UNIVERSITÉ PARIS VI. DÉPARTEMENT de MÉCANIQUE
LABORATOIRE DE MÉCANIQUE PHYSIQUE
CENTRE DE CALCUL ANALOGIQUE DU C.N.R.S.

RECHERCHES SUR LA PAROLE

Exposés
de

E. LEIPP, J.S. LIÉNARD, M. CASTELLENGO
J. SAPALY, D. TEIL, A. CALINET, M. MLOUKA

JANVIER 1971 N° = 53

GAM

BULLETIN DU GROUPE D'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES - TOUR 66 - PLACE JUSSIEU PARIS 5e

RECHERCHES SUR LA PAROLE

G.A.M. N° 53 Janvier 1971

LISTE DES EXPOSES

1° E. LEIPP - Les recherches sur la parole au Laboratoire d'Acoustique.

2° J.S. LIENARD - Application de l'ordinateur à la synthèse, la reconnaissance et l'étude statistique de la parole.

3° M. CASTELLENGO - Elaboration du dictionnaire des phonatomes et améliorations apportées à la voix de l'icophone.

4° J. SAPALY - Les Icophones à commandes optique et numérique.

5° D. TEIL - Développements numériques de l'Icophone.

6° A. CALINET - Description de deux programmes pour la correction et l'anamorphose des éléments phonétiques.

7° M. MLOUKA - Codage et analyse spectrale de la parole en vue de la reconnaissance automatique.

E. LEIPP

LES RECHERCHES SUR LA PAROLE AU LABORATOIRE D'ACOUSTIQUE

JANVIER 1971 N° 53

GAM

BULLETIN DU GROUPE d'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES . TOUR 66 . Place Jussieu . PARIS 5e

G. A. M.
Groupe d'Acoustique Musicale
Laboratoire d'Acoustique
Faculté des Sciences
Place Jussieu. Tour 66

Adresse Postale
9 Quai Saint Bernard PARIS 5°

Paris 10 Février 1971

BULLETIN N° 53

Réunion du vendredi 29 Janvier 1971

THEME : COLLOQUE SUR LA PAROLE

Exposé de M. LEIPP :

" L'Histoire de la parole au Laboratoire d'Acoustique "

M. le Doyen GAUTHIER n'a pu assister à notre réunion en raison de ses obligations.

Etaient présents :

M. le Professeur SIESTRUNCK, Président.
M. LEIPP, Secrétaire Général; Melle CASTELLENGO (Secrétaire).

Puis, par ordre d'arrivée :

M. MLOUKA (Informaticien au C.C.A.); M. CALINET (Informaticien CCA); M. LIENARD (Ingénieur Arts et Métiers); M. TEIL (ingénieur CCA); M. NAGAI (stagiaire au Laboratoire d'acoustique) M. GILOTAUX (Directeur, technique Pathé Marconi); M. VAL (J.L.) Maitre-assistant de Physi... M. F. FORET (compositeur); M. TOURTE (Professeur honoraire au Conservatoire); Mlle BOUE (orthophoniste); Melle DUTERTRE (Orthophoniste); M. HERVE; M. ROCHE (Institut de programmation); M. VIAL J.P. (Ingénieur SEMA); M. MERCIER (CNET, Lannion); M. BELLIER; M. DUPREY; M. DESGOUTTE (linguiste); M. DUPEYRAT (Stagiaire DEA informatique); M. TALAMON (Ed. MASSON) M. CARRE (Ingénieur, ENSERG, Grenoble); M. SAIEB (Musicologue); M. BRAURE (ingénieur informaticien CCA); M. BATISSIER (Secrétaire général SIERE); M. BELLEC (informaticien); Melle de CREVOISIER (orthophoniste); M. LANGEVIN (PG); Faculté des Sciences de Paris; M. LEOTHAUD; M. OUNA (S), traducteur; M. GUEROULT; M. CONDAMINES (Laboracoitre d'acoustique ORTF); Melle CHIKLER (orthophoniste); M. ETIENNE; M. REBOTIER (Conservatoire de Musique de Paris); M. SURUGUE (Etnomusicologue); M. GEUENS (IAD, Bruxelles); M. CEOEN (professeur IAD, Bruxelles); M. DEWEVRE (Rédacteur électro-acoustique); Melle JAMATI; Dr DORGEUILLE (Dr en médecine); M. OERLEMANS; M. OLIVA; Dr CLAVIE (Médecin); Mme STRAUS (Professeur de pédagogie au Lycée La Fontaine); Mme BOREL MAISONNY (orthophoniste); M. Akira TAMBA (Compositeur, musicologue); Melle MECZ (Assistante, Sorbonne); M. MOLES (Professeur, Faculté des Lettres de Strasbourg, Directeur de l'Institut de Psychologie sociale); Madame et Monsieur MARTINET (Professeur de linguistique à la Sorbonne); Dr POUBLAN (Médecin biologiste); M. J.J. DUPARC (Directeur de la Revue Musicale); M. Th. ANDRE TALAMON (Laboratoire de Mécanique Physique, Fac. des Sciences de Paris); Mme PRIM; Melle Danièle BRUERE; M. DEMARS; M. NAGLER (Stagiaire au Laboratoire d'acoustique, preneur de sons); M. CUVELIER; M. DREYFUS-GRAF (Ing. EPFZ); M. SERPOLET (Fac. Sc. BREST); M. BATIER; Mme CHARNASSE (CNRS); M.P. GERMAIN (Prof. Fac. Sc. Paris)

Etaient excusés : M. Charles MAILLOT (LYON); M. BLONDELET (Ets BUFFET-CRAMPON); M. CHARPEINE et JUNCK (AFIMA); M. JOLIVET; M. GROSSIN; M. CHAUVEAU; Mme MATTEODO; M. BOE; Melle E. WEBER; M. BLADIER; M. FAYEULLE; M. CASAL, M. CHENAUD; M. QUANCARD; M. GIRVES; M. ROGER; MM. BOUARD et GEIN; Melle COURTIN; M. PUJOLLE; M. FRANCOIS; M. TRAN VAN KHE; M.J. CHAILLEY.

