mais, sur un plan statique, sans apprentissage préalable, pour l'écoute de sons isolés, notre système de perception reprend sa référence personnelle et les caractéristiques physiques retrouvent toute leur importance comme uniques références permettant une identification. Dans ce cas, aucun contexte ne joue qui faciliterait cette identification.

Ces diverses observations conduisent à nous poser les questions suivantes :

Quels sont les paramètres nécessaires et suffisants permettant une identification des diverses voyelles d'un système (voyelles de locuteurs hommes, femmes et enfants) ; en particulier, les valeurs des deux premiers formants sont-elles suffisantes ?

Il ne le semble pas. En effet, dans les données de *PETERSON* par exemple, nous trouvons :

pour la voyelle de locuteurs hommes [ɔ] :  $F_1 = 730 \text{ Hz}$  ;  $F_2 = 1\ 090 \text{ Hz}$  ;

pour la voyelle de locuteurs enfants [o] :  $F_1 = 680 \text{ Hz}$  ;  $F_2 = 1\ 060 \text{ Hz}$  .

Ces valeurs assez voisines de  $F_1$  et  $F_2$  ne constituent donc pas un ensemble de données suffisant pour permettre la distinction entre les voyelles [ɔ] et [o]. Des couples de valeurs  $F_1$  et  $F_2$ , identiques pour des voyelles différentes, se retrouvent dans de nombreux cas, comme le montrent les chevauchements des plages de voyelles représentées dans le plan  $F_1 - F_2$  (voir, par exemple, la figure 9, page 26 - d'après *PETERSON* et *BARNEY*).

Il paraît donc nécessaire d'associer à  $F_1$  et  $F_2$ , un paramètre supplémentaire permettant, en statique, de distinguer entre elles les diverses voyelles. Le troisième formant pourrait jouer ce rôle dans certains cas, mais la fréquence fondamentale  $F_0$  qui est, en moyenne pour un homme, inférieure d'une octave à celle d'une femme, est un paramètre intéressant. Si l'on admet que les trois paramètres  $F_0$ ,  $F_1$  et  $F_2$  sont suffisants pour assurer toute distinction dans un système de voyelles, on est naturellement conduit à associer à chaque voyelle une fonction d'identification telle que, pour la voyelle [a] :

$$I_{[a]} = \Phi_{[a]}(F_0, F_1, F_2) .$$

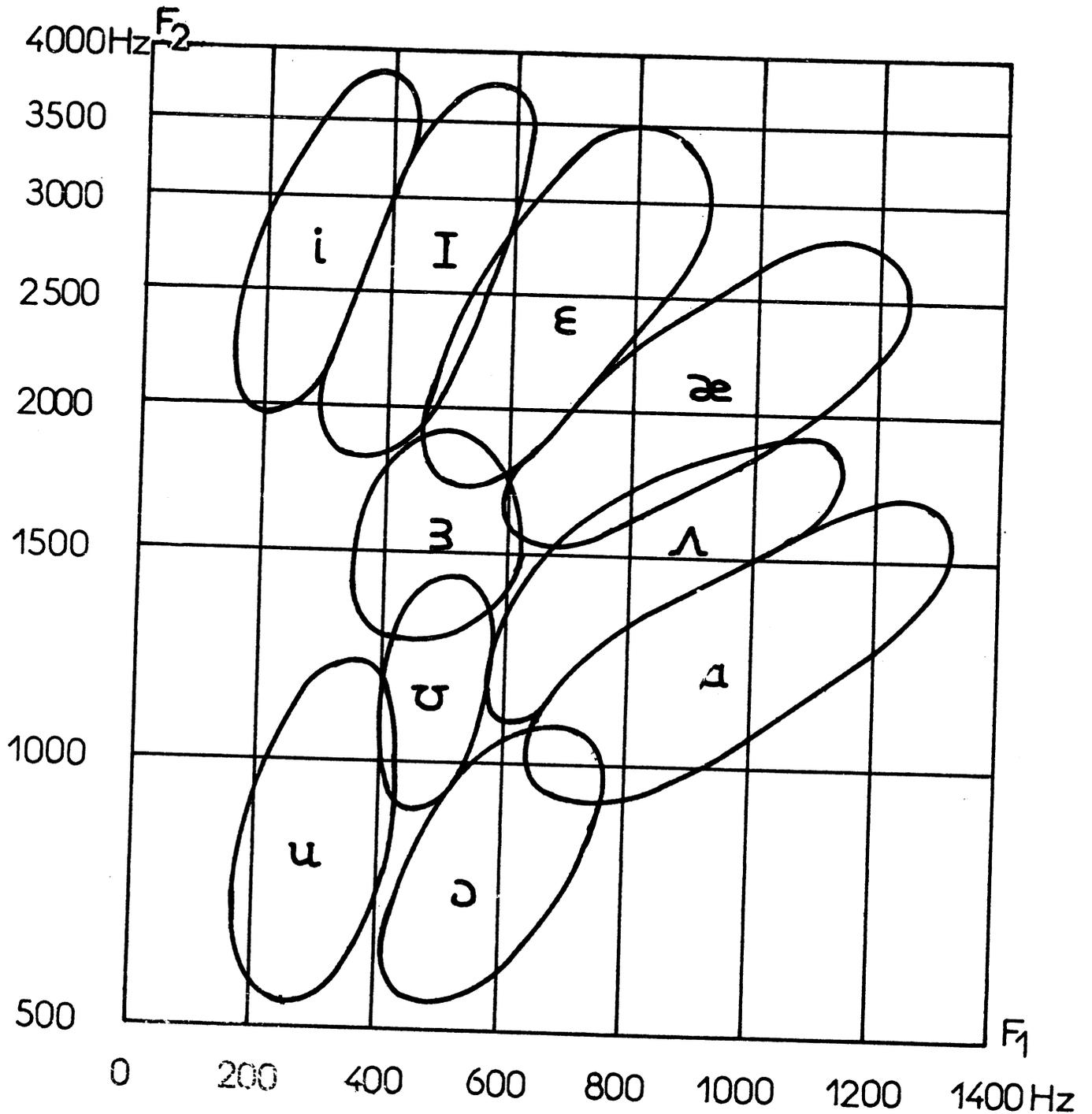
Pour chaque voyelle, cette fonction sera représentée, dans un espace à trois dimensions par un volume comprenant trois domaines principaux relatifs à la voyelle prononcée respectivement par des hommes, par des femmes, par des enfants.

Or, pour un même locuteur, la fréquence fondamentale peut varier dans de grandes proportions ; pour un homme, par exemple, elle peut atteindre la valeur moyenne admise pour les locuteurs femmes. Le son produit, au cours de cette évolution de fréquence, garde-t-il les mêmes caractéristiques du point de vue de la perception ? Si oui, la fonction d'identification  $I$  doit rester localisée à l'intérieur du volume considéré et nous devons alors poser comme hypothèse qu'une modification adéquate de la forme

et des dimensions du conduit vocal intervient, provoquant un déplacement des fréquences de résonance de ce conduit, donc une modification des fréquences des formants.

FIGURE 9

Plages de voyelles représentées dans le plan  $F_1 - F_2$ .



Nous savons que notre larynx est mobile et que son déplacement est fonction de la fréquence des cordes vocales. Ce phénomène vérifie-t-il l'hypothèse précédente ? (G.E. ARNOLD - 1957).

Pour répondre à cette interrogation, nous avons effectué des expériences à l'aide d'un appareil simulateur du conduit vocal, puis en mesurant les fréquences des formants pour une voyelle émise par un locuteur modifiant la fréquence de sa source vocale (R. CARRE, R. LANCIA, M. WAJSKOP - 1968).

## 2.2. ETUDE DE LA PRODUCTION DES VOYELLES A L'AIDE D'UN SIMULATEUR DU CONDUIT VOCAL

Nous avons enregistré les courbes de réponse en fréquence d'un simulateur du conduit vocal auquel nous faisons subir différentes modifications. Dans ce paragraphe, nous commenterons les résultats obtenus après avoir brièvement décrit notre appareil simulateur ainsi que les diverses phases de notre expérience.

### 2.2.1. SIMULATEUR VOCAL

Plusieurs travaux (H.K. DUNN - 1950 ; K.N. STEVENS, S. KASOWSKI, G. FANT - 1953 ; G. FANT - 1960) ont déjà précisé le fonctionnement d'un simulateur de conduit vocal et son intérêt pour les recherches relatives à la production de la parole.

Dans un tel appareil, le conduit vocal est simulé dans sa forme et ses dimensions.

Pour cela, le conduit vocal est découpé en tranches, centimètre par centimètre, par exemple, à partir des cordes vocales. La longueur d'un conduit étant de l'ordre de 17 cm chez un homme, on obtient ainsi 17 tubes de 1 centimètre de long et de section variable entre 0 et 20 cm<sup>2</sup> environ.

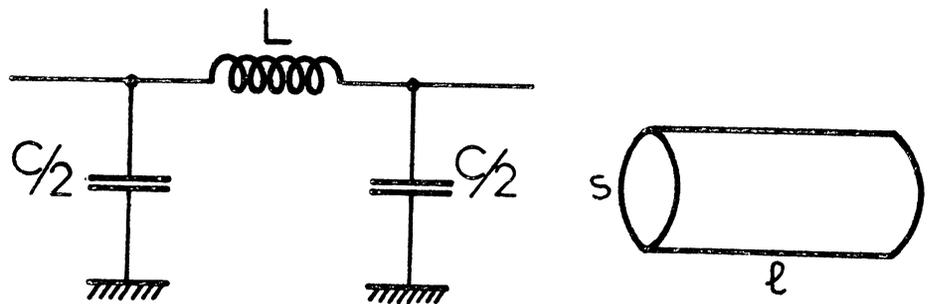
Chacun de ces tubes peut être simulé électriquement par une cellule constituée d'une inductance L et de deux condensateurs C (voir figure 10,

page 28), avec : 
$$L = \frac{\rho l}{S} \quad \text{et} \quad C = \frac{l S}{\rho c^2} ;$$

$\rho$  : densité de l'air,  $c$  : vitesse du son.

FIGURE 10

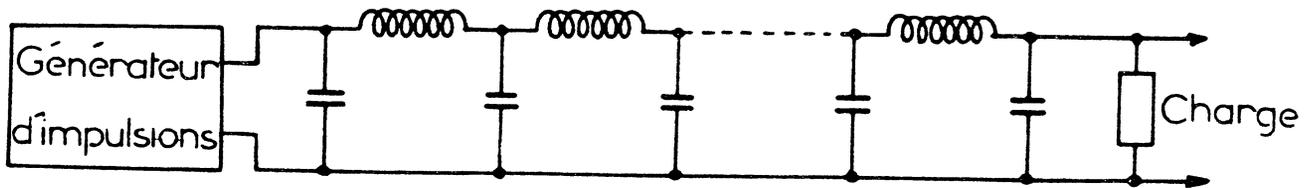
Simulation d'un tube de section  $S$  et de longueur  $l$ .



On a alors le schéma représenté ci-dessous :

FIGURE 11

Simulateur vocal



La fréquence de coupure de l'appareil dépend du nombre de cellules. Avec 20 cellules (donc une par centimètre), elle est de l'ordre de 5500 Hz, ce qui est peu si nous voulons étudier la production des sons non vocaux ; avec 40 cellules (une par demi-centimètre), elle est de 11 000 Hz.

L'appareil que nous avons réalisé (M. GAUDE, F. PULVERIC, J.Y. VALET - 1966) est constitué de 20 cellules. La fréquence de coupure de cet appareil est suffisamment élevée pour l'étude des deux premiers formants des voyelles.

Chaque cellule peut simuler une section droite de conduit vocal dont l'aire peut être réglée par bonds (10) entre  $0,16 \text{ cm}^2$  et  $16 \text{ cm}^2$ .

### 2.2.2. EXPERIENCES DE SIMULATION DU CONDUIT VOCAL

Nous avons simulé le conduit vocal masculin sous diverses formes adaptées chacune à la production de l'une des voyelles françaises suivantes :

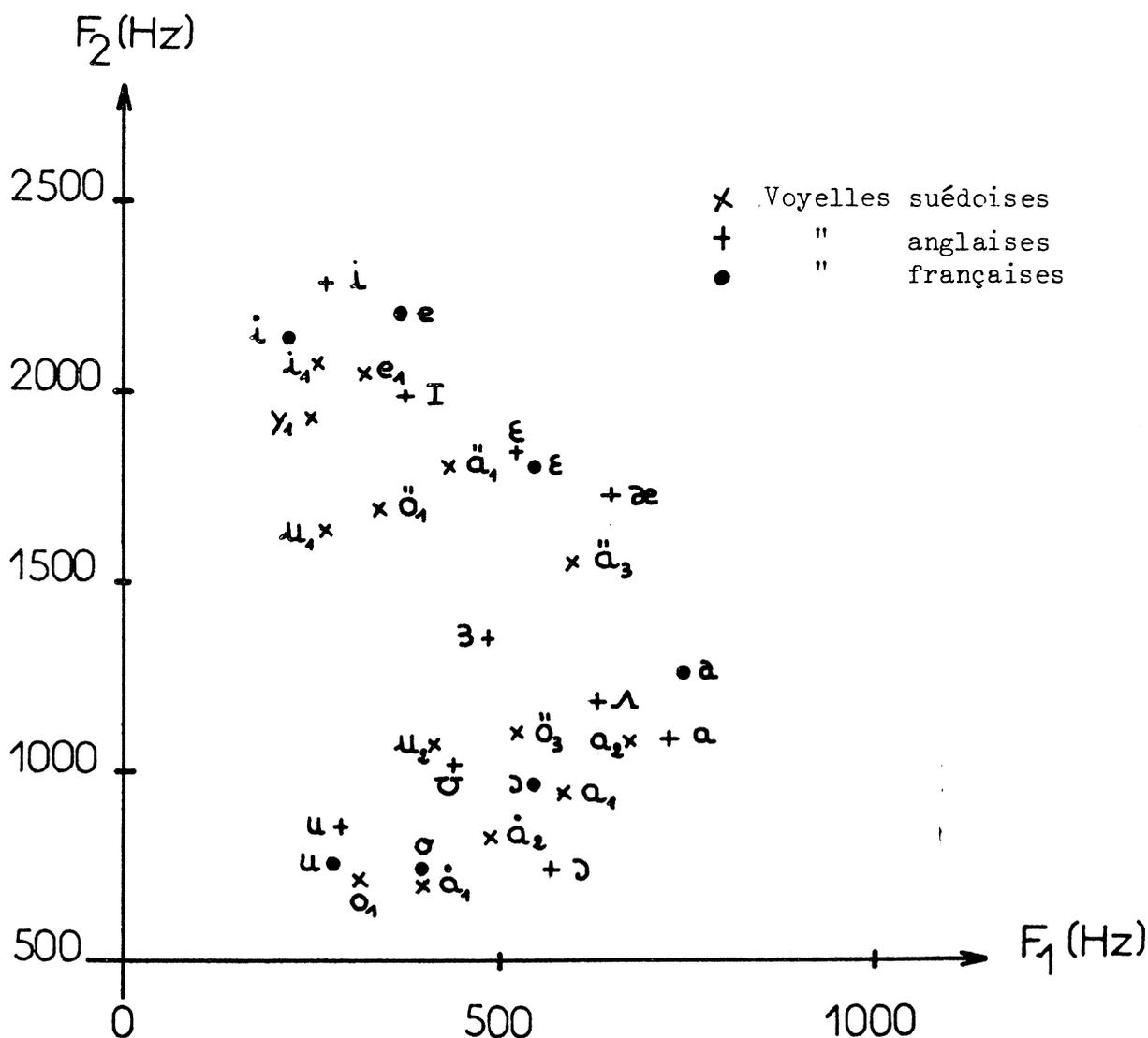
	[u]	[ɔ]	[o]	[a]	[ɛ]	[e]	[i]
F <sub>1</sub>	275 Hz	375 Hz	550 Hz	750 Hz	550 Hz	375 Hz	225 Hz
F <sub>2</sub>	750 Hz	750 Hz	950 Hz	1 250 Hz	1 800 Hz	2 200 Hz	2 125 Hz

Les fréquences des formants ont été déterminées à partir de données publiées par DELATTRE (1948) et de données qui nous furent communiquées par M. WAJSKOP.

La figure 12 représente, dans le plan  $F_1 - F_2$ , la position de ces voyelles et, à titre de comparaison, celle des voyelles anglaises de PETERSON et des voyelles suédoises de FANT.

FIGURE 12

Voyelles représentées dans le plan  $F_1 - F_2$ .



Pour reconstituer les formes appropriées du conduit vocal, nous nous sommes inspirés des travaux de CHIBA et KAJIYAMA (1941), de FANT (1960) et de STEVENS, KASOWSKI et FANT (1953).