PERIODIQUE : 6 numés annuels.
Prix de vente : service gratuit.
Imprimerie : Laboratoire de Mécanique Physique Fac. Sciences - PARIS.
Nom du Directeur : M. le Professeur SIESTRUNCK.
N° d'inscription à la commission paritaire : 42 283.

LEIPP GAM Janvier 1971

L'HISTOIRE DE LA PAROLE AU LABORATOIRE D'ACOUSTIQUE

par E. LEIPP

I. AVANT-PROPOS

Les idées, expériences et réalisations qui seront évoquées ici ont toutes été exposées lors de réunions ou publiées dans des comptes-rendus de congrès, bulletins GAM, Annales des Télécommunications, etc... , et on trouvera à la fin de ce texte une bibliographie exhaustive à ce sujet. L'intérêt de la présente réunion est de regrouper en un tout cohérent ce que nous avons fait au Laboratoire d'Acoustique depuis sa création dans le domaine de la parole : étude de la structure physique et perceptive des signaux de parole, réalisation de synthétiseurs pour démontrer la validité de la doctrine de synthèse que nous avons imaginée, idées ayant conduit à la mise en machine de notre méthode de synthèse, ainsi que celles qui sont développées actuellement quant à la reconnaissance automatique de la parole d'un locuteur quelconque. Ce tour d'horizon est pour nous l'occasion de faire le point, pour voir où nous en sommes. D'autre part ceux que nos recherches intéressent, seront informés ainsi à la source, toutes références utiles à l'appui.

Nous parlerons essentiellement des travaux faits au Laboratoire d'Acoustique et au Laboratoire de Mécanique Physique; la partie concernant les développements informatiques sera développée en détail plus loin, par Melle CASTELLENGO - M. LIENARD (J.S.) et les spécialistes du Centre de Calcul Analogique du C.N.R.S.

Ceux qui nous connaissent ne seront pas surpris de trouver ici un certain nombre d'opinions et d'idées passablement hétérodoxes, d'apparence parfois même saugrenue. Nous avons eu la chance que quelqu'un y ait cru suffisamment pour nous fournir les moyens de recherches indispensables : j'ai nommé M. Le Profresseur SIESTRUNCK, qui nous a constamment encouragés dès la première heure. Etant très ignorants de ce que font et disent les spécialistes de la parole : phonéticiens, linguistes, chercheurs et techniciens des télécommunications, etc... nous n'avons évidemment qu'un seul moyen, pour justifier nos idées sur la structure de la parole, c'était de montrer qu'elles sont utilisables pour faire de la synthèse. On peut vraiment prétendre avoir compris un mécanisme lorsqu'on a réussi à en faire à la fois l'analyse et la synthèse. C'est notre seul argument, mais il est de poids.

Il peut évidemment sembler étrange que des chercheurs spécialisés en acoustique musicale, s'intéressant au fonctionnemert et au rayonnement des instruments de musique et à la perception des sons musicaux, abordent les problèmes de l'analyse, de la synthèse et de la reconnaissance de la parole sur lesquels de très nombreux spécialistes mondiaux, disposant de moyens souvent illimités en personnel et en matériel, font des recherches systématiques depuis plusieurs décennies. Mais en y regardant de plus près, cela n'a vraiment rien d'étonnant. En effet, qu'est-ce donc que la "machine à parler" humaine sinon un instrument de musique comme les autres, avec anche, tuyaux, résonateurs etc... et fonctionnant strictement selon les mêmes lois qu'un instrument à vent classique Qu'est-ce donc que le rayonnement acoustique de l'appareil phonatoire et que représente-t-il acoustiquement de particulier comparativement à celui d'une trompette ou d'une "voix humaine" d'orgue ? Qu'est-ce donc que le mécanisme de la perception et de l'"intégration" de la parole; n'est-il pas le même en perception et en intégration de la musique Si les points de vue que nous avons sont originaux par rapport à ceux des spécialistes de la parole, cela ne peut être que favorable. Dans combien de domaines des "incompétents" n'ont-ils pas apporté de la lumière ! KEMPELEN et FABER étaient-ils phonéticiens ou linguistes ? Charles CROS était-il acousticien ? On le sait bien : il suffit parfois d'un hasard pour que des non-spécialistes s'engagent dans une voie, connexe à la leur en général, et, du fait même de leurs points de vue particuliers trouvent des façons

...../

prégnantes d'aborder un problème. Ce hasard, pour nous porte un nom : il s'agit d'un instrument de musique très particulier, d'un "monstre" de la lutherie; il s'agit de la guimbarde en un mot, sur laquelle j'avais fait des recherches systématiques dès 1959; on y reviendra plus loin.

Il n'est certes pas hors de propos ici de dire quelques mots sur la question des antériorités. Si l'on veut bien se donner la peine de chercher un peu, en reculant dans le temps, on est à peu près certain, en tous domaines, de trouver toujours un poète, un penseur, un chercheur imaginatif ayant eu telle ou telle idée de "génie"... réalisé telle ou telle machine.... Souvent ils n'ont pu passer à la réalisation pour des raisons d'ordre technologique ou financier, mais l'idée était bel et bien "en l'air"... On pourrait citer des cas innombrables, entre la "parole gelée" dont parle RABELAIS, le voyage dans la lune avec Jules Verne, qui décrivit aussi bien la télévision dans le Chateau des Carpathes.... La recherche des antériorités est donc un jeu stérile : le chercheur honnête cite toujours ses sources; il ne peut le faire lorsqu'il n'en a pas, et qu'il a retrouvé tout seul ce que d'autres avaient découvert avant lui. En fait, nous avons mis sur pied une doctrine cohérente de la phonation et de la structure physique de la parole; nous avons imaginé une méthode de synthèse à l'aide d'un appareillage que nous avons conçu et réalisé sans même connaitre les théories de DELATTRE ou le Playback de HASKINS ou autres. Ce que nous avons trouvé représente en fait toujours des choses très simples que tout chercheur digne de ce nom est capable de réinventer sans aller copier ce qu'on fait ses collègues ....

Ces réflexions ne sont certes pas inutiles dans le domaine qui nous concerne ici... Mais passons donc aux faits .... que nous prendrons à peu près dans l'ordre historique, ce qui montrera l'enchainement des faits et des circonstances.

II. L'HISTOIRE DE LA GUIMBARDE

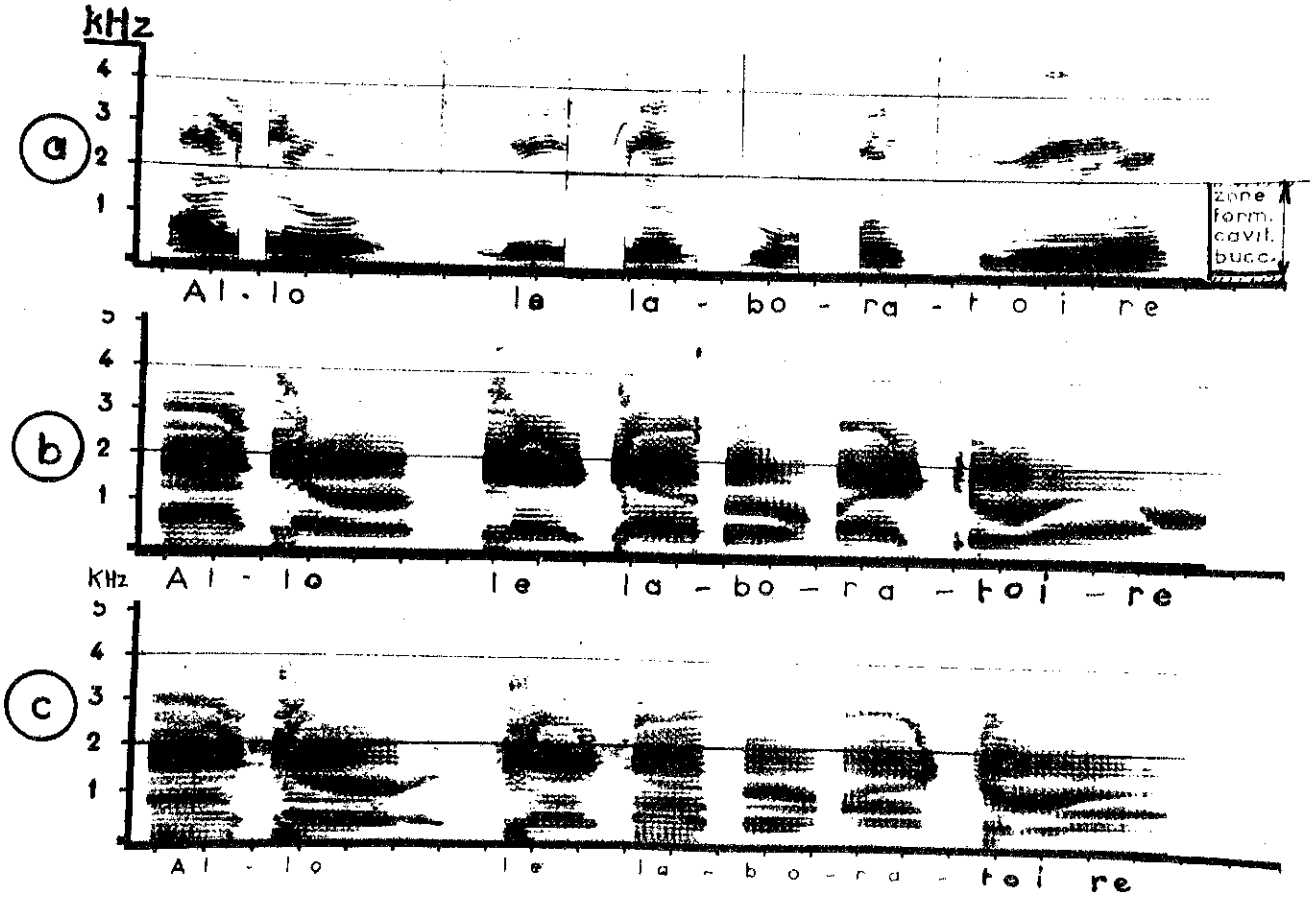
C'est à partir de 1960 que les choses prirent une tournure sérieuse. Je faisais depuis longtemps des recherches sur les instruments de musique, étudiant, avec les moyens de l'électro-acoustique commençait à mettre à notre disposition, le fonctionnement et le rayonnement des instruments de musique. J'ai eu l'occasion de donner quelques précisions sur mes activités antérieures à la création du Laboratoire d'Acoustique lors de la réunion du GAM sur le GAM (Bulletin N° 50, octobre 1970). Je rappelle à cette occasion le rôle déterminant que jouèrent dans ces recherches M. Charles MAILLOT de Lyon et M. A. MOLES, qui me fournirent alors des idées et des moyens de recherche. Mais tout commença après la soutenance de ma thèse sur les instruments à cordes frottées en 1960 à la suite de laquelle M. le Professeur SIESTRUNCK, directeur de ma thèse, me confia la responsabilité du laboratoire d'acoustique. Je m'intéressais alors précisément à divers "monstres" d'apparence simpliste ex : ocarina, orgue à bouche japonais, harmonica et... guimbarde.