## 2.2.3. RESULTATS

Après avoir enregistré les courbes de réponse en fréquence des différents conduits correspondant à la production des sept voyelles précédents, nous avons court-circuité deux cellules du simulateur vocal, représentant donc deux centimètres du conduit vocal, au niveau du début de la cavité du pharynx, côté larynx. Ces deux centimètres peuvent représenter d'une manière valable le déplacement du larynx d'un conduit vocal d'homme lorsque la fréquence des cordes vocales augmente d'une octave. Nous avons enregistré les nouvelles courbes de réponse et calculé le coefficient  $k$ , écart relatif entre les valeurs mesurées dans les deux cas pour les deux premières résonances ( $F_1$  et  $F_2$ ).

L'évolution du coefficient  $k_1$  (pour  $F_1$ ) est représentée figure 13 (page 31) et celle du coefficient  $k_2$  (pour  $F_2$ ) par la figure 14 (page 32). Nous avons également représenté les résultats obtenus par *FANT* (1959) et *PETERSON* (1952) à partir d'analyses de voyelles naturelles prononcées par des hommes et par des femmes.

Pour le classement en abscisse des voyelles, nous nous sommes référés à la figure 12 (page 29) ainsi qu'à la classification adoptée par *FANT* (1966).

Les valeurs des coefficients  $k$  obtenus par *FANT* et *PETERSON* varient dans d'assez grandes limites d'une voyelle à une autre.

*FANT* a donné les raisons de tels écarts (1966) :

le conduit vocal de la femme n'est pas rigoureusement homothétique de celui de l'homme pour la production d'une même voyelle ;  
la longueur du conduit vocal est, chez la femme, plus petite que chez l'homme (de l'ordre de 2 cm) mais cette réduction est plus sensible sur la longueur de la cavité du pharynx que sur la longueur de la cavité buccale ;  
de plus, dans certains cas, cette raison est insuffisante et d'autres modifications portant sur les sections des constriction doivent être prises en considération.

Donc, lorsque la fréquence des cordes vocales, pour un homme, augmente, le déplacement du larynx réduit la longueur du conduit vocal par diminution de la longueur de la cavité du pharynx. Ce fait est un premier élément qui montre que les dimensions du conduit vocal d'un homme tendent, lorsque la fréquence des cordes vocales augmente, vers celles d'un conduit vocal de femme.

FIGURE 13

Evolution du coefficient  $k_1$  obtenu à partir du simulateur.

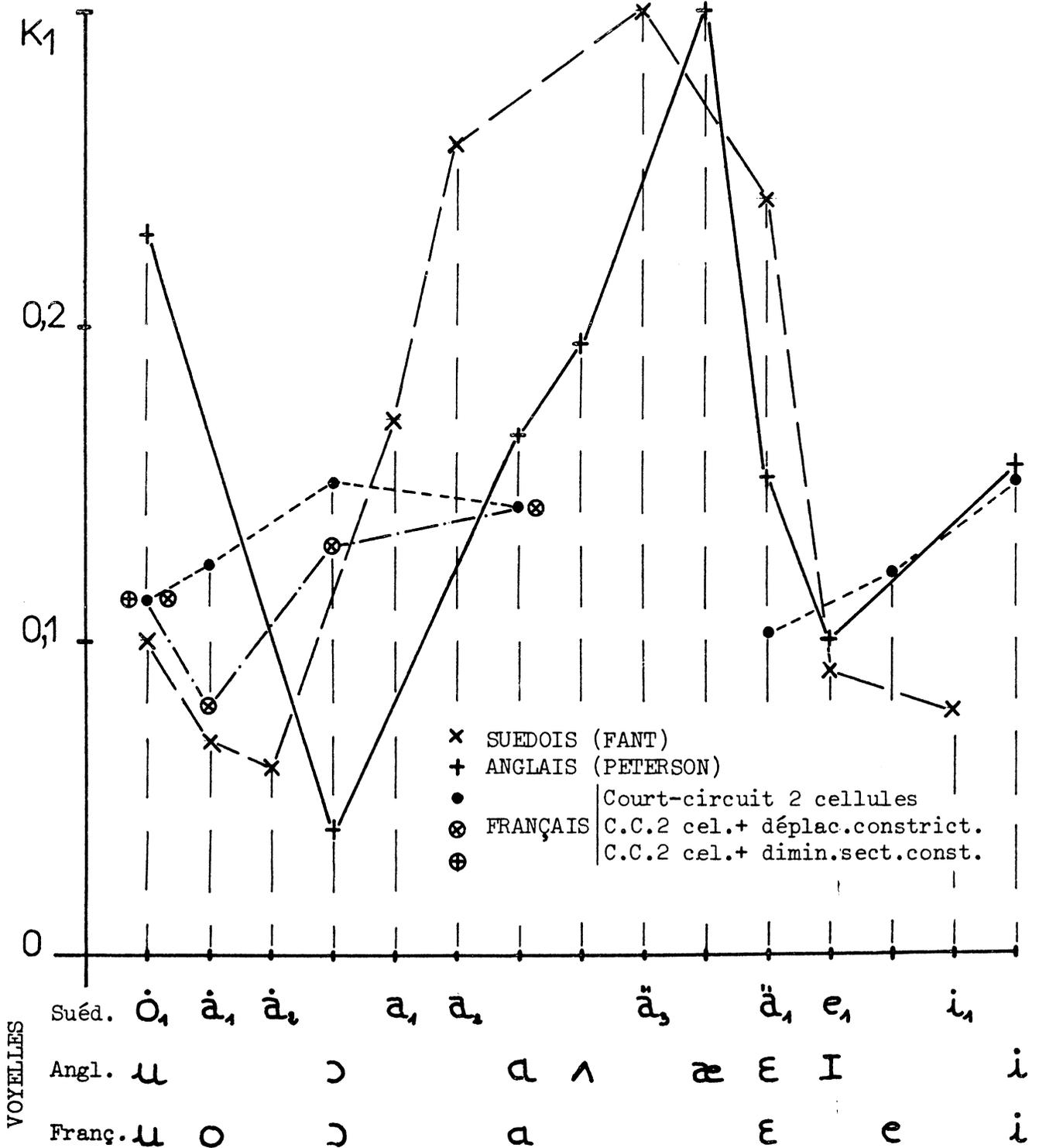
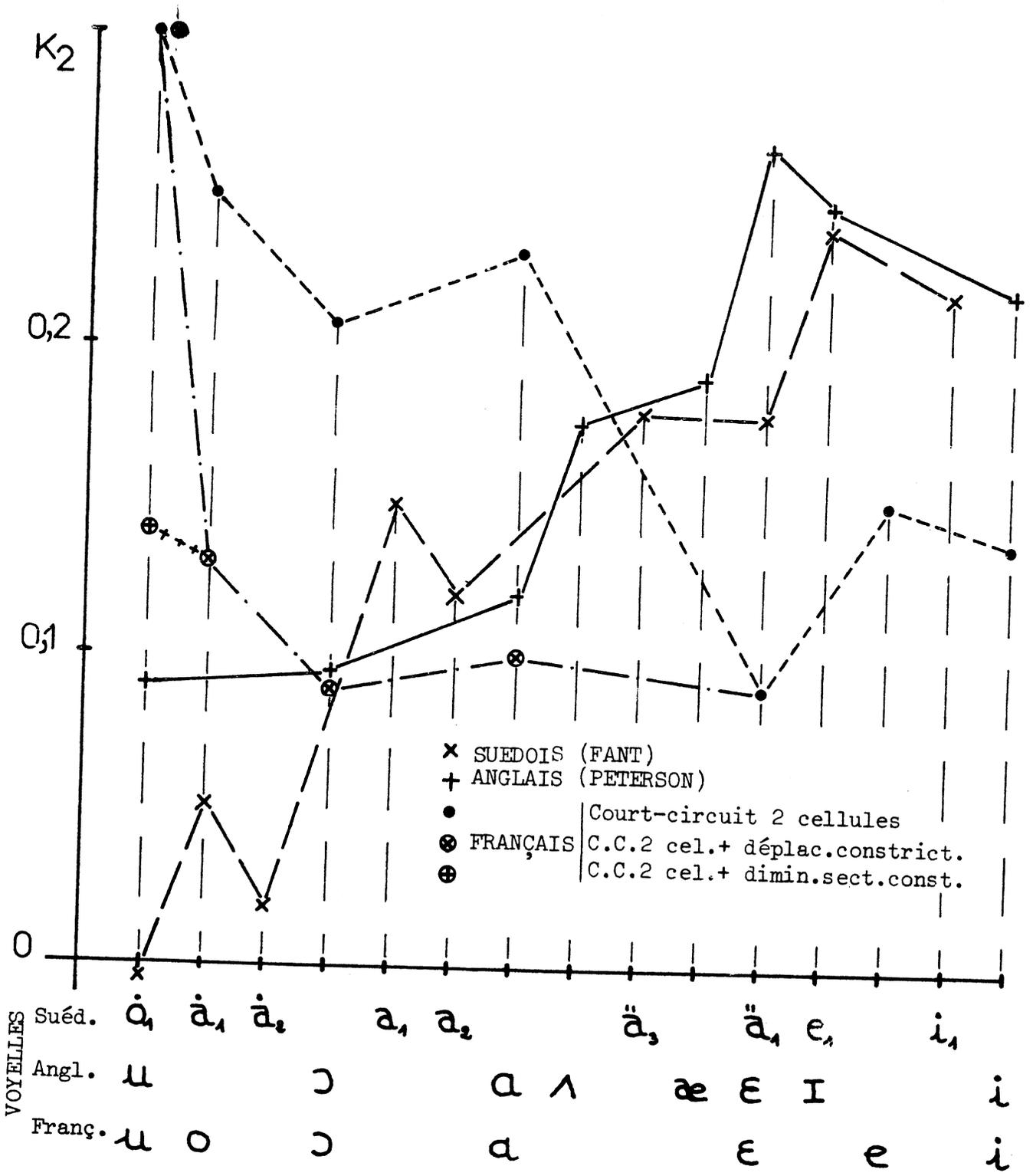


FIGURE 14

Evolution du coefficient  $k_2$  obtenu à partir du simulateur



Les résultats obtenus par simulation, représentés figures 13 et 14, vont dans le sens des résultats obtenus par *FANT* ou *PETERSON*, en particulier pour le coefficient  $k_1$ .

Pour  $k_2$ , les résultats sont assez peu concordants en ce qui concerne les voyelles [u], [o], [ɔ] et [a] du type postérieur. Afin de découvrir les raisons de ces divergences, nous avons jugé utile d'examiner de près l'évolution du conduit vocal lors d'une élévation du larynx.

Les coupes du conduit vocal proposées par *FANT* (1964) et représentées ci-dessous nous montrent que, pour les voyelles postérieures, la constriction se trouve située à proximité du larynx. Il est évident qu'une élévation de ce dernier va provoquer une élévation des tissus environnants et, pour les voyelles postérieures, une élévation de la constriction.

FIGURE 15  
Coupes du  
conduit vocal.



Etant donné la non-rigidité du conduit vocal, on peut penser que ce dernier déplacement sera moindre que celui du larynx. Compte tenu de cette déduction, il apparaît que les modifications tendent alors à assimiler, avec encore plus de précision, un conduit vocal d'homme à un conduit vocal de femme. Nous avons mis à profit cette remarque en modifiant dans le sens voulu, la position de la constriction sur notre simulateur. Les résultats obtenus (nous avons déplacé la constriction de 1 cm), sont représentés figures 13 et 14 pour les voyelles [o], [ɔ] et [a]. Par contre, les coefficients  $k_1$  et  $k_2$  pour la voyelle [u] n'ont pas sensiblement varié. Mais, en nous reportant une nouvelle fois à la figure 15, nous constatons que, pour cette voyelle, une élévation du larynx a tendance à diminuer la section de la constriction plutôt qu'à la déplacer.

Après avoir tenu compte de ce phénomène, nous trouvons :

$$k_1 = 11 \%$$

$$k_2 = 14,2 \%$$

Tous les résultats que nous avons obtenus sont alors en assez bon accord avec ceux de *FANT* ou *PETERSON*.

Nous avons montré, à partir du phénomène connu d'élévation du larynx liée à l'augmentation de la fréquence des cordes vocales et en nous appuyant sur certaines hypothèses fondées sur nos observations, que cette élévation du larynx est un déplacement très important qui tend à transformer un conduit vocal d'homme en un conduit vocal de femme. Il est possible que d'autres réflexes compensateurs jouent également mais ils n'ont pas été envisagés ici (légère modification de l'articulation par exemple).

Il semble donc que, pour un locuteur donné, des modifications anatomiques apparaissent lorsque la fréquence des cordes vocales varie, modification qui facilite l'identification du son produit.

Ces phénomènes de compensation valables pour l'homme sont aussi valables pour la femme dans la mesure où le mécanisme est inversé (abaissement du larynx).

Pour accepter sans réticences l'ensemble des propositions ci-dessus, il faudrait évidemment effectuer une étude complète de la forme du conduit vocal à l'aide de rayons X. Mais on peut s'assurer, par l'analyse spectrale, de l'existence des déplacements des formants au cours de la production d'une voyelle naturelle, à mélodie variable, par un même locuteur.

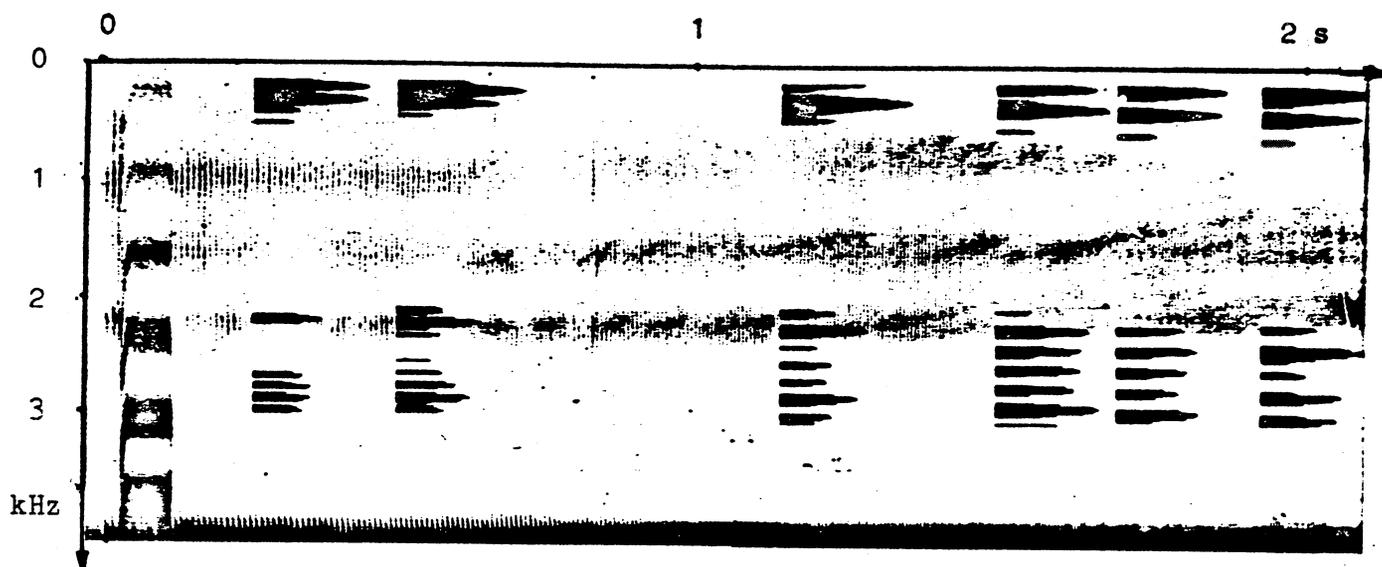
## 2.3. ANALYSE SPECTRALE DE VOYELLES NATURELLES

Nous avons enregistré les signaux images des sept voyelles précédentes prononcées, à mélodie variable (la mélodie variant de 100 Hz à 200 Hz), par un locuteur masculin et effectué une analyse spectrale de ces signaux à l'aide d'un sonographe.

Une photographie du sonagramme de la voyelle [i] est représentée ci-après :

FIGURE 16

Sonagramme de la voyelle [i] ; la fréquence fondamentale croît.



Nous avons effectué des sections aux instants correspondant à une mélodie de 100 Hz, 200 Hz et à des fréquences intermédiaires. La fréquence des formants a été déterminée à l'aide de courbes standards selon la technique décrite par FANT (1962).