Au Congrès international d'acoustique de LAUSANNE en 1957 (bib.1) et au Congrès International d'Acoustique de STUTTGART (bib.2), j'avais déjà présenté, en collaboration avec M. MOLES, une méthode d'appréciation de la qualité des instruments de musique à l'aide du sonagraphe. Cet appareil qui venait d'apparaître en France, apportait alors une réponse à de nombreuses questions, car c'était la seule méthode efficace pour analyser des sons à caractère fortement évolutifs. Rappelons que le sonagraphe était primitivement destiné à l'étude de la structure physique de la parole ...

Le sonagraphe et la représentation sonagraphique resteront désormais notre méthode de choix, tant en musique qu'en parole. Grâce à lui il me fut possible de montrer le rôle de la cavité buccale dans le timbre des instruments de musique à vents où une anche vibre dans la cavité buccale. Celle-ci détermine en fait un "formant", un "harmonique sensible" des spectres rayonnés par de tels instruments. Le musicien peut régler à volonté celui-ci pour moduler le timbre des sons qu'il produit, et les rendre

...../

①



La figure donnée dans les Annales Télec. T.I8 nº5-6 (I963) bib 6

PAROLE SEMI-SYNTHETIQUE REALISEE AVEC LA GUIMBARDE

a) . On a tiré un sonagramme de la phrase : "Allo!Le laboratoire" où l'on a indiqué la zone formantique de la cavité buccale (200-2000 Hz).Ce sonagramme montre bien d'une part les "barres de résonance",les formants...et on y lit aisément les raies harmoniques délivrées par les cordes vocales.Les "formants" ,dans leur évolution,"dessinent" bien des graphismes ,qui supportent l'information sémantique de la parole.

b) . On a remplacé ici les cordes vocales par la lame vibrante d'une guimbarde,qui délivre,elle aussi,un spectre de raies harmoniques,parallèles ,puisque la fréquence de la lame ne change pas.On a articulé la même phrase :"allo! Le laboratoire" -sans débit d'air,et cordes vocales immobilisées.On entend alors une voix synthétique "sonnant " le vocoder....Les formes dessinées par les formants sont très similaires à celles de la parole normale (a)....

c) On a utilisé ici une autre guimbarde:le résultat est le même.

auditivement intéressants. Les résultats, très résumés, ont fait l'objet d'une communication au Congrès international de Copenhague, en 1962 (bib.3), et si nous citons ce travail, c'est bien entendu parce que je vais retrouver le rôle de la cavité buccale lors de l'étude de la guimbarde et de la parole semi-synthétique qu'elle permet de réaliser.

La guimbarde, simple lame d'acier à fréquence quasi-fixe, et que l'on fait vibrer entre les dents afin d'exciter la cavité buccale, avait fortement excité ma curiosité à l'époque,et en 1962 j'avais envoyé une bande documentaire à un concours de "chasseurs" de son, qui fut alors primée et diffusée lors d'une émission radiophonique (14/9/1962). Je venais justement de trouver le moyen de réaliser avec la guimbarde une parole semi-synthétique, sonnant très curieusement à la manière d'un vocoder. La bande présentée débutait précisément par des échantillons de phrases synthétisées ainsi, en français, allemande et anglais. Cette bande existe toujours au laboratoire d'acoustique... Ce fut la première expérience de synthèse de parole ! Et ceci me conduisit à étudier de plus près le fonctionnement et le rayonnement de la guimbarde. Plusieurs publications, en 1963, rendent compte des résultats de mes recherches (Revue du Son, Acustica, Annales des Télécommunications; bib. 4-5-6). Cette dernière étude a pour titre " Un vocoder mécanique, la guimbarde" et résume une communication faite au Groupement des acousticiens de langue française en 1962. Un certain nombre d'idées étaient dès lors apparues, par exemple celle de la notion de l'analogie entre signes acoustiques de la parole et signes graphiques de la sténographie, considérés comme des supersignes résultant d'association de signes élémentaires. La publication comporte un document faisant le parallèle entre le sonagramme de la parole normale et celui de la parole synthétique réalisée avec la guimbarde (phrase : "allo, le laboratoire"; fig. 1).

Dans ce texte on retrouvera l'idée que "l'essentiel de l'information sémantique est contenue simultanément dans les deux sonagrammes". Celle, aussi, qu'il existe, en sténographie Duployé, "28 signes correspondant au phonèmes et que l'on peut associer selon les besoins". Il est explicitement précisé que les "barres de résonance" (formants selon FLETCHER) véhiculent en fait l'information sémantique", et que "grâce à ce modeste instrument de musique, des personnes privées de cordes vocales peuvent retrouver la parole moyennant un apprentissage sommaire". etc...

Ces extraits montrent qu'un certain nombre de bases étaient dès lors posées, débouchant sur une recherche systématique en parole. Le "vocoder à deux francs cinquante" que constitue une guimbarde ne pouvait bien entendu séduire personne, quoique nous l'ayons proposé à des médecins spécialisés en rééducation de parole, laryngectomies etc...

Ceci est l'histoire de la guimbarde, et les choses en étaient là lorsque se produisit un autre évènement, déterminant pour la suite de nos recherches. Il s'agit de la thèse de GEISSERT, soutenue à l'Institut National des Sourds Muets, Rue St-Jacques, le 21 Octobre 1964.

II. L'HISTOIRE DE LA THESE DE GEISSERT

Sous la direction de M. MOLES, un étudiant en rééducation des Sourds-muets avait fait le travail suivant. Un texte était enregistré au magnétophone (La petite chèvre de M. SEGUIN, de A. DAUDET). La bande était ensuite coupée en fragments de 2,4 secondes, dont on tira les sonagrammes. Ceux-ci, collés bout à bout, représentaient en diagramme continu fréquence-temps le texte intégral. Deux rouleaux permettaient de faire défiler cette "bande sonagraphique" qui fut filmée intégralement. L'expérience consistait à faire écouter le texte en projetant simultanément la séquence cinématographique des sonagrammes. Pour nous qui étions bien habitués déjà à la représentation sonagraphique, une question se posait : "est-il possible de suivre le discours à partir des images ?". Nous étions présents tous les trois : Melle CASTELLENGO, M. J.S.LIENARD et moi-même. La réponse fut nette : il était impossible de "lire" les sonagrammes à cette cadence.

...../

Les raccordements avec les recherches faites sur la guimbarde furent immédiats : l'information sémantique était supportée uniquement par les "barres de résonance", les formants, et tout le reste spectres de raies, bruits etc..., ne véhiculait que de l'information esthétique, sans intérêt dans le cas considéré... Bref, la redondance énorme de la parole était à la base du fait qu'on ne savait absolument pas quoi "lire" sur les sonagrammes qui défilaient. Il y avait beaucoup trop de choses inutiles dans le signal de la parole normale si on s'intéresse exclusivement à l'information sémantique. La forêt cachait l'arbre ici....

Cette réunion fut très stimulante, au point que le même soir, il fut tenu entre nous trois un "colloque" où certaines idées furent élaborées, qui s'avèreront fructueuses par la suite, en particulier celle de l'étude de la parole chuchotée.

Ceci est l'histoire de la thèse de GEISSERT. Voyons à présent les suites qu'elle eut pour nous.

III. L'HISTOIRE DE LA PAROLE CHUCHOTEE

Lors de ce colloque, au soir du 21 octobre 1964, les échanges d'idées aboutirent à la conclusion suivante. La parole normale véhicule trop d'information; comment faire émerger les "formes sémantiques" simples que nécessairement notre système auditif sait extraire à l'audition normale. L'accord se réalisa rapidement sur un point : ce sont les raies du spectre qui représentent la musique de la parole. Or cette musique n'a aucune signification sémantique; elles ne conditionnent que le caractère esthétique de la parole. Comment supprimer cette musique ? Nous savions qu'elle était engendrée par les cordes vocales, instrument à anche musculaire double responsable des spectres de raies de la parole. Pour supprimer ces raies rien de plus simple : on parlera sans faire fonctionner les cordes vocales. Un moyen : la guimbarde, qui met très bien en évidence les formes sémantiques. Celles-ci sont faciles à étudier, à "lire" sur le sonagramme, car les raies engendrées par la guimbarde restent fixes, ce qui simplifie tout du point de vue dépouillement des sonagrammes. Si ce spectre reste fixe, il ne véhicule, perceptivement, pas d'information. En effet, c'est ce qui change dans le temps, la fréquence en particulier, qui est susceptible de supporter des messages rapides... Et si on simplifiait encore le problème en supprimant complètement le spectre de raies sans aucun intérêt pour nous, puisque seul nous intéresse provisoirement le message sémantique ? Rien de plus facile ! Après tout il suffit de parler en voix chuchotée... L'expérience montre que si elle manque d'agrément et de portée (avantage pour les élèves...) elle est totalement intelligible. Donc elle véhicule bien toute l'information sémantique. C'est une bonne raison pour étudier la parole chuchotée systématiquement. Le même soir, des analyses de phrases chuchotées nous avaient montré la légitimité de nos hypothèses et notre agenda porte à cette date la mention :