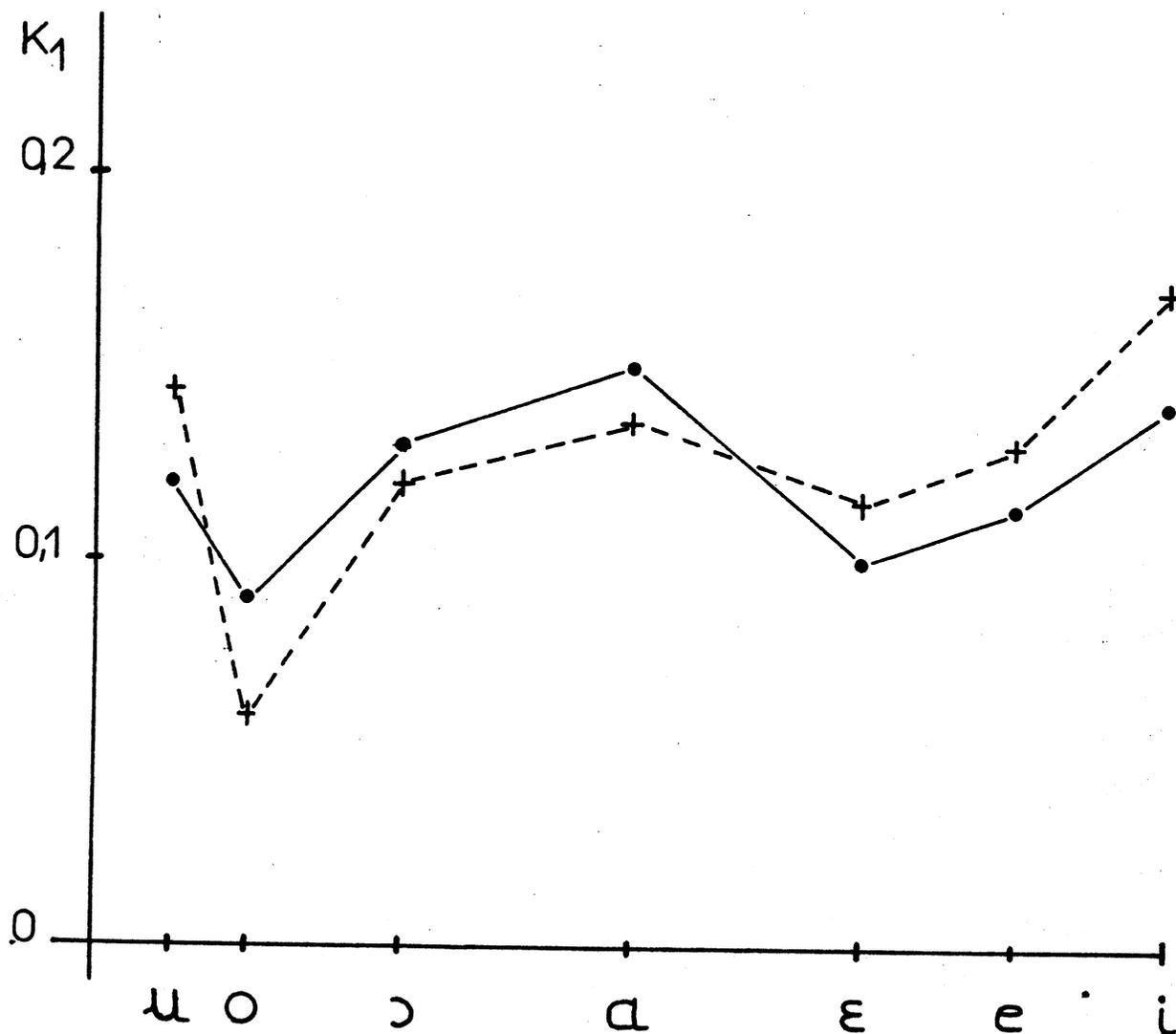
Les résultats obtenus sont représentés figures 17 et 18 (pages 36 et 37).

Ces résultats correspondent bien avec nos propres mesures sur le simulateur vocal.

Nos expériences pourront être étendues à un plus grand nombre de locuteurs hommes ainsi qu'à des femmes prononçant les voyelles précédentes à mélodie décroissante.

FIGURE 17

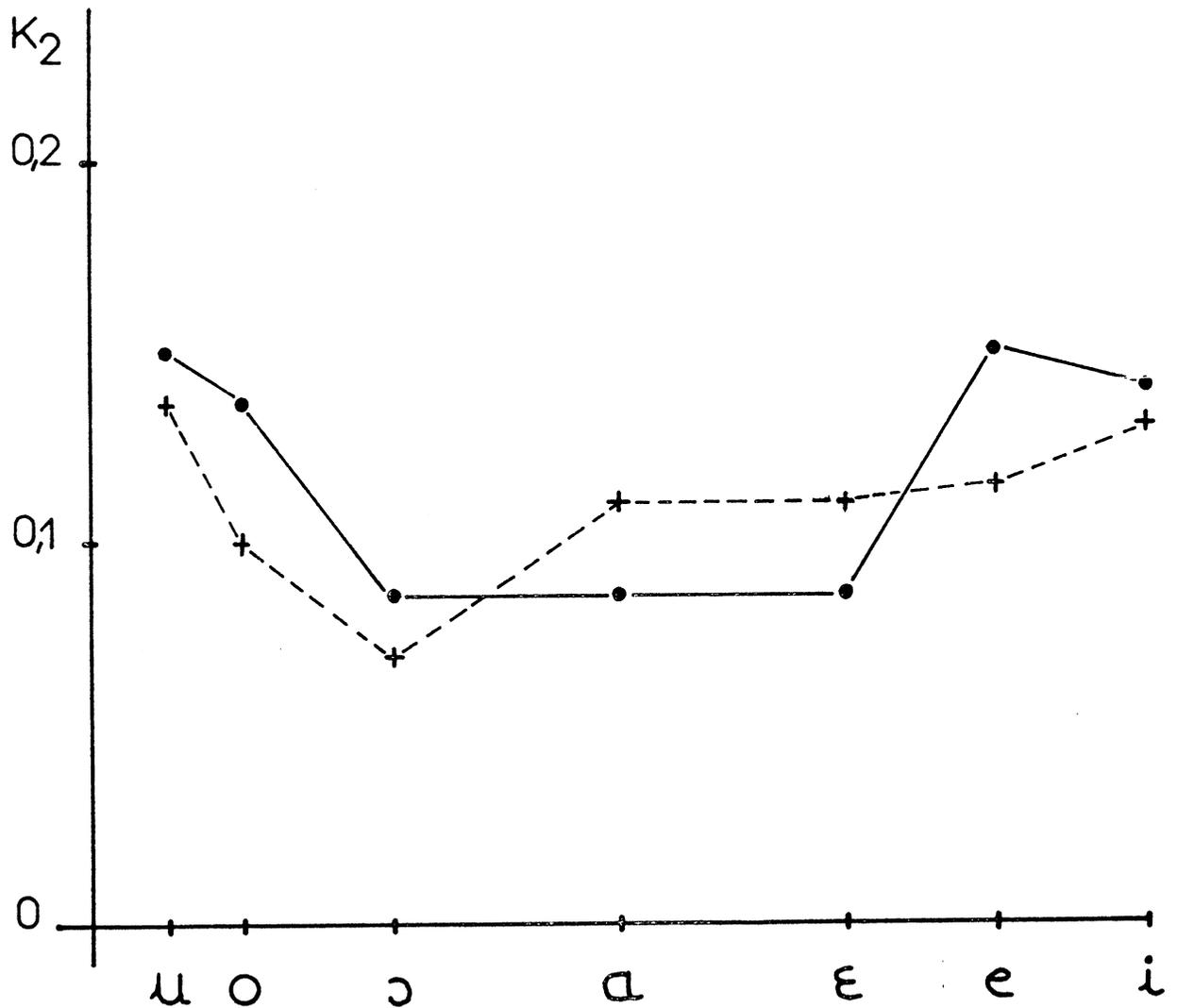
Evolution du coefficient  $k_1$  dans le cas des voyelles naturelles.



- Français - simulation
- + Français - naturel

FIGURE 18

Evolution du coefficient  $k_2$  dans le cas des voyelles naturelles.



- Français - simulation
- + Français - naturel

## 2.4. CONCLUSIONS

Nos expériences montrent que le conduit vocal de l'homme tend à se transformer en conduit vocal de femme lorsque la mélodie augmente ; réciproquement, le conduit vocal de la femme s'allonge lorsque la mélodie diminue. Elles pourront être complétées en diversifiant les locuteurs et en étudiant le conduit vocal à l'aide de rayons X.

Les résultats obtenus peuvent être rapprochés de ceux rapportés par *MEYER-EPPLER* (1957). *MEYER-EPPLER* demande à des locuteurs de chuchoter des voyelles et d'en modifier la mélodie, le mot mélodie ayant une signification très subjective. Les cordes vocales n'étant pas en action, les modifications apportées sont au niveau du spectre et de l'amplitude.

Si *MEYER-EPPLER* constate, pour une mélodie plus "élevée", une augmentation de l'amplitude, il remarque aussi une élévation des fréquences des formants. Il explique ces déplacements de formants par une élévation constatée du larynx.

Mais il faut noter que ces phénomènes de déplacement du larynx ne paraissent pas essentiels.

Ces déplacements sont constatés dans la plupart des cas mais l'amplitude du déplacement varie d'un locuteur à un autre. Dans un autre domaine, les chanteurs professionnels conservent souvent leur larynx immobile. On peut penser que c'est la raison pour laquelle leur chant est souvent difficile à comprendre.

Malgré ces dernières remarques, nous pensons que les résultats que nous avons obtenus devront être pris en considération lors de la mise au point d'un système de reconnaissance de parole.

## - CHAPITRE 3 -

LES FORMANTS,  
ELEMENTS D'IDENTIFICATION DES LOCUTEURS

## 3.1. INTRODUCTION

La parole d'un locuteur donné est caractérisée, essentiellement, par sa source vocale (fréquence moyenne et évolutions de cette fréquence) et par son conduit vocal (formes et évolutions). L'action de ce dernier sur le signal de source pouvant être représentée par sa fonction de transfert ou, plus simplement, par la donnée des pôles correspondants aux formants de cette fonction de transfert, on peut donc penser qu'une étude sur les formants et leurs évolutions apportera des renseignements intéressants pour l'identification des locuteurs.

Nous avons étudié les évolutions des formants, en particulier, dans les plans  $F_1 - F_2$  et  $F_2 - F_3$ . L'emploi de ces paramètres et de ce mode de représentation, original dans le cas de l'étude de l'identification, simplifie les comparaisons et donne des résultats intéressants.

La recherche des paramètres de la parole jouant un rôle dans l'identification d'un locuteur est un problème d'actualité. Cette identification, à partir du signal vocal, éveille beaucoup la curiosité du grand public. On parle d'*empreintes vocales* qui pourraient concurrencer les empreintes digitales. Disons tout de suite qu'il est difficile de faire un parallèle entre ces deux types d'empreintes et que la méfiance actuelle des hommes de justice devant les techniques d'identification d'un locuteur est tout-à-fait justifiée (R.H. BOLT, F.S. COOPER, E.E. DAVIS, P.B. DENES, J.M. PICKETT, K.N. STEVENS - 1970).

Lors des procès en justice, on dispose parfois d'un enregistrement magnétique d'une voix normale ou déformée, soit volontairement soit par une mauvaise transmission téléphonique par exemple, et on compare cette voix avec une ou plusieurs autres. On souhaite alors pouvoir dire : la voix inconnue est celle de tel ou tel, ou bien la voix inconnue ne correspond à aucune des voix des locuteurs choisis pour la comparaison. Or, en l'état actuel des travaux dans ce domaine, il est impossible d'assurer de manière formelle l'identification d'un locuteur.

Actuellement, à part quelques rares travaux sur l'imitation des voix (*L.G. KERSTA - 1962*) et quelques études où la voix à identifier n'est pas dans le catalogue des échantillons références (*J.E.LUCK - 1965 ; K.P.LI, J.E.DAMMANN, W.D.CHAPMAN - 1966 ; K.N.STEVENS, C.E.WILLIAMS, J.R.CARBONELL, B.WOODS - 1968*), les recherches portent essentiellement sur l'identification d'un locuteur dont on dispose d'un échantillon de parole à l'intérieur d'un catalogue réduit (une dizaine d'échantillons par exemple). Par ailleurs, on ne tient pas compte des modifications de la voix avec l'âge.

Ces travaux sur l'identification des locuteurs apportent une contribution d'un genre nouveau à notre connaissance des paramètres de la parole. Il s'agit de préciser le rôle et l'importance de chacun d'entre eux selon que l'on s'intéresse au message proprement dit, au contenu esthétique, à l'identification du locuteur...

On a cherché à étudier les possibilités d'identification chez un auditeur, à qualifier les paramètres subjectifs entrant en jeu et à étudier le rôle de ces paramètres dans l'identification (*W.D.VOIERS - 1964 ; I.POLLACK, J.M.PICKETT, W.M.SUMBY - 1954*).

On a aussi cherché à identifier les locuteurs à partir d'une analyse physique du signal de parole. *KERSTA (1962)*, *YOUNG* et *CAMPBELL (1967)* effectuent visuellement des comparaisons de spectres du signal de parole (sonagrammes).

*HARGREAVES* et *STARKWEATHER (1963)* et *PRUZANSKY (1963)* effectuent ces comparaisons mathématiquement.

Par ailleurs, des études effectuées à partir de la mesure du cepstre ont été faites (*J.E.LUCK - 1969*).

Mais on a également cherché à déterminer dans le spectre très riche de la parole quels pouvaient être les éléments intéressants pour l'identification des

locuteurs. C'est ainsi que les caractéristiques de certains sons peuvent dépendre plus particulièrement d'un locuteur donné. *GLENN* et *KLEINER* (1968) ont comparé les spectres des consonnes nasales (la forme et les dimensions d'un conduit nasal sont des caractéristiques propres à un locuteur donné).

On a aussi recherché, dans l'un et l'autre spectre de deux échantillons de parole d'un même locuteur, les surfaces élémentaires, déterminées par quantification des échelles temps et fréquence, dont les états sont analogues mais qui, par contre, changent notablement d'un locuteur à l'autre (*S. PRUZANSKY, M.V. MATHEWS - 1964*).

Cette recherche d'éléments essentiels et leur comparaison entre eux ne sont pas simples. On peut simplifier le spectre en ne retenant que les données relatives aux formants, les fréquences des formants dépendant de la forme et des évolutions du conduit vocal d'un locuteur donné.

Des recherches (*J.J. WOLF - 1970*) ont d'ailleurs révélé une corrélation entre les formants et les surfaces élémentaires du plan temps-fréquence les plus significatives sur le plan de l'identification.

Quelles sont alors les importances relatives des formants dans l'identification des locuteurs ? Comment effectuer les comparaisons des données sur les divers formants ?

Nous allons présenter les résultats des comparaisons des fréquences de formants correspondant à dix locuteurs hommes, repérés 1 à 10, prononçant deux fois successivement un même échantillon de parole (les dix locuteurs proviennent de milieux peu différents).

Des sonagrammes des échantillons de parole sont effectués, sur lesquels on repère les fréquences  $F_1$ ,  $F_2$  et  $F_3$  des trois premiers formants des séquences voisées. Ces séquences sont au nombre de deux :

- . l'une ( $S_1$ ) correspond à une voyelle d'une durée moyenne de 150 ms,
- . l'autre ( $S_2$ ) correspond à trois syllabes d'une durée moyenne de 500 ms.

Nous avons aussi repéré les évolutions de la fréquence fondamentale  $F_0$  en vue d'étudier le rôle de ce paramètre important lorsqu'il est associé aux formants pour l'identification d'un locuteur.

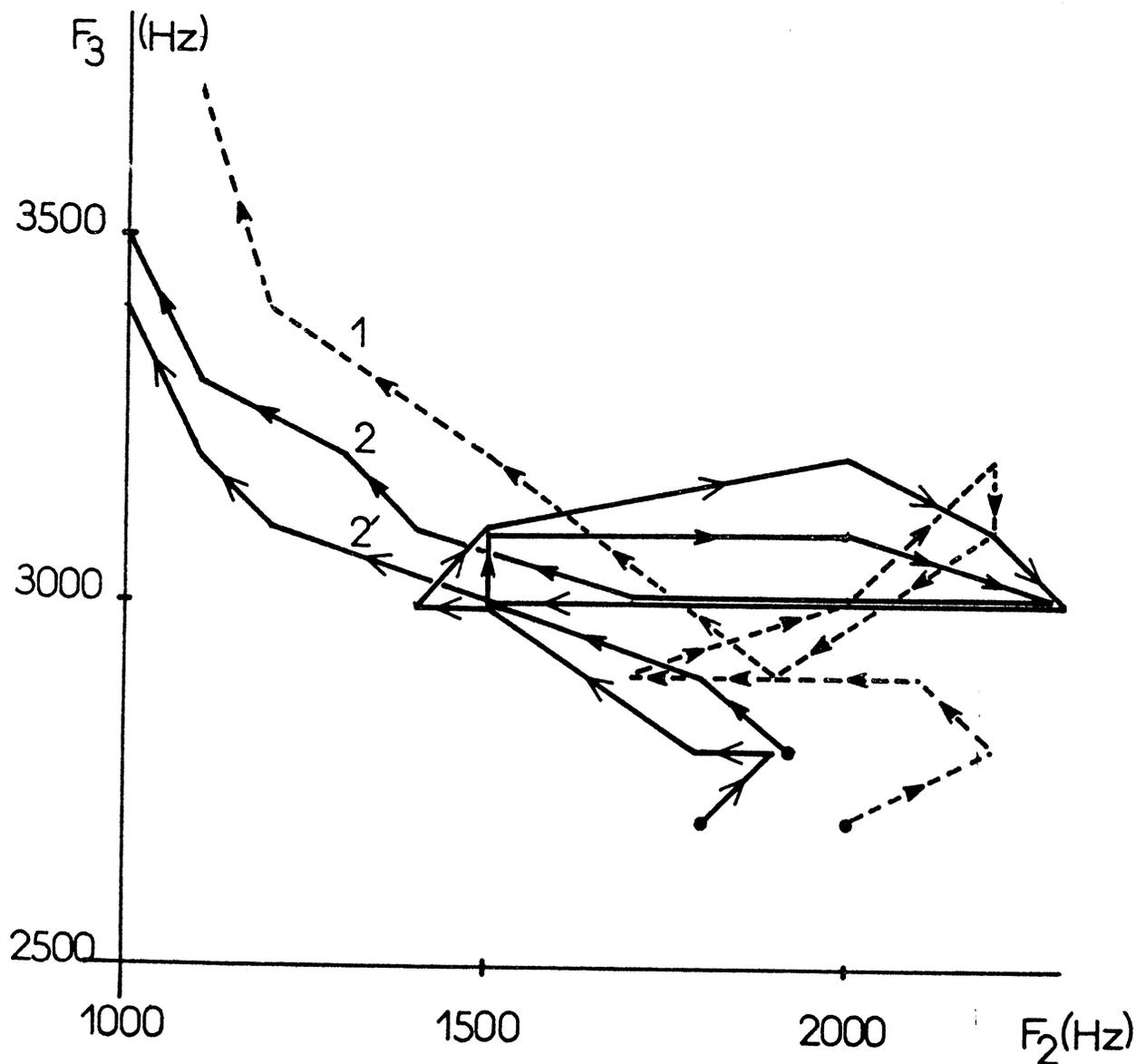
Les comparaisons des données ont été effectuées dans les plans  $F_1$ - $F_2$ ,  $F_2$ - $F_3$  et  $F_0$ - $F_3$ . Elles ont aussi été effectuées à partir d'une étude des courbes de répartition et de probabilité sur l'échelle des fréquences pour un paramètre donné.

3.2. COMPARAISON DES FREQUENCES DES PARAMETRES  
DANS LES PLANS  $F_1 - F_2$ ,  $F_2 - F_3$  et  $F_0 - F_3$

Les évolutions des fréquences des formants  $F_2$ ,  $F_3$  de la deuxième séquence voisée  $S_2$ , ont été reportées (figure 19) dans le plan  $F_2 - F_3$  pour le locuteur repéré "1" (échantillon n° 1) et pour le locuteur repéré "2" (deux échantillons n° 2 et 2').

FIGURE 19

*Evolutions des fréquences des formants dans le plan  $F_2 - F_3$  - Séquence  $S_2$ .  
Locuteur "1" (échantillon n° 1) ; Locuteur "2" (deux échantillons n° 2 et 2').*



Dans notre étude, les comparaisons des tracés deux à deux ont été effectuées d'une part par dix observateurs non entraînés et, d'autre part, par corrélation. Pour cette corrélation, les plans  $F_1-F_2$ ,  $F_2-F_3$  et  $F_0-F_3$  sont quantifiés, selon les axes, par bonds de 50 Hz. Les courbes sont alors représentées selon cette quantification.

On calcule alors pour chaque comparaison les coefficients  $k$  :

$$k = \frac{1}{N} \frac{\sum_{n=1}^N x_n y_n}{\sqrt{m_x m_y}}$$

$x_1 \dots x_n$  représentent les valeurs du plan considéré pour une courbe déterminée,  
 $y_1 \dots y_n$  représentent les valeurs correspondantes pour la courbe à comparer à la précédente,  
 $m_x$  et  $m_y$  sont les moyennes.

Pour effectuer la corrélation, nous avons admis des translations d'amplitude d'une courbe par rapport à une autre, pouvant aller jusqu'à 150 Hz selon les axes. Ces translations permettent de tenir compte des différences entre les prononciations, par un même locuteur, d'une même phrase. On retient le coefficient de corrélation le plus grand.

### 3.2.1. COMPARAISONS DANS LE PLAN $F_1-F_2$

Les tableaux 20 et 21 (pages 44 et 45) recouvrent les résultats obtenus respectivement pour les séquences voisées  $S_1$  et  $S_2$  de durée respective 150 et 500 ms environ, à partir des courbes du plan  $F_1-F_2$ , les comparaisons étant effectuées par dix observateurs non entraînés. Les courbes références sont soit les courbes 1, 2, ... 10, soit les courbes 1', 2', ... 10'.

Le tableau 22 (page 46) donne les coefficients de corrélation, pour la séquence  $S_2$ , obtenus comme il a été indiqué à partir des courbes du plan  $F_1-F_2$ .

La comparaison de ce tableau et des tableaux 20 et 21 montre qu'il existe une certaine concordance entre les résultats annoncés par les observateurs et ceux obtenus par corrélation, sans qu'il nous soit possible de préciser laquelle des deux méthodes est la meilleure.

Dans la suite, nous étudierons plus spécialement les résultats obtenus par corrélation, cette méthode étant plus systématique pour comparer rapidement et objectivement les données.



TABLEAUX 21 - Séquence S<sub>2</sub> ;  
Comparaisons des observateurs dans le plan F<sub>1</sub>-F<sub>2</sub>.

réf- érence →	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	10									
2		7					2	1		
3			8		1		1			
4				5		1		2		2
5					10					
6			1		1	3			5	
7		1					7	2		
8		4						6		
9				2					3	5
10	1			1			1		1	6

réf- érence ↓	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	10					1				
2		7				1	2		1	
3			7						1	
4				4						
5			1		10					
6			2			3			3	
7				1		3	4			
8		3		1			4	9		
9				2		2			3	1
10				2				1	2	9

TABLEAU 22 - Séquence S<sub>2</sub> ;  
Comparaisons par corrélation dans le plan F<sub>1</sub>-F<sub>2</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>0,62</u>	0,19	0,13	0,18	0,10	0,09	0,13	0,17	0,07	0,19
2	0,16	<u>0,54</u>	0,33	0,37	0,30	0,13	0,34	0,35	0,42	0,27
3	0,10	0,32	<u>0,47</u>	0,17	0,23	0,08	0,28	0,21	0,16	0,18
4	0,20	0,35	0,28	0,35	0,27	0,32	0,30	0,24	0,27	<u>0,36</u>
5	0,15	0,17	0,35	0,26	<u>0,53</u>	0,28	0,28	0,17	0,41	0,23
6	0,10	0,14	0,18	<u>0,43</u>	0,25	<u>0,34</u>	0,29	0,15	0,26	0,31
7	0,25	0,31	0,22	0,26	0,18	0,26	<u>0,55</u>	0,22	0,32	0,26
8	0,05	<u>0,62</u>	0,26	0,25	0,16	0,24	0,50	<u>0,54</u>	0,34	0,34
9	0,06	0,35	0,36	0,27	0,32	0,25	<u>0,50</u>	0,35	<u>0,48</u>	0,22
10	0,17	0,32	0,30	0,25	0,16	0,26	<u>0,58</u>	0,24	0,22	<u>0,38</u>

### 3.2.2. COMPARAISONS DANS LE PLAN F<sub>2</sub>-F<sub>3</sub>

Les tableaux 23 et 24 (pages 47 et 48) représentent les résultats obtenus pour les séquences voisées S<sub>1</sub> et S<sub>2</sub>, dans le plan F<sub>2</sub>-F<sub>3</sub>, les comparaisons étant effectuées par les observateurs.

Le tableau 25 (page 49) donne les coefficients de corrélation obtenus, dans le plan F<sub>2</sub>-F<sub>3</sub>, pour la séquence S<sub>2</sub>. On peut voir que les résultats obtenus à partir de la séquence S<sub>2</sub> sont bien plus intéressants du point de vue de l'identification comparés à ceux obtenus à partir de la séquence S<sub>1</sub>.

### 3.2.3. COMPARAISONS DANS LE PLAN F<sub>0</sub>-F<sub>3</sub>

Pour tenir compte des informations susceptibles d'être apportées par la fréquence fondamentale, nous avons comparé les courbes tracées dans le plan F<sub>0</sub>-F<sub>3</sub>. Le tableau 26 (page 49) donne les coefficients de corrélation pour la séquence S<sub>2</sub>.





TABLEAU 25 - Séquence S<sub>2</sub> ;  
Comparaisons par corrélation dans le plan F<sub>2</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>0,64</u>	0,28	0,08	0,02	0,03	0,01	0,05	0,09	0,02	0,07
2	0,34	<u>0,60</u>	0,02	0,01	0,07	0,04	0,05	0,09	0,00	0,13
3	0,33	0,24	<u>0,62</u>	0,32	0,61	0,39	0,37	0,40	0,61	0,26
4	0,11	0,04	0,39	<u>0,55</u>	0,36	0,44	0,50	0,32	0,34	0,36
5	0,28	0,23	0,34	0,30	<u>0,63</u>	0,25	0,18	0,28	0,53	0,29
6	0,07	0,07	0,34	0,36	0,38	<u>0,67</u>	0,33	0,35	0,31	0,27
7	0,16	0,07	0,50	0,22	0,24	0,53	<u>0,54</u>	0,27	0,40	0,26
8	0,29	0,18	0,57	0,26	0,40	0,38	0,50	<u>0,61</u>	0,57	0,49
9	0,06	0,10	0,48	0,27	0,62	0,35	0,32	0,44	<u>0,65</u>	0,34
10	0,02	0,19	0,46	0,28	0,49	0,43	0,33	0,28	0,40	<u>0,51</u>

TABLEAU 26 - Séquence S<sub>2</sub> ;  
Comparaisons par corrélation dans le plan F<sub>0</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>0,46</u>	0,06	0,00	0,00	0,06	0,17	0,21	0,03	0,00	0,15
2	0,08	<u>0,38</u>	0,00	0,03	0,29	0,00	0,00	0,17	0,09	0,20
3	0,00	0,03	<u>0,44</u>	0,16	0,00	0,00	0,00	0,00	0,01	0,00
4	0,10	0,10	0,09	<u>0,50</u>	0,23	0,41	0,19	0,24	0,20	0,21
5	0,06	0,27	0,03	0,28	<u>0,56</u>	0,24	0,19	0,28	0,41	<u>0,66</u>
6	0,00	0,11	0,22	0,47	0,12	<u>0,48</u>	0,06	0,40	0,16	0,13
7	0,25	0,15	0,00	0,35	0,36	0,26	<u>0,57</u>	0,25	0,36	0,33
8	0,03	0,13	0,17	0,43	0,36	0,24	0,19	0,46	<u>0,53</u>	0,28
9	0,00	0,07	0,32	0,40	0,34	0,15	0,17	0,32	0,48	0,36
10	0,05	0,28	0,07	0,36	<u>0,44</u>	0,32	0,23	0,30	0,35	0,39

## 3.2.4. DISCUSSION

Il est difficile d'exprimer le taux d'identification obtenu car le nombre de comparaisons est très insuffisant mais on peut noter que, dans notre expérience, l'identification est totale à partir des deux paramètres  $F_2$  et  $F_3$  (tableau 25, page 49).

Une meilleure confiance paraît être obtenue en fondant celle-ci sur l'examen des sommes, terme à terme, des coefficients de corrélation relatifs aux plans  $F_1-F_2$ ,  $F_2-F_3$  (tableau 27, page 51), puis aux plans  $F_1-F_2$ ,  $F_2-F_3$ ,  $F_0-F_3$  (tableau 28, page 51).

Les résultats de cette sommation montrent que chacune des comparaisons n'apporte pas le même type d'information en ce qui concerne l'identification. On peut alors penser qu'une multiplication terme à terme des coefficients est préférable comparée à l'addition.

Les tableaux 29 et 30 (page 52) donnent les résultats d'une telle multiplication pour les plans  $F_1-F_2$  et  $F_2-F_3$ , puis pour les plans  $F_1-F_2$ ,  $F_2-F_3$  et  $F_0-F_3$ . Leur examen montre qu'effectivement le produit est préférable à la somme.

Enfin, le fait que les formants  $F_1$  et  $F_2$  soient porteurs d'informations sur le locuteur peut expliquer les difficultés rencontrées en reconnaissance de parole à partir de ces formants.

On constate aussi que les tableaux ne sont pas symétriques. Ce résultat, conforme à celui obtenu par *BRICKER* (1966), montre que, pour un même locuteur, deux échantillons d'une même séquence peuvent, du point de vue de l'identification, se rapprocher d'échantillons de la même séquence prononcés chacun par deux autres locuteurs.

On remarque aussi que certaines voix se comparent difficilement à d'autres sur le plan des formants alors que d'autres voix semblent plus "standard" et sont, par suite, plus difficiles à identifier.

Enfin, le taux d'identification dépend de la longueur de l'échantillon étudié. Une durée de 150 ms est nettement insuffisante alors qu'avec 500 ms on obtient des résultats satisfaisants. Ce constat est en accord avec les observations rapportées par *POLLACK*, *PICKETT*, *SUMBY* (1954), *STEVENS et al.* (1968), bien que ceux-ci présentent comme optimale une durée de 1,2 seconde environ.

TABLEAU 27 - Séquence S<sub>2</sub> ;  
Somme, terme à terme, des tableaux  
correspondant aux plans F<sub>1</sub>-F<sub>2</sub> et F<sub>2</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>1,26</u>	0,47	0,21	0,20	0,13	0,10	0,18	0,26	0,09	0,26
2	0,50	<u>1,14</u>	0,35	0,38	0,37	0,17	0,40	0,44	0,42	0,40
3	0,43	0,56	<u>1,09</u>	0,49	0,84	0,47	0,65	0,61	0,77	0,44
4	0,31	0,39	0,67	<u>0,90</u>	0,63	0,76	0,80	0,56	0,61	0,72
5	0,43	0,40	0,69	0,56	<u>1,16</u>	0,53	0,46	0,45	0,94	0,52
6	0,17	0,21	0,52	0,79	0,63	<u>1,01</u>	0,62	0,50	0,57	0,58
7	0,41	0,38	0,72	0,48	0,42	0,79	<u>1,09</u>	0,49	0,72	0,52
8	0,34	0,80	0,83	0,51	0,56	0,62	1,00	<u>1,15</u>	0,91	0,83
9	0,12	0,45	0,84	0,54	0,94	0,60	0,82	0,79	<u>1,13</u>	0,56
10	0,19	0,51	0,76	0,53	0,65	0,69	0,91	0,52	0,62	<u>0,89</u>

TABLEAU 28 - Séquence S<sub>2</sub> ;  
Somme, terme à terme, des tableaux  
correspondant aux plans F<sub>1</sub>-F<sub>2</sub>, F<sub>2</sub>-F<sub>3</sub> et F<sub>0</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>1,72</u>	0,53	0,21	0,20	0,19	0,27	0,39	0,29	0,09	0,41
2	0,58	<u>1,52</u>	0,35	0,41	0,66	0,17	0,40	0,61	0,51	0,60
3	0,43	0,59	<u>1,53</u>	0,65	0,84	0,47	0,65	0,61	0,78	0,44
4	0,41	0,49	0,76	<u>1,40</u>	0,86	1,17	0,89	0,80	0,81	0,93
5	0,49	0,67	0,72	0,84	<u>1,72</u>	0,77	0,65	0,73	1,35	1,18
6	0,17	0,32	0,74	1,26	0,75	<u>1,49</u>	0,68	0,90	0,73	0,71
7	0,66	0,53	0,72	0,83	0,78	1,05	<u>1,66</u>	0,74	1,08	0,85
8	0,37	0,93	1,00	0,94	0,92	0,86	1,19	<u>1,61</u>	1,43	1,11
9	0,12	0,52	1,16	0,94	1,28	0,75	0,99	1,11	<u>1,61</u>	0,92
10	0,24	0,79	0,83	0,89	1,09	1,01	1,14	0,82	0,97	<u>1,28</u>