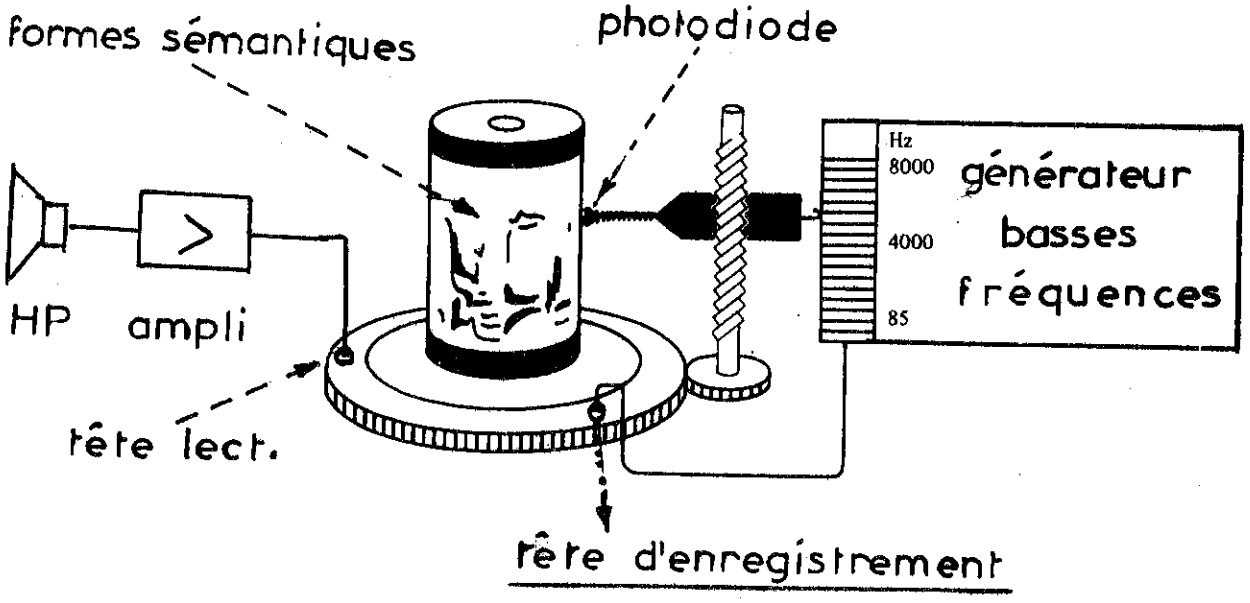
" La voix chuchotée entre 500 et 2000 Hz contient toute l'information sémantique; d'où étude des formes vocales selon le schéma sténographie. Expérience faite par filtrage de la bande entre 500 et 2000 Hz en voix chuchotée ".

La bande annoncée étaient un peu optimiste; mais si l'on veut bien tenir compte du fait que les filtres ont nécessairement une pente d'atténuation plus ou moins marquée, on peut dire qu'en fait toute l'information sémantique est contenue entre quelque 100 et 4000 Hz. Ce fait est confirmé, d'ailleurs, par l'exploitation systématique, empirique, de la "bande téléphonique", dès les premiers modèles de téléphone et de phonographes. Cette bande correspond à celle du maximum de sensibilité de l'oreille, à laquelle la parole humaine et les instruments de musique traditionnels se sont adaptés au mieux, par expériences, erreurs, corrections successives.

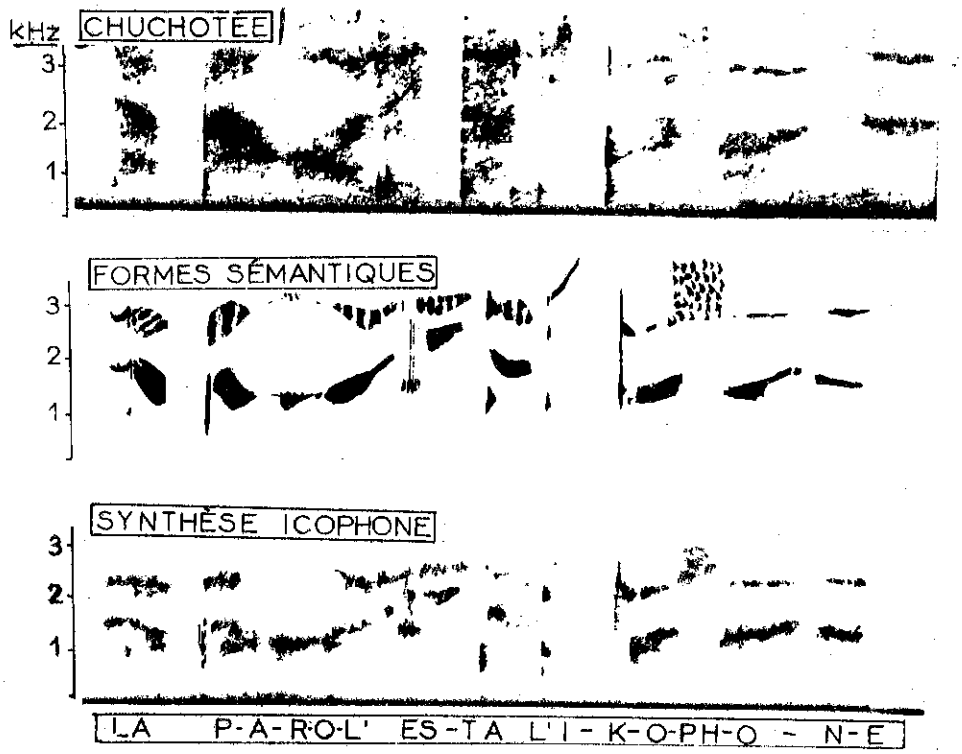
La décision fut prise alors de faire une étude systématique de la parole chuchotée, dont on couperait ce qui est inférieur à 300 Hz et supérieur à 4500 Hz.

...../

②



L'ICOPHONE I. Une photodiode explore graduellement ,du bas vers le haut,un "icogramme" dessiné à l'encre de chine sur une feuille disposée sur le cylindre tournant.On enregistre par rerecording les signaux délivrés par le générateur de sinusoïdes.A la fin de l'opération,on relit la boucle magnétique;on entend la phrase dessinée sur l'icogramme.



Voici le document présenté à LANNION (Colloque GALF,Juin 1965) En haut,le sonagramme en voix chuchotée de la phrase:"la parole est à l'Icophone";au milieu les formes sémantiques extraites de cette analyse;en bas,la synthèse réalisée à l'Icophone I.

En procédant ainsi, on supprimait à la fois l'information esthétique bien gênante en raison de son importance énergétique, ainsi que la redondance fréquentielles dans les bandes marginales de la zone sensible de l'oreille. En filtrant de cette manière la voix chuchotée on supprime en fait plus de 90 % de l'énergie d'un mot et dès lors celui-ci devient alors un phénomène acoustique relativement simple, qui se traduit sur les sonagrammes par un graphisme, une forme, à la fois peu compliquée et très floue, mais en tout cas reconnaissable sans difficulté et sans ambiguité parmi les formes des autres mots. En peu de temps cette idée se précisa, et nous commençions à repasser à l'encre de Chine sur les sonagrammes, les "formes" de mots et de phrases, ces schématisations supprimant encore une grande partie de l'information inutile contenue dans la parole chuchotée-filtrée que nous utilisions.

Très rapidement nous fûmes conscients que, pour démontrer la réalité de nos hypothèses, il fallait réaliser une machine à relire acoustiquement nos graphismes et vérifier s'ils correspondaient à de la parole intelligible.

Il était important, d'abord, de trouver un nom, car une chose qui n'a pas de nom n'existe pas .... Antikay ? puisque c'est le contraire du sonagraphe fabriqué par Kay ! ... Pharganos, puisque c'est le mot "sonagraphe" prononcé à l'envers .... N Finalement ce fut "ICOPHONE" qui fut adopté, c'est-à-dire, avec une petite entorse aux habitudes étymologiques : " l'appareil qui transforme des images en sons ".

Les choses ne traînèrent pas. M. SIESTRUNCK nous envoya son spécialiste électronicien, M. SAPALY et un conseiller précieux : M. MOUTET. Peu de crédits ... il faut les compenser par de l'imagination. En Juillet 1964, l'appareil était en cours de construction; c'était l'ICOPHONE I (fig. 2). On utilisait le moteur, le cylindre, le système d'entrainement et la tête d'enregistrement du sonagraphe normal, auquel furent adaptés divers organes très simples : un micro-projecteur de spot, éclairant fortement l'image des formes sémantiques fixée autour du cylindre du sonagraphe, à la place normale du papier à étincelage; une cellule photo-électrique, qui, lors de la rotation du cylindre captait les taches noires des dessins au passage et déclenchait au même moment un générateur de sinusoïdes dont les sons montaient pendant que le projecteur et la cellule se déplaçaient de bas en haut. Les "morceaux" des formes acoustiques relues ainsi, délivrés par le générateur, étaient ensuite "empilées" par re-recording sur le disque magnétique du sonagraphe. Quand l'opération était terminée, on pouvait écouter l'enregistrement global.

Etant donné la rusticité des moyens employés, on ne pouvait guère escompter de synthèse de bonne qualité. Les essais préalables, faits en utilisant au lieu de parole des dessins à l'encre de Chine, traits, cercles, triangles etc..., montrèrent que la "fidélité" avec laquelle l'ICOPHONE I transformait des images optiques en images acoustiques était très douteuse .... Cependant, la suggestion aidant, nous avons réussi à synthétiser quelques phrases, intelligibles en y mettant de la bonne volonté, et qui l'étaient en tout cas suffisamment pour montrer que notre idée directrice était bonne. Le problème de l'intelligibilité de la parole synthétique, en tout cas, était posé dès lors. Nous comprenions très clairement chaque mot synthétisé; pour nous c'était parfaitement intelligible, parce que nous savions à l'écoute, ce qui allait arriver et que très peu d'information acoustique était nécessaire pour reconnaitre alors les "formes" en question. Mais M. SIESTRUNCK qui arriva au moment de notre premier "eureka" ne comprit pas la moindre trace de phonème ou de mot ....

Ce fut stimulant pour nous de chercher à perfectionner notre technique-acrobatique - pour améliorer les résultats. Dès cette époque, en tout cas, un certain nombre d'idées se firent jour. En particulier, si la parole est composée de "formes", il devrait être intéressant de l'étudier sous l'aspect de la GESTALTTHEORIE sur laquelle A. MOLES, quelques années auparavant, avait attiré mon attention.

Il s'avéra rapidement que cette voie était intéressante, et la lecture de divers ouvrages traitant de ces questions nous apporta les compléments d'information nécessaires. En fait, les lois fondamentales de la Gestaltthéorie, de la psychologie

...../

des formes perçues, s'appliquent intégralement ici :

- une forme est un tout, une intégralité, qui n'est pas égale perceptivement à la somme de ses parties; c'est quelque chose de plus, et dont l'originalité vient de la manière de raccorder les parties et de choisir les points de jonction. Ainsi la figure 3 donne-t-elle 5 "sommes de parties identiques" ... mais une seule façon de raccorder les éléments reconstitue la forme voulue : celle d'un lapin !