TABLEAU 29 - Séquence S<sub>2</sub> ;  
 Multiplication, terme à terme, des tableaux  
 correspondant aux plans F<sub>1</sub>-F<sub>2</sub> et F<sub>2</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>0,398</u>	0,053	0,010	0,035	0,003	0,001	0,006	0,015	0,001	0,013
2	0,054	<u>0,324</u>	0,006	0,003	0,021	0,005	0,024	0,038	0,000	0,035
3	0,033	0,077	<u>0,292</u>	0,054	0,140	0,031	0,104	0,084	0,098	0,047
4	0,022	0,014	0,109	<u>0,194</u>	0,097	0,140	0,150	0,077	0,110	0,130
5	0,042	0,039	0,118	0,078	<u>0,334</u>	0,070	0,051	0,048	0,218	0,067
6	0,007	0,010	0,061	0,155	0,095	<u>0,228</u>	0,096	0,053	0,081	0,084
7	0,040	0,022	0,110	0,057	0,043	0,138	<u>0,298</u>	0,060	0,128	0,068
8	0,014	0,114	0,148	0,065	0,064	0,091	0,250	<u>0,330</u>	0,194	0,166
9	0,003	0,035	0,176	0,073	0,198	0,088	0,160	0,154	<u>0,312</u>	0,075
10	0,003	0,061	0,138	0,070	0,078	0,112	0,192	0,067	0,088	<u>0,194</u>

TABLEAU 30 - Séquence S<sub>2</sub> ;  
 Multiplication, terme à terme, des tableaux  
 correspondant aux plans F<sub>1</sub>-F<sub>2</sub>, F<sub>2</sub>-F<sub>3</sub> et F<sub>0</sub>-F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>0,183</u>	0,003	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,002
2	0,004	<u>0,123</u>	0,000	0,000	0,006	0,000	0,000	0,008	0,000	0,007
3	0,000	0,002	<u>0,128</u>	0,009	0,000	0,000	0,000	0,000	0,001	0,000
4	0,002	0,001	0,010	<u>0,097</u>	0,022	0,057	0,030	0,018	0,022	0,026
5	0,002	0,011	0,003	0,022	<u>0,187</u>	0,017	0,010	0,013	0,089	0,041
6	0,000	0,001	0,013	0,073	0,001	<u>0,110</u>	0,006	0,020	0,013	0,011
7	0,010	0,003	0,000	0,027	0,015	0,036	<u>0,170</u>	0,015	0,046	0,022
8	0,000	0,015	0,025	0,028	0,023	0,022	0,048	<u>0,152</u>	0,100	0,047
9	0,000	0,002	0,052	0,029	0,067	0,013	0,027	0,049	<u>0,150</u>	0,027
10	0,000	0,017	0,009	0,025	0,034	0,036	0,044	0,020	0,031	<u>0,075</u>

En comparant les tableaux 22 (page 46) et 25 (page 49), on constate que la conjugaison des formants  $F_2$  et  $F_3$  permet d'effectuer une meilleure différenciation des locuteurs que celle des formants  $F_1$  et  $F_2$ .

Ceci est en accord avec les premières observations rapportées par *WOLF* (1970) et par *STEVENS et al.* (1968).

*STEVENS* constate que les voyelles antérieures sont plus facilement identifiées. Or, précisément, pour ces voyelles, les deuxième et troisième formants sont accentués.

### 3.3. COMPARAISON DES FREQUENCES DES PARAMETRES DANS LES PLANS REPARTITION - FREQUENCE ET PROBABILITE - FREQUENCE

Il a été indiqué au Premier chapitre que la répartition des fréquences des formants est différente d'un locuteur à un autre. Cette différence peut être exploitée pour l'identification.

Nous avons calculé l'écart quadratique entre les vingt courbes de répartition relatives aux évolutions des formants relevées pour l'expérience précédente. Ces vingt courbes correspondent aux 20 échantillons, résultat de la prononciation par dix locuteurs, deux fois successivement, de la séquence  $S_2$ .

Nous avons aussi calculé cet écart quadratique à partir des courbes de probabilité déduites des courbes de répartition.

#### 3.3.1. ECART QUADRATIQUE DANS LE CAS DU PREMIER FORMANT

Les tableaux 31 et 32 (page 54) représentent les résultats obtenus à partir, respectivement, des courbes de répartition et de probabilité.

Nous pouvons constater que le taux d'identification est pratiquement le même que les écarts soient calculés à partir des courbes de répartition ou à partir de celles de probabilité.

#### 3.3.2. ECART QUADRATIQUE DANS LE CAS DU DEUXIEME FORMANT

Les tableaux 33 et 34 (page 55) représentent les résultats obtenus à partir, respectivement, des courbes de répartition et de probabilité.

On peut constater que, pour le deuxième formant, le taux d'identification est supérieur à celui obtenu précédemment pour le premier formant.

TABLEAU 31 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de répartition dans le cas de  $F_1$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>12,4</u>	40,7	41,9	36,3	16,6	35,7	72,7	71,0	53,4	71,9
2	<u>12,2</u>	<u>39,3</u>	40,7	34,5	<u>11,5</u>	33,6	73,0	71,3	53,0	72,1
3	30,1	12,2	<u>11,3</u>	16,6	41,0	23,4	32,9	29,0	12,4	31,0
4	25,1	4,6	6,2	<u>1,4</u>	30,5	13,2	42,5	39,6	20,1	41,1
5	20,6	32,3	33,9	26,3	<u>8,3</u>	24,4	68,6	67,3	47,8	67,8
6	23,0	21,5	22,8	17,0	21,7	<u>7,5</u>	52,6	54,0	35,3	52,5
7	30,0	<u>2,2</u>	<u>0,6</u>	8,2	37,0	17,1	36,3	32,8	13,3	34,6
8	61,8	39,4	37,9	44,8	71,0	45,4	<u>2,5</u>	18,0	26,5	7,0
9	51,0	26,3	24,7	32,3	60,6	38,2	17,8	<u>8,7</u>	<u>10,8</u>	14,0
10	61,1	38,3	36,8	43,8	70,5	45,1	<u>0,0</u>	15,5	24,9	<u>4,5</u>

TABLEAU 32 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de  $F_1$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>12,6</u>	40,8	42,6	36,1	24,5	30,4	70,9	81,4	57,7	73,6
2	<u>15,7</u>	40,9	43,2	34,7	15,8	26,6	74,9	85,3	60,2	77,6
3	36,3	22,1	19,8	29,5	59,4	43,9	27,7	33,4	<u>12,9</u>	27,7
4	25,0	7,1	9,9	<u>1,4</u>	36,4	21,4	49,3	56,6	29,7	50,6
5	26,3	45,4	48,2	37,0	<u>10,7</u>	25,5	85,1	94,8	68,0	87,6
6	28,9	36,9	39,4	29,4	16,8	<u>10,7</u>	71,5	84,2	58,1	75,0
7	30,1	<u>2,8</u>	<u>0,0</u>	11,3	45,8	29,4	40,4	46,7	19,8	41,2
8	59,1	42,9	40,5	50,3	80,1	59,8	<u>2,8</u>	25,5	26,2	9,9
9	57,9	36,8	33,9	45,3	78,9	61,2	19,6	<u>12,7</u>	14,1	14,4
10	59,9	42,9	40,4	50,6	81,1	61,0	<u>0,0</u>	22,6	25,0	<u>7,1</u>

TABLEAU 33 - Séquence S<sub>2</sub> ;

Ecart quadratique entre les courbes  
de répartition dans le cas de F<sub>2</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>13,1</u>	22,6	43,0	48,1	48,3	36,0	36,6	18,8	36,4	20,6
2	27,6	<u>8,7</u>	26,2	37,1	35,8	30,6	18,5	20,0	32,4	19,9
3	46,8	28,4	<u>11,6</u>	29,2	25,2	37,3	16,4	31,5	35,7	36,9
4	49,9	34,0	26,5	<u>10,5</u>	<u>7,2</u>	29,6	30,8	39,6	29,7	45,9
5	44,9	34,6	22,9	15,2	<u>9,2</u>	28,5	28,7	36,0	28,2	40,3
6	24,9	23,3	28,1	24,9	26,1	<u>13,0</u>	26,7	24,7	19,6	26,3
7	42,4	24,0	15,1	33,6	27,6	27,3	<u>4,2</u>	24,3	33,3	26,4
8	22,2	22,4	35,5	47,0	49,0	31,1	28,2	14,1	27,6	<u>13,2</u>
9	30,7	30,6	31,4	31,4	35,3	16,9	30,9	25,0	<u>8,1</u>	30,0
10	28,2	17,7	29,4	43,8	45,7	34,2	19,2	<u>10,1</u>	32,0	<u>9,9</u>

TABLEAU 34 - Séquence S<sub>2</sub> ;

Ecart quadratique entre les courbes  
de probabilité dans le cas de F<sub>2</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>19,9</u>	32,9	43,5	34,4	33,1	33,9	38,1	22,1	36,7	29,5
2	29,2	<u>11,8</u>	23,3	29,2	29,5	26,3	19,0	27,1	30,0	25,8
3	33,5	19,5	<u>17,2</u>	27,9	25,8	30,3	19,1	24,7	24,8	31,0
4	36,2	33,6	36,8	<u>12,5</u>	<u>12,4</u>	26,5	37,5	33,2	19,5	42,1
5	33,7	31,1	33,6	17,6	<u>10,9</u>	27,9	38,2	34,1	20,6	44,1
6	<u>17,6</u>	22,2	29,1	20,2	17,7	<u>14,2</u>	29,5	27,3	24,1	29,9
7	38,0	24,9	23,5	37,0	40,5	32,9	<u>6,3</u>	22,7	34,1	17,8
8	28,8	21,1	21,0	33,8	35,4	27,2	<u>14,8</u>	18,8	25,8	19,5
9	36,0	29,7	26,5	24,5	24,2	25,1	31,9	33,5	<u>12,1</u>	40,1
10	32,4	25,6	28,8	37,5	40,8	33,0	13,9	<u>14,2</u>	33,7	<u>11,0</u>

### 3.3.3. ECART QUADRATIQUE DANS LE CAS DU TROISIEME FORMANT

Les tableaux 35 et 36 (page 57) représentent les résultats obtenus à partir, respectivement, des courbes de répartition et de probabilité.

### 3.3.4. ECART QUADRATIQUE DANS LE CAS DU FONDAMENTAL

Les tableaux 37 et 38 (page 58) représentent les résultats obtenus à partir, respectivement, des courbes de répartition et de probabilité.

### 3.3.5. DISCUSSION

Nous constatons tout d'abord que les résultats obtenus pour le deuxième formant sont meilleurs que ceux obtenus pour les premier et troisième formants. Les différences peuvent s'expliquer mais l'on ne peut toutefois conclure que le deuxième formant est plus particulièrement significatif du point de vue de l'identification.

La plage de variation du deuxième formant est très importante comparée à celles des premier et troisième formants et un examen des courbes de répartition représentées au premier chapitre montre les difficultés de comparaison des courbes relatives aux premier et troisième formants. Une erreur de mesure de 100 Hz sur les fréquences de formants modifie en effet notablement les courbes de répartition et de probabilité dans le cas de  $F_1$  et de  $F_3$ .

Nous allons comparer, tout en tenant compte de la remarque précédente, les résultats obtenus par la méthode exploitant les courbes de probabilité avec ceux obtenus par corrélation. Pour cela, nous calculons l'écart quadratique pour l'ensemble  $F_1-F_2$  et  $F_2-F_3$  (tableaux 39 et 40, page 59).

Dans le cas de l'étude des paramètres  $F_1$  et  $F_2$ , nous pouvons constater que les résultats obtenus sont les mêmes en valeur absolue. Il y a 13 locuteurs identifiés sur 20 (tableaux 22 et 39, pages 46 et 59), mais les erreurs ne concernent pas toujours les mêmes locuteurs (il est d'ailleurs très difficile d'interpréter ces différences), et la confiance dans l'identification est meilleure dans le cas de la comparaison dans le plan  $F_1-F_2$  (tableau 22, p. 46).

L'observation des tableaux 25 et 40 (pages 49 et 59) conduit aux mêmes conclusions avec un taux d'identification nettement inférieur dans le cas du tableau 40 (page 59).

TABLEAU 35 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de répartition dans le cas de  $F_3$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>18,1</u>	36,6	125,1	128,8	79,5	143,0	148,2	90,9	114,5	97,5
2	<u>13,5</u>	<u>18,6</u>	131,0	135,2	82,5	148,8	153,8	96,4	120,8	103,5
3	56,6	58,4	73,7	75,9	<u>22,3</u>	93,9	98,8	44,9	61,9	47,6
4	147,5	153,2	34,4	42,3	83,4	29,3	<u>12,9</u>	78,9	55,2	78,5
5	71,1	75,6	64,8	64,6	<u>12,6</u>	87,0	93,5	32,7	48,4	29,2
6	135,5	141,8	18,0	32,7	70,2	<u>16,8</u>	15,7	65,6	42,2	62,3
7	131,4	137,9	<u>11,3</u>	<u>28,9</u>	65,3	17,0	22,9	60,6	37,0	56,8
8	100,3	109,0	36,8	34,5	31,1	62,0	67,1	<u>21,8</u>	15,9	<u>15,0</u>
9	104,8	113,1	27,9	29,8	35,3	52,1	59,4	27,7	<u>13,4</u>	18,5
10	92,7	98,8	32,1	37,8	25,7	53,5	59,9	31,8	<u>26,3</u>	27,2

TABLEAU 36 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de  $F_3$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>19,4</u>	46,6	62,1	80,9	47,4	72,0	81,2	45,3	68,2	57,3
2	<u>18,4</u>	<u>29,4</u>	70,8	88,6	50,4	78,7	85,9	58,5	76,0	68,9
3	45,1	36,7	43,0	53,0	<u>14,6</u>	53,4	60,5	51,2	44,0	44,9
4	91,1	96,2	43,1	49,7	65,3	51,5	<u>15,7</u>	69,3	55,6	82,2
5	52,7	51,3	44,3	49,9	<u>15,7</u>	55,1	70,2	51,0	37,8	27,4
6	71,4	79,2	<u>15,9</u>	<u>32,3</u>	42,4	28,9	21,2	51,5	31,9	53,7
7	70,6	78,9	<u>10,1</u>	34,3	42,2	<u>20,8</u>	29,3	50,9	32,2	50,7
8	59,0	75,6	28,7	35,3	31,6	48,4	56,1	<u>31,8</u>	<u>16,3</u>	23,6
9	58,1	73,6	23,3	36,2	30,9	39,6	55,5	35,6	<u>18,9</u>	<u>19,7</u>
10	54,4	60,2	<u>18,5</u>	34,8	21,6	31,6	45,9	47,4	25,1	33,4

TABLEAU 37 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de répartition dans le cas de  $F_0$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>16,9</u>	172,0	250,0	155,0	135,0	47,5	38,0	120,0	145,0	105,0
2	203,0	<u>31,3</u>	108,0	73,6	100,0	202,0	214,0	118,0	91,0	147,0
3	297,0	170,0	<u>52,4</u>	211,0	232,0	298,0	307,0	244,0	277,0	281,0
4	149,0	80,0	192,0	42,0	24,6	148,0	157,0	34,6	<u>19,5</u>	65,5
5	153,0	65,0	184,0	28,0	<u>20,0</u>	147,0	162,0	39,7	<u>10,4</u>	73,2
6	114,0	90,0	196,0	63,0	40,5	111,0	125,0	<u>30,4</u>	50,4	<u>44,6</u>
7	<u>23,0</u>	181,0	266,0	158,0	135,0	<u>42,0</u>	<u>35,4</u>	118,6	146,0	98,7
8	166,0	66,0	131,0	<u>17,4</u>	29,0	158,0	175,0	59,0	<u>24,6</u>	91,5
9	176,0	52,0	173,0	<u>21,0</u>	48,0	171,0	186,0	74,0	38,0	106,5
10	163,0	66,0	185,0	23,0	26,0	<u>158,0</u>	<u>171,0</u>	53,0	<u>14,9</u>	85,0

TABLEAU 38 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de  $F_0$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>10,8</u>	51,5	45,8	54,3	54,1	<u>43,5</u>	<u>33,3</u>	52,1	53,9	53,8
2	53,6	<u>18,2</u>	37,7	31,1	37,0	60,5	52,0	43,7	34,8	52,9
3	56,9	49,6	<u>25,2</u>	58,1	60,8	64,0	55,5	59,5	60,7	64,0
4	64,1	51,8	60,4	43,4	43,7	71,0	59,9	40,3	<u>20,1</u>	49,7
5	54,0	28,2	48,2	26,8	25,3	62,4	50,7	<u>26,8</u>	<u>13,7</u>	37,4
6	42,2	33,6	42,8	35,0	34,6	44,9	38,7	<u>29,0</u>	32,4	<u>33,3</u>
7	<u>25,7</u>	56,4	56,0	51,5	50,7	52,8	36,9	52,6	52,9	53,9
8	66,9	51,3	60,9	<u>18,1</u>	<u>21,5</u>	72,8	61,1	50,8	38,2	59,1
9	64,8	40,6	57,5	<u>28,1</u>	35,5	69,6	61,2	50,3	40,8	60,9
10	60,2	34,2	54,5	<u>17,4</u>	24,9	67,1	55,8	41,7	22,4	49,7

TABLEAU 39 - Séquence S<sub>2</sub> ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de F<sub>1</sub> et F<sub>2</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>23,6</u>	52,4	60,9	49,9	41,2	45,5	80,5	84,3	68,4	79,3
2	<u>33,1</u>	42,1	49,1	45,4	33,5	37,4	77,3	89,5	67,3	81,8
3	49,4	<u>29,8</u>	<u>26,2</u>	40,6	64,8	53,3	33,6	41,5	28,0	41,6
4	44,0	34,3	38,1	<u>12,6</u>	38,5	34,1	61,9	65,6	35,5	65,8
5	42,7	55,0	58,8	41,0	<u>15,3</u>	31,8	93,3	100,7	71,1	98,1
6	33,8	43,1	49,0	35,7	24,4	<u>17,8</u>	77,3	88,5	62,9	80,7
7	48,5	25,1	<u>23,5</u>	38,7	61,1	44,1	40,9	51,9	39,4	44,9
8	65,7	47,8	45,6	60,6	87,6	65,7	<u>15,1</u>	31,7	36,8	21,9
9	68,2	41,3	43,0	51,5	82,5	66,1	37,4	35,8	<u>18,6</u>	42,6
10	68,1	50,0	49,6	63,0	90,8	69,4	<u>13,9</u>	<u>26,7</u>	42,0	<u>13,1</u>

TABLEAU 40 - Séquence S<sub>2</sub> ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de F<sub>2</sub> et F<sub>3</sub>.

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>29,5</u>	56,6	82,0	87,5	58,0	79,5	90,0	50,5	76,0	64,5
2	34,5	<u>31,6</u>	74,5	93,5	56,5	83,0	88,0	64,5	81,7	78,5
3	53,2	41,5	46,5	59,8	<u>29,6</u>	61,5	63,5	57,6	50,5	54,4
4	98,0	102,0	56,7	51,0	65,5	58,0	<u>40,5</u>	77,0	59,0	92,0
5	62,5	60,0	62,5	53,0	<u>19,2</u>	62,0	80,0	61,2	43,0	52,0
6	73,7	82,0	33,0	<u>38,0</u>	46,0	<u>32,0</u>	36,4	58,2	40,0	61,5
7	80,4	82,8	<u>25,5</u>	50,5	58,5	39,0	<u>30,0</u>	55,8	47,0	53,6
8	65,7	78,5	35,5	48,7	47,5	55,5	58,0	<u>37,0</u>	<u>30,6</u>	<u>30,6</u>
9	68,4	79,2	35,4	43,7	39,4	46,5	65,0	49,0	<u>22,4</u>	44,5
10	63,0	65,0	<u>34,2</u>	51,0	47,0	46,0	47,6	49,5	42,0	35,0

Les études sur les courbes de probabilité, dans les conditions de l'expérience, paraissent donc donner des résultats inférieurs à ceux obtenus dans les plans  $F_1-F_2$ ,  $F_2-F_3$ .

PRUZANSKY (1963) ayant effectué une intégration des données fréquentielles dans le temps applique un procédé se rapprochant de nos techniques d'exploitation des courbes de probabilité. Elle juge les résultats ainsi obtenus intéressants comparativement à d'autres obtenus en tenant compte explicitement du facteur temps. En effet, dans son expérience, l'exploitation directe, dans le temps, des sorties d'un banc de filtres conduit à des comparaisons difficiles.

En étudiant l'écart quadratique sur  $F_1$ ,  $F_2$  et  $F_3$  (tableau 41 ci-dessous), nous devrions obtenir un taux d'identification proche de celui de PRUZANSKY à savoir : 89 %.

Le "taux" que nous obtenons est nettement inférieur ; ce résultat semble indiquer que nos observations précédentes sur la nécessité de mesures très précises dans le cas des premier et troisième formants sont fondées.

TABLEAU 41 - Séquence  $S_2$  ;  
Ecart quadratique entre les courbes  
de probabilité dans le cas de  $F_1$ ,  $F_2$  et  $F_3$ .

	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'
1	<u>32,0</u>	69,5	92,5	95,0	63,0	85,0	114,0	96,0	95,0	98,0
2	<u>38,0</u>	51,6	86,0	99,5	59,0	87,0	115,0	107,0	101,0	107,0
3	64,5	<u>47,0</u>	50,3	66,6	66,0	75,5	69,5	66,0	52,4	61,0
4	101,0	102,0	57,5	<u>51,0</u>	76,0	62,5	72,0	95,0	66,0	105,0
5	67,8	75,0	73,6	54,6	<u>22,0</u>	67,0	117,0	113,0	80,4	102,0
6	79,0	90,5	51,4	<u>48,0</u>	49,0	<u>34,0</u>	80,0	102,0	70,5	97,0
7	85,8	82,9	<u>25,5</u>	51,7	74,5	48,8	51,0	73,0	51,0	67,8
8	88,5	89,5	53,8	69,0	93,0	81,5	58,0	<u>45,0</u>	40,2	<u>32,2</u>
9	89,6	87,5	49,0	63,0	88,0	77,0	89,0	50,5	<u>26,5</u>	47,0
10	87,0	78,0	53,0	72,0	94,0	76,0	<u>47,6</u>	54,5	48,7	<u>36,0</u>

### 3.4. CONCLUSIONS

Il n'était pas dans notre intention d'élaborer une méthode d'identification des locuteurs mais plutôt d'étudier le rôle des formants et, accessoirement, de la fréquence fondamentale comme paramètres porteurs d'informations spécifiques d'un locuteur donné.

Par ailleurs, nos travaux se sont limités à l'étude de dix locuteurs seulement. Aussi évitons-nous, à leur sujet, de parler de taux d'identification.

Nous avons cherché à déterminer des tendances et nous ne comparerons pas nos résultats avec ceux obtenus par d'autres auteurs bien qu'ils aient souvent fondé leurs travaux sur un petit nombre de données, mais nous nous permettrons certains rapprochements.

Rappelons que l'essentiel de notre travail a été effectué sur une très courte séquence de parole et qu'à partir de celle-ci des résultats significatifs ont été obtenus quoique, d'après *POLLACK (1954)*, il faille au moins 1,2 seconde de signal pour effectuer dans de bonnes conditions une identification par des auditeurs ; cette durée a aussi été proposée par *STEVENS et al. (1968)* pour des études sur sonagrammes.

Les résultats obtenus au moyen de la première méthode - étude dans les plans  $F_1-F_2$ ,  $F_2-F_3$ ,  $F_0-F_3$  - sont bien meilleurs que ceux obtenus à partir des courbes de répartition ou de probabilité.

En appliquant cette première méthode, nous conservons des données sur les transitions que nous pouvons facilement comparer. On peut penser que lorsque le débit de parole du locuteur augmente les transitions deviennent plus rapides d'où il pourrait résulter de notables modifications, pour un même locuteur, des tracés dans l'un des plans considérés. En fait, il n'en est rien (*T. GAY - 1968*), les réductions de durée se font essentiellement au niveau des voyelles donc des états stables. Seul un débit très accéléré modifie les évolutions des formants ; dans ce cas, les états stables correspondant aux voyelles ne sont pas toujours atteints (*LINDBLOM - 1963*).

Par ailleurs, il ne semble pas que le paramètre temps soit très important pour l'identification (*S. PRUZANSKY - 1963*).

On peut aussi remarquer que l'application de la première méthode conduit à des résultats corrects malgré les erreurs inévitables de mesure sur les formants et malgré les différences existant entre diverses prononciations d'un même échantillon par un locuteur donné.

Pour localiser l'origine de l'information spécifique d'un locuteur donné, il faudrait faire des essais plus systématiques mais nous pouvons déjà noter que les paramètres  $F_2$  et  $F_3$  sont les plus significatifs sans toutefois pouvoir préciser d'une manière sûre l'apport respectif de chacun d'eux. Les renseignements obtenus à partir des courbes de répartition ne peuvent nous conduire à reconsidérer ce résultat, même s'ils étaient fondés sur l'exploitation de données très précises. En effet, les transitions du troisième formant qui, dans le cas des courbes de répartition, n'interviennent pas, peuvent jouer un rôle essentiel dans l'identification.

POLLACK (1954) montre qu'un signal de parole dont la bande de fréquences est limitée entre 500 et 5000 Hz est aussi facilement identifié par des auditeurs que lorsque cette bande n'est pas limitée. Mais peut-être existe-t-il des paramètres physiques significatifs et exploitables objectivement qui soient sans signification pour l'auditeur et son système de perception.

Si nous voulons identifier objectivement un locuteur, avec un taux d'identification de 99,9 %, nous devons traiter plus finement les paramètres auxquels un auditeur est sensible, voire exploiter d'autres paramètres car les possibilités d'identification pour un auditeur sont assez limitées.

Les résultats que nous avons obtenus à partir des plans  $F_0-F_3$ ,  $F_1-F_2$ ,  $F_2-F_3$  sont très bons (tableau 30, page 52) bien que le nombre de paramètres exploités soit réduit. Avec 9 paramètres, le taux d'identification obtenu par WOLF (1970) est de 98,5 % ; avec 17 paramètres, le taux est de 100 %.

Une meilleure confiance dans l'identification serait peut-être obtenue en traitant certains facteurs particuliers : fréquence du quatrième formant, durées des consonnes nasales, des fricatives, spectre des fricatives, etc... mais il paraît aussi nécessaire d'améliorer les techniques de comparaison, lesquelles, dans notre étude, sont du domaine de la reconnaissance des formes.

Plutôt qu'étudier ces facteurs particuliers, il nous paraît actuellement plus opportun d'augmenter notablement le nombre de nos locuteurs et de préciser les résultats obtenus avec un locuteur dont nous ne posséderions pas d'échantillon de référence. Le tableau 30 nous montre déjà qu'il est possible de préciser la valeur d'un seuil permettant une identification absolue.

## - CHAPITRE 4 -

## ETUDE SUR LES TRANSITIONS D'AMPLITUDE

## 4.1. INTRODUCTION

Nous avons vu que, dans le cas des voix de femmes, le spectre du signal de parole est "pauvre" et, par suite, les formants sont difficiles à détecter. Mais, le phénomène à analyser étant essentiellement transitoire, en particulier sur les plans de l'amplitude et de la fréquence de la source vocale (en transitoire, l'étude du signal émis doit rendre compte de l'ensemble de la fonction de transfert du conduit vocal), on peut envisager d'étudier ces transitions au niveau de la perception.

Dans ce chapitre, nous nous sommes limités à l'étude de la transition d'amplitude sur le plan de la perception.

Lors de la production de la parole, au début de la phonation, le signal au niveau de la source vocale subit une variation d'amplitude plus ou moins importante, cette variation se retrouvant déformée dans le signal émis (les cavités jouant le rôle de filtres).

Au cours de cette transition d'amplitude, le conduit vocal est totalement excité et le spectre du signal de sortie est continu avant d'atteindre progressivement sa structure harmonique. Nous avons abordé l'étude de ce phénomène, sur le plan de la perception, à l'aide d'un synthétiseur de parole. A notre connaissance, aucun travail sur ce sujet n'a encore été publié.

#### 4.2. MODE OPERATOIRE

Dans un synthétiseur à formants dont la fonction de transfert correspond à celle du conduit vocal, nous retrouvons le phénomène évoqué précédemment si les circuits de formants sont attaqués par des impulsions dont nous modifions brusquement l'amplitude.

Pour produire cette variation, nous utilisons une porte placée à l'entrée des circuits de formants. Si cette porte est placée en sortie, les circuits de formants ne sont pas excités par la variation d'amplitude et fonctionnent en régime permanent.

Dans le cas de la production de voyelles, nous pourrions obtenir deux types de signaux de même durée dont les spectres en régime établi sont les mêmes mais dont les spectres au cours de la transition sont différents. Dans le premier cas, les circuits de formants étant soumis à la variation d'amplitude, le signal de sortie, au moment de la transition, dépend de la fonction de transfert globale du synthétiseur ; le spectre est continu. Dans le deuxième cas, le spectre du signal de sortie est toujours constitué d'harmoniques, la transition se traduisant, au niveau de la perception, en particulier par des distorsions d'amplitude sans aucun rapport avec la réponse du synthétiseur.

La figure 42 (page 65) représente le schéma synoptique de l'expérience.

Le synthétiseur à formants a été décrit par ailleurs (PAILLÉ, BEAUVIALA, CARRÉ - 1970). La porte est un multiplieur dont le signal de commande  $V_c$  est un signal rectangulaire de durée déterminée correspondant à la durée de la voyelle à étudier et de temps de montée et de descente réglables.

Avec l'appareillage ainsi mis en oeuvre, nous pouvons élaborer des voyelles de synthèse qui seront testées par des auditeurs.

Les voyelles que nous avons synthétisées sont au nombre de 10 et sont facilement discernables les unes des autres, exceptées les voyelles [ɑ] et [a] ou bien [ø] et [œ], pour certains auditeurs.

Les formants de nos voyelles sont ceux adoptés, en général, pour la langue française dans le cas des voix d'hommes (voir tableau 43, page 65). Chacune des voyelles synthétisées est étudiée dans trois cas correspondant chacun à une valeur particulière de la fréquence fondamentale  $F_0$ . Les valeurs adoptées pour  $F_0$  sont : 100, 200 et 300 Hz.

FIGURE 42

Schéma synoptique de l'expérience.

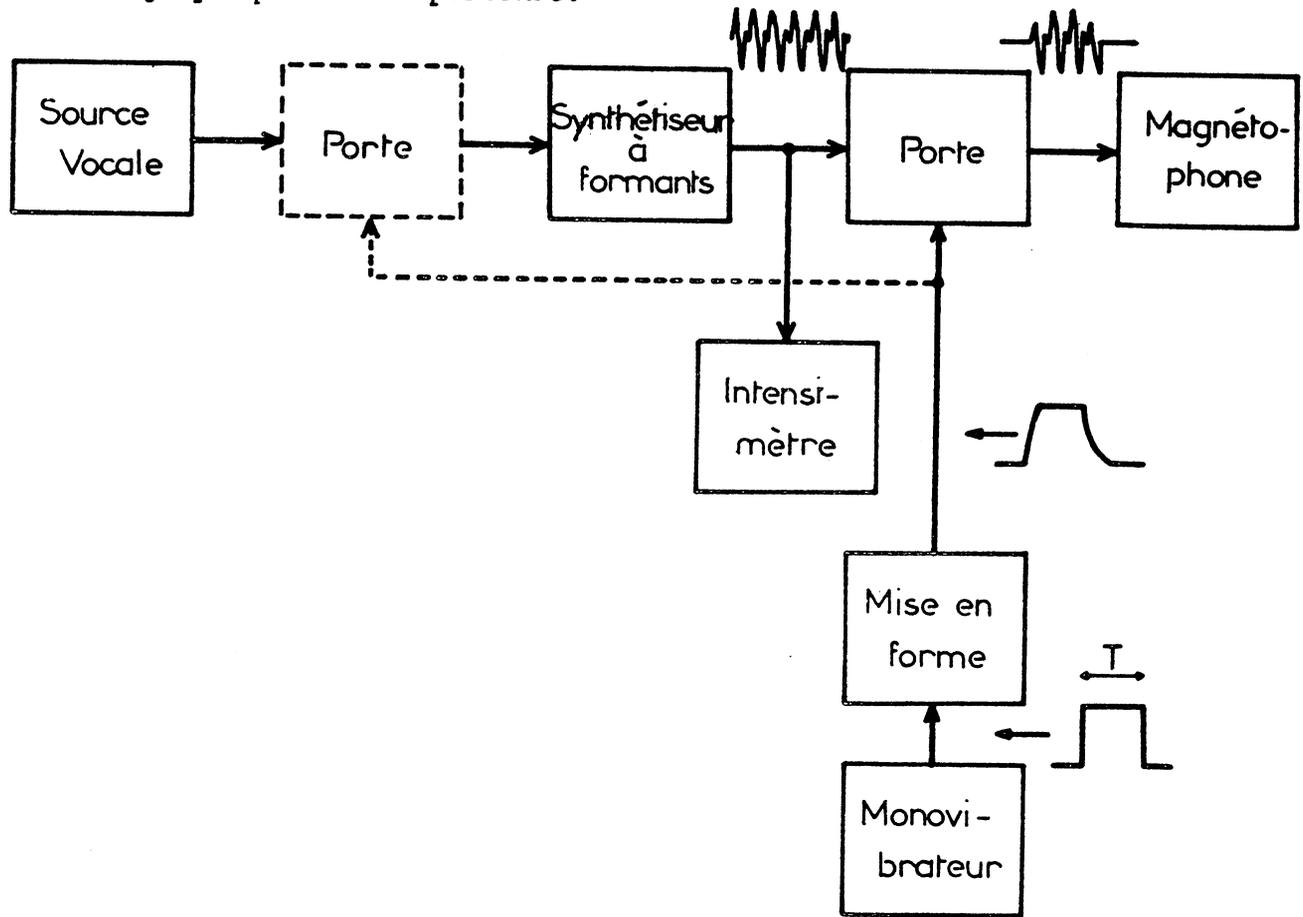


TABLEAU 43

Fréquences des trois premiers formants des voyelles synthétisées.

	[u]	[o]	[ø]	[oe]	[a]	[ɑ]	[e]	[ε]	[y]	[i]
F <sub>1</sub>	250	350	320	550	850	800	380	550	250	210
F <sub>2</sub>	750	700	1500	1400	1300	1000	2500	2000	2000	2900
F <sub>3</sub>	2500	2800	2300	3000	3700	3000	2900	3000	2100	3400

La porte de commande d'amplitude est disposée soit à l'entrée, soit en sortie des circuits de formants. Nous avons donc 60 sons différents, de même durée T, enregistrés dans un ordre aléatoire.

Nos voyelles sont écoutées au casque par des auditeurs témoins selon le programme suivant : la voyelle n° 1 est produite deux fois successivement, un

écart d'une seconde séparant les deux émissions ; l'auditeur dispose ensuite de deux secondes avant l'émission de la voyelle n° 2 pour indiquer, parmi les dix voyelles, celle qu'il a perçue.

Ce type de test est intéressant car il permet une étude plus complète, comme il sera précisé, que celle du taux de reconnaissance de voyelles de synthèse dont les caractéristiques (formants et fréquence fondamentale) correspondraient dans tous les cas à celles des voyelles naturelles.

Par ailleurs, rappelons que dans le cas d'une mélodie élevée ( $F_0 = 300$ ); le nombre de composantes spectrales est faible.

#### 4.3. RESULTATS

Nous avons représenté tableau 44 le nombre d'erreurs obtenu lorsque la durée des voyelles est de 100 ms et celle des constantes de temps de montée et de descente du signal de commande de 10 ms.

Pour cette expérience, les tests ont été effectués par 10 auditeurs ; nous avons donc disposé de 600 tests élémentaires. Dans notre tableau récapitulatif, chaque terme d'erreur est relatif à 100 tests élémentaires.

TABLEAU 44

	$F_0 =$			TOTAL
	100 Hz	200 Hz	300 Hz	
Porte à l'entrée des circuits de formants	10 (sur 100)	8	36	54 (sur 300)
Porte en sortie des circuits de formants	18	9	25	52

L'examen de ce tableau révèle que, pour  $F_0 = 100$  Hz, il y a moins d'erreurs lorsque la porte est à l'entrée des circuits de formants que lorsqu'elle est en sortie. Par contre, pour  $F_0 = 300$  Hz, le nombre d'erreurs est plus important lorsque la porte est à l'entrée des circuits de formants. Ces résultats semblent indiquer que la transition d'amplitude apporte, lorsqu'elle est effectuée avant les circuits de formants, des informations complémentaires, soit par enrichissement du spectre, soit en rendant les voyelles plus naturelles.

Les formants de nos voyelles de synthèse étant ceux d'une voix d'homme, avec  $F_0 = 100$  Hz le taux d'erreur est plus faible lorsque la porte est montée à l'entrée des circuits de formants. Avec cette fréquence fondamentale pourtant, le spectre des voyelles, en régime stationnaire, est suffisamment riche et les formants doivent pouvoir être détectés avec précision. On peut alors penser qu'un apport de naturel résultant du régime transitoire des circuits de formants facilite la reconnaissance ou bien que la quantité d'informations exploitée par notre organe de perception ne dépend pas uniquement de la richesse spectrale.

WAJSKOP (1967) obtient une meilleure identification des voyelles prononcées par une femme comparativement à celle des voyelles prononcées par un homme ; le taux d'erreurs varie du simple au double. La durée des voyelles testées étant la même dans les deux cas, il pense que l'on dispose, dans le premier cas, d'un nombre d'informations plus élevé que dans le second, ceci pouvant découler d'un nombre de périodes du signal plus grand (fréquence plus élevée).

Dans nos tests, avec la fréquence fondamentale de 300 Hz, le taux d'erreurs est plus grand lorsque la porte est montée à l'entrée des circuits de formants.

Ce résultat peut être interprété. En effet, ainsi que nous l'avons exposé au paragraphe 2.3. (page 35), les fréquences des formants d'une voyelle dans le cas des voix de femmes sont supérieures en moyenne de 15 % à celles des formants dans le cas des voix d'hommes. Or, pour  $F_0 = 300$  Hz, la fréquence fondamentale correspond à celle d'une voix de femme et un auditeur annoncera une voyelle, si elle existe, qui sera dotée de fréquences de formants (la référence étant la voix d'homme) inférieures de 15 % à celles indiquées tableau 43 (page 65).

Il est donc normal de trouver des erreurs et, lorsque la porte est montée à l'entrée des circuits de formants, des informations complémentaires facilitent ces erreurs. Mais, inversement, il est intéressant de noter que, la porte étant montée à la sortie du synthétiseur et bien que la mélodie soit élevée, le taux d'erreur reste inférieur à celui obtenu porte à l'entrée.

Lorsque la fréquence fondamentale est de 200 Hz, nous nous trouvons à la transition permettant de passer d'une interprétation à une autre. Par ailleurs, on constate que, pour cette fréquence, le taux d'erreur est le plus faible.

Ce résultat est, a priori, surprenant. Une interprétation peut cependant être donnée : admettant que, dans ce cas, la fréquence fondamentale n'est pas suffisamment caractéristique d'une voix de femme, nous retenons comme seul fait nouveau par rapport aux tests effectués avec  $F_0 = 100$  Hz, le doublement de la fréquence fondamentale ; selon l'hypothèse de *WAJSKOP*, on fournirait alors à l'auditeur plus d'informations et, par là, une meilleure possibilité d'identification.

Le tableau 45 représente les résultats que nous avons obtenus lors d'une étude préliminaire effectuée à partir de onze voyelles. Les résultats obtenus sont intéressants à rapporter car ils confirment ceux donnés tableau 44. Dans cette expérience, les tests ont été effectués par six auditeurs.

TABLEAU 45

Porte à l'entrée des circuits de formants

Porte en sortie des circuits de formants

	$F_0 =$			TOTAL
	100 Hz	200 Hz	300 Hz	
Porte à l'entrée des circuits de formants	13 (sur 60)	10	30	53 (sur 180)
Porte en sortie des circuits de formants	18	10	26	54

Dans le but de compliquer le processus de reconnaissance et pour rendre moins naturel l'ensemble des sons à tester, nous avons réduit la longueur des voyelles jusqu'à 10 ms. Dans cette expérience, la constante de temps du circuit de commande était de 2 ms.

Nous avons représenté tableau 46 (page 69) les erreurs obtenues avec vingt auditeurs.

Là encore, les résultats correspondent à ceux obtenus précédemment mais, ici, il est plus difficile de parler de naturel pour les voyelles testées dans cette expérience.

TABLEAU 46

	F <sub>0</sub> =			TOTAL
	100 Hz	200 Hz	300 Hz	
Porte à l'entrée des circuits de formants	37 (sur 200)	31	70	138 (sur 600)
Porte en sortie des circuits de formants	62	40	54	156

#### 4.4. CONCLUSIONS

Il semble donc que la transition d'amplitude au niveau du signal de source, c'est-à-dire à l'entrée de notre conduit vocal, apporte des informations complémentaires qui facilitent l'identification, mais on ne peut préciser sous quelle forme ces informations sont détectées par notre organe de perception.

Il serait intéressant, pour compléter les renseignements que nous avons obtenus, de répéter les expériences décrites à partir de données sur les formants de voyelles prononcées par des femmes.

Il serait également intéressant, pour mieux étudier les origines du minimum d'erreurs observé à F<sub>0</sub> = 200 Hz, de reprendre nos tests de perception pour des fréquences fondamentales de 150 à 250 Hz.

Par ailleurs, il ne semble pas que les formants doivent être définis avec grande précision dans le cas de l'identification des voyelles, bien que la transition d'amplitude apporte des informations complémentaires sur ces formants. Ces informations ne semblent pas être traitées par notre organe de perception comme le ferait un analyseur spectral classique.

Dans un ensemble de reconnaissance automatique de la parole, il serait peut-être utile de tenir compte de ce résultat.

Par contre, dans le cas de la commande d'un synthétiseur à formants, nous ne devons pas nous inspirer uniquement du comportement de notre organe de perception pour détecter automatiquement les formants.



## C O N C L U S I O N

Les travaux présentés dans ce mémoire nous ont conduits à poser de nouveaux problèmes, lesquels sont évoqués dans les conclusions des divers chapitres. Nous nous proposons d'étudier ultérieurement ces problèmes d'une manière systématique à la lumière des résultats que nous avons obtenus.

Ces résultats nous ont permis de préciser le rôle et l'importance des formants comme paramètres porteurs directs ou indirects d'informations sur le message proprement dit et sur les caractéristiques individuelles de la voix. L'étude du rôle et de l'importance des formants sur le plan de la perception a été abordée.

Les résultats de nos travaux doivent faciliter l'élaboration d'un détecteur automatique de formants. Par ailleurs, ils apportent des éclairages nouveaux sur des problèmes concernant la reconnaissance de la parole.

Les études sur le rôle et l'importance des formants comme porteurs d'informations sur le message et sur les caractéristiques du locuteur nous ont permis d'observer que :

- . la loi de répartition des fréquences des formants varie d'un locuteur à un autre (étude statistique) ;
- . l'identification d'un locuteur peut être envisagée à partir des évolutions des formants ; des résultats significatifs ont été obtenus à partir des formants  $F_2$  et  $F_3$  ;
- . les évolutions des formants  $F_1$  et  $F_2$  réputés porteurs d'informations sur le message proprement dit sont aussi caractéristiques, dans une certaine mesure, du locuteur :

Sur le plan de la perception, nos premiers résultats montrent que :

- . les transitions d'amplitude apportent des informations complémentaires facilitant l'identification ;
- . si l'enrichissement du spectre, résultant des transitions d'amplitude, permet de mieux préciser la structure formantique, il ne semble pas cependant que notre organe de perception exploite directement cet enrichissement spectral.

Pour l'élaboration d'un détecteur automatique des formants, les résultats obtenus au premier chapitre sont très utiles. L'étude statistique nous a montré que les plages des formants sont bien déterminées, que la distance entre les formants, à un instant donné, passe par un minimum et que la vitesse d'évolution des formants est limitée.

Par ailleurs, l'existence de liens entre  $F_1$  et  $F_2$  d'une part et  $F_3$  d'autre part, nous conduit à penser que le troisième formant pourrait être calculé au moyen d'une expression mathématique fonction du locuteur. Cette possibilité pourrait rendre inutile la détection du troisième formant et, en conséquence, permettrait de réduire la capacité d'une ligne de transmission d'un vocoder à formants.

En reconnaissance de parole, nous pourrions tirer parti des résultats obtenus aux chapitres 2 et 3. Ces résultats montrent qu'une normalisation des données formantiques doit tenir compte de la fréquence fondamentale du signal et du fait que les paramètres  $F_1$  et  $F_2$  sont aussi porteurs d'informations sur les caractéristiques du locuteur.

## ANNEXE 1

## RAPPELS SUR LA THEORIE DU FONCTIONNEMENT DE L'APPAREIL VOCAL

La parole, en tant que phénomène physique, résulte de l'excitation des cavités de l'appareil vocal, soit par une source d'impulsions pour les sons vocaux, soit par une source de bruit pour les sons non vocaux. Dans le premier cas, la source est constituée par l'ensemble poumons - cordes vocales ; dans le deuxième cas, le bruit est produit par l'action du courant d'air sur certaines parois de l'appareil vocal (à l'endroit d'un resserrement du conduit vocal par exemple). Ces deux phénomènes peuvent intervenir simultanément avec des intensités relatives plus ou moins grandes.

En parallèle sur le conduit vocal est greffée la cavité nasale, laquelle est limitée d'un côté par les narines et de l'autre par le voile du palais. Ce dernier règle le couplage acoustique entre le conduit vocal et le conduit nasal.

Nous allons brièvement indiquer les caractéristiques principales des sources ainsi que celles du conduit vocal.

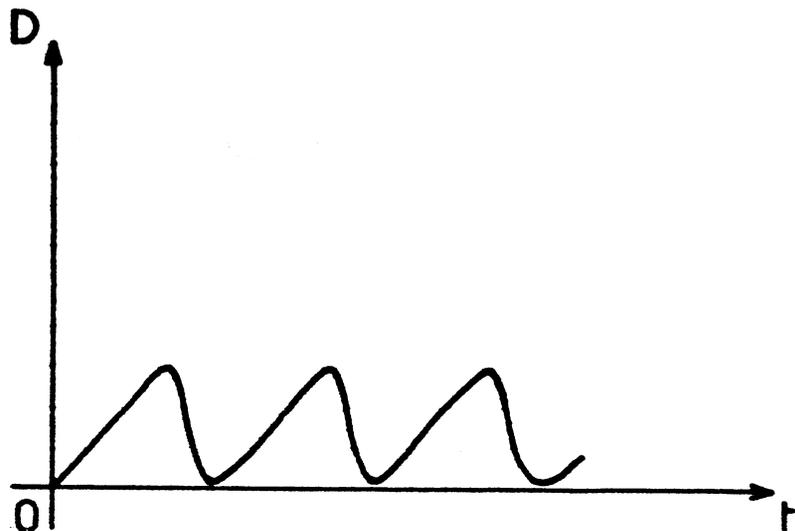
## SOURCE VOCALE ET SOURCE DE BRUIT

La source vocale est essentiellement constituée par les poumons qui compriment l'air à l'intérieur des bronches et de la trachée, et par les cordes vocales composées de deux muscles plus ou moins tendus côte à côte dans la phase de phonation.

Ces deux muscles se séparent et se rapprochent alternativement sous l'action de la pression de l'air à l'intérieur de la trachée. La fréquence de ce phénomène de relaxation est essentiellement déterminée par la tension et la masse des cordes vocales ainsi que par la pression subglottique.

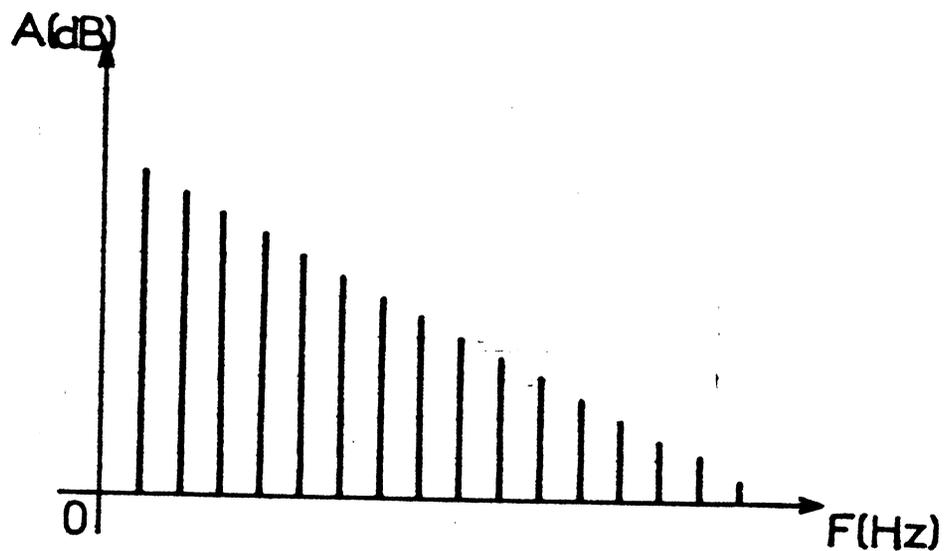
Le débit d'air  $D$  au niveau des cordes vocales est donc modulé à la fréquence de vibration de ces dernières (figure 47).

FIGURE 47



La forme du signal image du débit ne subit pas de modifications importantes en cours de phonation et l'enveloppe du spectre de ce signal de source présente un affaiblissement de l'ordre de 12 dB/octave (figure 48).

FIGURE 48



L'impédance interne de la source vocale est essentiellement variable et supérieure à 1000 ohms acoustiques. L'impédance de charge apportée par le conduit vocal étant inférieure à 100 ohms, on considère, en première approximation, la source

vocale comme étant un générateur de grande impédance interne, donc non perturbé par l'évolution des cavités de la bouche (FLANAGAN - 1965).

Le bruit peut être provoqué par un rétrécissement du conduit vocal, dans la mesure où certaines conditions de pression interne et de section du conduit sont réalisées.

La source de bruit est localisée sur le parcours du conduit vocal et sa situation peut varier avec le son prononcé.

L'impédance interne de cette source est considérée comme faible et purement réelle. Quant au spectre, on s'accorde à le trouver plat aux moyennes fréquences, l'affaiblissement se manifestant côté basses fréquences au-dessous de 1 kHz et côté hautes fréquences au-dessus de 8 kHz (FLANAGAN - 1965).

#### CONDUIT VOCAL

Le calcul de la fonction de transfert du conduit vocal  $T(p) = D_s(p)/D_e(p)$  ( $D_s(p)$  et  $D_e(p)$  sont les transformées de Laplace du débit en sortie  $D_s(t)$  et du débit à l'entrée  $D_e(t)$  respectivement) a été effectué par G. FANT (1960).

Dans le cas de pertes faibles et dans le cas d'une propagation suivant une seule dimension, la fonction de transfert peut être mise sous la forme :

$$T(p) = \prod_{n=1}^{\infty} \frac{1}{\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1} \quad \text{avec } \tau_n = \frac{1}{\omega_n} \quad (\omega_n \text{ est une pulsation propre}).$$

Cette expression peut être simplifiée en ne tenant compte que des quatre premiers termes du produit et en faisant intervenir un terme correctif :

$$T(p) \approx \prod_{n=1}^4 \frac{1}{\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1} \quad Cr_4 .$$

Le terme correctif  $Cr_4$  a été déterminé par FANT ; il correspond à la fonction de transfert d'un circuit résonnant et d'un filtre passe-haut montés en cascade.

Les cavités du conduit vocal modifient le spectre du signal transmis en favorisant certaines composantes qui correspondent aux fréquences des pôles de la fonction de transfert  $T(p)$ .

Sur la figure 49.a, nous avons représenté le module de la fonction de transfert du conduit vocal et sur la figure 49.b, le spectre du signal résultant de l'action de la source sur le conduit vocal.

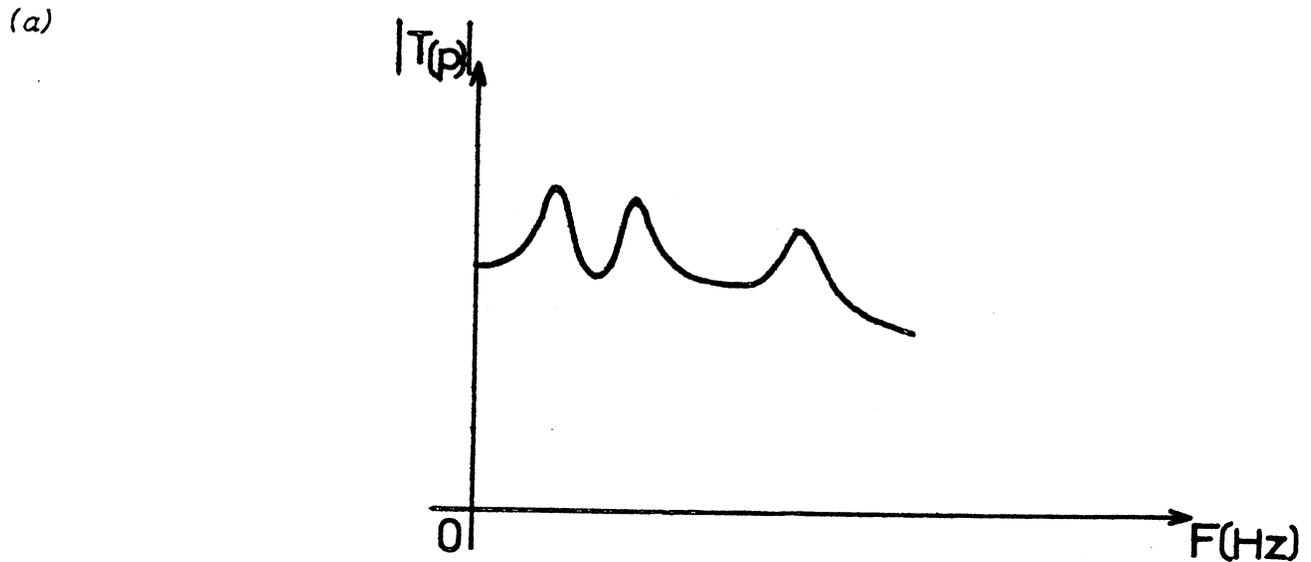
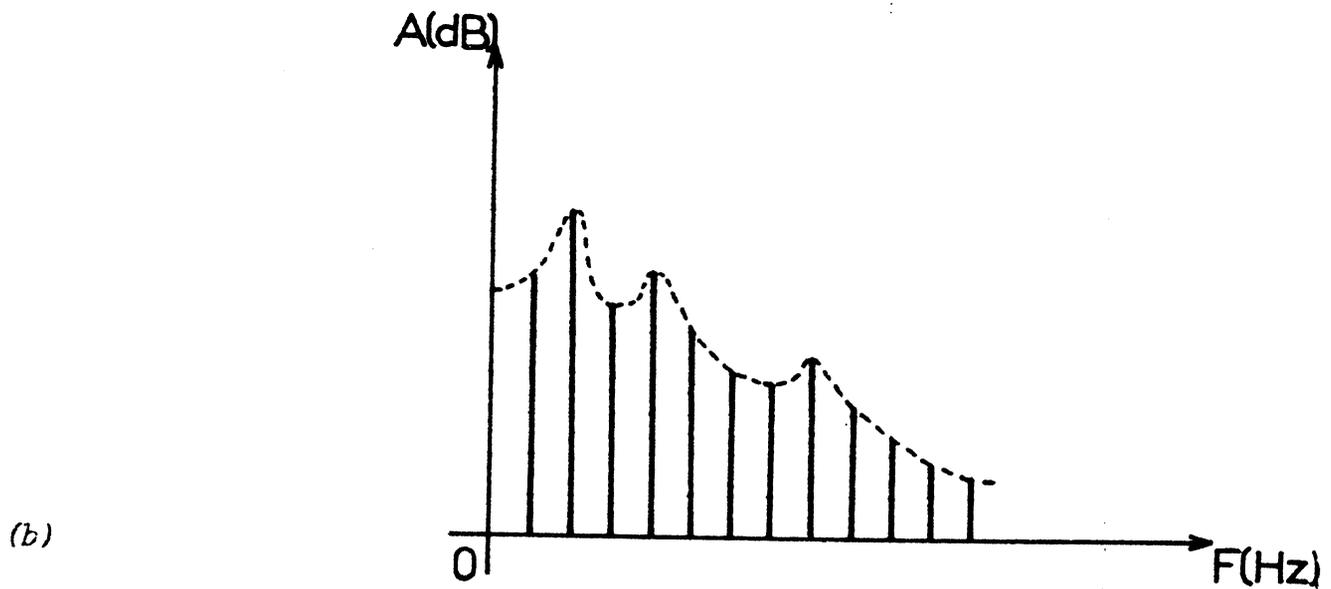


FIGURE 49



Les régions fréquentielles correspondant à des maximums d'amplitude sont appelées formants. La fréquence  $F_n$  d'un formant est l'une des fréquences de résonance du conduit vocal et correspond à l'un des pôles de la fonction de transfert  $T(p)$ . On a donc :  $F_n = \frac{1}{2 \pi \tau_n}$ .

De même, la bande passante "à 3 dB", exprimée en Hertz, d'un formant est :

$$\Delta F_n = \frac{\zeta_n}{\pi \tau_n}$$

Cette bande passante est, en première approximation, constante dans la plage de variation d'un formant déterminé.

Dans le cas de production de bruit, la fonction de transfert à considérer est de la forme :

$$T(p) = \prod_{n=1}^{\infty} \frac{1}{\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1} \times \left( \prod_{m=1}^{\infty} \tau_m^2 p^2 + 2 \zeta_m \tau_m p + 1 \right) \times (p - \tau_g).$$

L'apparition de zéros dans la fonction de transfert est due aux cavités en amont de la source de bruit. Le terme  $\tau_g$  est un zéro réel qui correspond aux pertes glottiques.

Dans le cas d'une excitation des cavités nasales, la fonction de transfert peut se mettre sous la forme suivante :

$$T(p) = K \prod_{n=1}^f \frac{(\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1)}{(\tau'_n{}^2 p^2 + 2 \zeta'_n \tau'_n p + 1)} \frac{Cr_g}{\prod_{m=1}^g (\tau_m^2 p^2 + 2 \zeta_m \tau_m p + 1)}.$$

Les fonctions de transfert évoquées précédemment permettent la détermination du débit  $D_s$  au niveau des lèvres en fonction du débit de la source.

Normalement, dans le cadre du processus phonation-audition, l'oreille réceptrice n'est pas excitée par le débit  $D_s$  au niveau des lèvres du locuteur mais par la pression  $P_s$  qui en résulte, au niveau de l'oreille, à une distance  $l$  des lèvres. Il apparaît donc nécessaire d'associer aux fonctions de transfert précédentes celle de rayonnement  $P_s(p)/D_s(p)$  de la bouche.

Si on assimile la tête à une sphère, on a :  $P_s/D_s = \frac{\rho p}{4 \pi l} Cr(p)$ ,  $Cr(p)$  étant équivalent à une fonction de transfert comportant un pôle à 1000 Hz et un zéro à 500 Hz.



## ANNEXE 2

TABLEAU DE CORRESPONDANCE des symboles que nous avons utilisés  
pour représenter les voyelles.  
Ces symboles sont extraits de l'Alphabet  
de l'Association Phonétique Internationale (I.P.A.).

[i]	comme dans	ni	;
[e]	"	dé	;
[ɛ]	"	fait	;
[a]	"	part	;
[ɑ]	"	pas	;
[ɔ]	"	tort	;
[o]	"	tôt	;
[u]	"	tout	;
[ø]	"	peu	;
[y]	"	pu	;
[ə]	"	de	;
[œ]	"	seul	;

(in Manual of Phonetics - éd. L. KAISER - 1957).



## B I B L I O G R A P H I E

- ARNOLD G.E.  
Morphology and physiology of the speech organs  
Manual of Phonetics - éd. L. KAISER, 31-64 (1957)
- BOLT R.H., COOPER F.S., DAVID E.E., DENES P.B., PICKETT J.M., STEVENS K.N.  
Speaker identification by speech spectrograms : A scientists' view of its  
reliability for legal purposes  
J.A.S.A., 47, 597-612 (1970)
- BRICKER P.D., PRUZANSKY S.  
Effects of stimulus content and duration on talker identification  
J.A.S.A., 40, 1441-1449 (1966)
- CARRÉ R.  
Analyse de la parole : détection automatique des trois premiers formants  
Acustica, 16, 243-247 (1966)
- CARRÉ R.  
Etude et réalisation d'un détecteur, à fonctionnement continu, automati-  
que des paramètres de la parole  
Thèse de Docteur-Ingénieur - Grenoble (1966)
- CARRÉ R.  
Etude statistique sur les formants vocaux  
Revue d'Acoustique, 3/4, 230-234 (1968)
- CARRÉ R., LANCIA R., WAJSKOP M.  
Sur la production des voyelles par des locuteurs hommes et femmes  
6ème Congrès International d'Acoustique - Tokyo (1968)
- CARRÉ R.  
Identification des locuteurs : exploitation des données relatives aux  
fréquences des formants  
7ème Congrès International d'Acoustique - Budapest (1971)
- CHIBA T., KAJIYAMA M.  
The vowel : its nature and structure  
Tokyo - KAISEIKAN (1941)
- DELATTRE P.  
Un triangle acoustique des voyelles orales du français  
French Review, 21 (1948)
- DUNN H.K.  
The calculation of vowel resonances, and an electrical vocal tract  
J.A.S.A., 22, 740-753 (1950)
- FANT G.  
Acoustic analysis and synthesis of speech with applications to swedish  
Ericson Technics, 1, 1-108 (1959)

- FANT G.*  
Acoustic theory of speech production  
S'Gravenhage - Mouton et Cie (1960)
- FANT G.*  
Formants and cavities  
Proc. of the 5th Cong. of Phonetics Sciences - Münster (1964)
- FANT G.*  
A note on vocal tract size factors and non-uniform F-Pattern scalings  
Speech Transmission Laboratory, Q.P.S.R. 4/1966 - Stockholm
- FLANAGAN J.L.*  
Automatic extraction of formant frequencies from continuous speech  
J.A.S.A., 28, 592-596 (1956)
- FLANAGAN J.L.*  
Speech analysis, synthesis and perception  
Springer Verlag (1965)
- GAY T.*  
Effect of speaking rate on diphthong formant movements  
J.A.S.A., 44, 1570-1573 (1968)
- GAUDE M., PULVERIC F., VALET J.Y.*  
Simulateur vocal  
Rapport interne (1966)
- GLENN J.W., KLEINER N.*  
Speaker identification based on nasal phonation  
J.A.S.A., 43, 368-372 (1968)
- HARGREAVES W.A., STARKWEATHER J.A.*  
Recognition of speaker identity  
Language and Speech, 6, 63-67 (1963)
- KERSTA L.G.*  
Voiceprint identification  
Nature, 196, 1254-1257 (1962)
- LADEFOGED P., VANDERSLICE R.*  
The "Voiceprint" Mystique  
Working Papers in Phonetics - Dept Linguistics - Univ. Calif. Los Angeles,  
126-142 (nov. 1967)
- LI K.P., DAMMANN J.E., CHAPMAN W.D.*  
Experimental studies in speaker verification, using an adaptative system  
J.A.S.A., 40, 966-978 (1966)
- LINDBLOM B.*  
Spectrographic study of vowel reduction  
J.A.S.A., 35, 1773-1781 (1963)
- LUCK J.E.*  
Automatic speaker verification using cepstral measurements  
J.A.S.A., 46, 1026-1032 (1969)
- MEYER-EPPLER W.*  
Realization of Prosodic features in whispered speech  
J.A.S.A., 29, 104-106 (1957)
- PAILLE J., BEAUVIALA J.P., CARRE R.*  
Synthèse de la parole : description et utilisation d'un synthétiseur du  
type "à formants"  
Revue de Physique Appliquée, 5, 785-793 (1970)
- PAUL A.P., HOUSE A.S., STEVENS K.N.*  
Automatic reduction of vowel spectra - An analysis by synthesis method and  
its evaluation  
J.A.S.A., 36, 303-308 (1964)
- PETERSON G.E., BARNEY H.L.*  
Control methods used in a study of the vowels  
J.A.S.A., 24, 175-184 (1952)

POLLACK I., PICKETT J.P., SUMBY W.H.

On the identification of speakers by voice

J.A.S.A., 26, 403-406 (1954)

PRUZANSKY S.

Pattern matching procedure for automatic talker recognition

J.A.S.A., 35, 354-358 (1963)

PRUZANSKY S., MATHEWS M.V.

Talker-recognition procedure based on analysis of variance

J.A.S.A., 36, 2041-2047 (1964)

SCHROEDER M.R., NOLL A.M.

Recent studies in speech research at Bell Telephone Laboratories

5ème Congrès International d'Acoustique - Liège (1965)

STEVENS K.N., KASOWSKI S., FANT G.

An electrical analog of the vocal tract

J.A.S.A., 25, 734-742 (1953)

STEVENS K.N., WILLIAMS C.E., CARBONELL J.R., WOODS B.

Speaker authentication and identification : a comparison of spectrographic and auditory presentations of speech material

J.A.S.A., 44, 1596-1607 (1968)

VOIERS W.D.

Perceptual bases of speaker identity

J.A.S.A., 36, 1065-1073 (1964)

WAJSKOP M.

Identification de voyelles en fonction de leur durée

Proc. of the 6th Intern. Cong. of Phonetics Sciences - Prague (1967)

WOLF J.J.

Choice of speaker recognition parameters

Res. Lab. of Electronics M.I.T. - Q.P.R. 97, 125-133 (1970)

YOUNG M.A., CAMPBELL R.A.

Effects of context on talker identification

J.A.S.A., 42, 1250-1254 (1967).

o o  
o



Dernière page d'une thèse

VU

Grenoble, le 17 juin 1971

le Président de la thèse

J. BERT

VU, et permis d'imprimer,

Grenoble, le 18 Juin 1971

le Président de l'Université  
Scientifique et Médicale

Le Vice-Président,

G. CAU