- de même, une forme est transposable sans cesser d'être reconnue (fig.4). Von EHRENFELS raisonnait sur des formes mélodiques que l'on peut transposer vers l'aigu ou le grave qui ne perdent rien de leur caractère... Ici, il en est de même ! Formes mélodiques ou graphiques : c'est la même chose, strictement. Toute l'information que nous envoie le monde extérieur nous arrive en fait toujours sous l'aspect de "formes", de "gestalts" à trois dimensions : intensité (ou grosseur), composition de l'objet (spectre acoustique, optique ou chimique) et durée. Seuls les capteurs périphériques du cerveau changent, le reste est codage en impulsions et traitement de l'information par ordinateur .....

- une forme, une gestalt, est anamorphosable très largement sans cesser d'être reconnue. Même quand on dépasse les limites usuelles de reconnaissance de ces formes, on peut encore deviner qu'il s'agit bien ici d'un lapin et non d'une maison ou d'un sapin ....

- le niveau des signaux de parole ne joue aucun rôle, dès que la "forme" acoustique est complète (fig.5). Quelle que soit la grosseur des traits, et même leur degré de flou, la gestalt est toujours reconnue ...

Ces réflexions furent très fructueuses, mais une question se posa dès lors.

Synthétiser de la parole en recopiant des sonagrammes, ce n'est guère mieux que de relire une bande magnétique.... ce n'est pas vraiment faire de la synthèse, c'est-à-dire reconstituer une forme, un corps, à partir de ses éléments. Puisqu'une gestalt n'est pas égale à la somme de ses éléments comment allons-nous pouvoir trouver une réponse à ce problème ? Comment pourrons nous trouver le secret des raccordements entre éléments ?

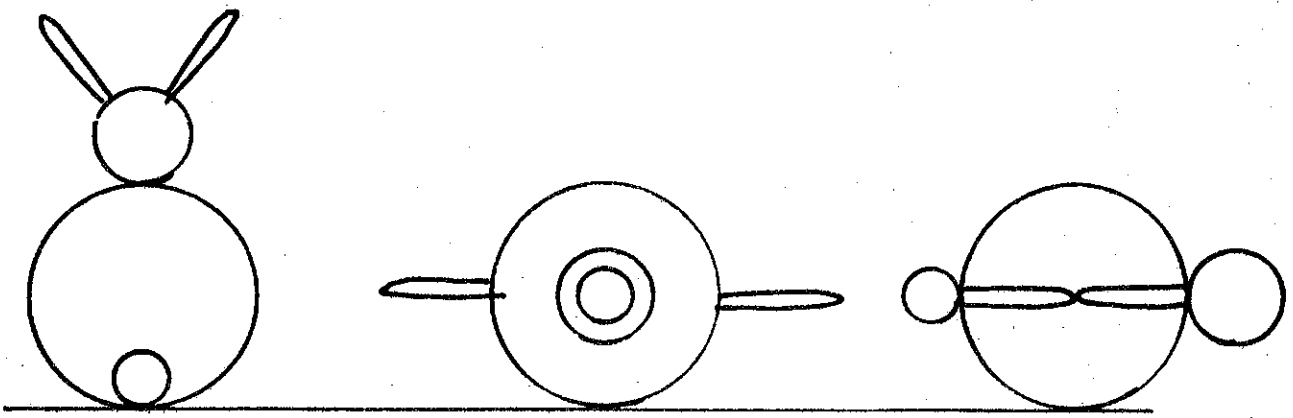
La réponse apparut bientôt, à la suite d'observations sur "l'instrument de musique" à faire de la parole. Une observation banale montre en effet que lorsque nous parlons, l'appareil phonatoire n'est jamais au repos. Il n'y a donc strictement aucun phénomène périodique en parole normale, et le découpage habituel des phonéticiens en "phonèmes" n'est, dès lors, plus légitime. On peut bien prononcer en continu des "voyelles" acoustiquement et en faire un signal acoustique quasi périodique et stable; mais il s'agit là d'un artefact qui n'a rien à voir avec la parole normale. La parole est d'abord envoi d'informations par modulation de la fréquence dans le temps; si rien ne change dans le temps, il n'y a plus d'information, donc plus de parole.... Les voyelles de la phonétique ne sont donc pas de la parole.... De toutes façons les consonnes, signaux transitoires très rapides par excellence, et qui véhiculent justement beaucoup d'information pour cela, ne peuvent être étudiées comme des "unités" de parole. Leur forme, ainsi d'ailleurs que celle des voyelles, est effectivement différente selon le "voisinage" immédiat. Si on considère des mots différents, mais comportant les mêmes "phonèmes", on vérifie sans difficulté que ces phonèmes n'ont absolument pas la même forme, parce que les points de raccordement, pour un même phonème, sont différents. Que l'on fasse donc les sonagrammes chuchotés de : pari, irap, pira, païr, piar etc... pour s'en convaincre ....

Mais comment alors passer à la "synthèse vraie" ! Si les phonèmes ne sont pas les "éléments" de la parole, quels sont donc ceux-ci ?

La solution fut trouvée en raisonnant de la façon suivante. L'appareil phonatoire est composé d'organes mobiles, dont chacun possède un certain "champ de liberté". La langue peut se placer à un certain nombre de positions limites, et pour parler, elle va

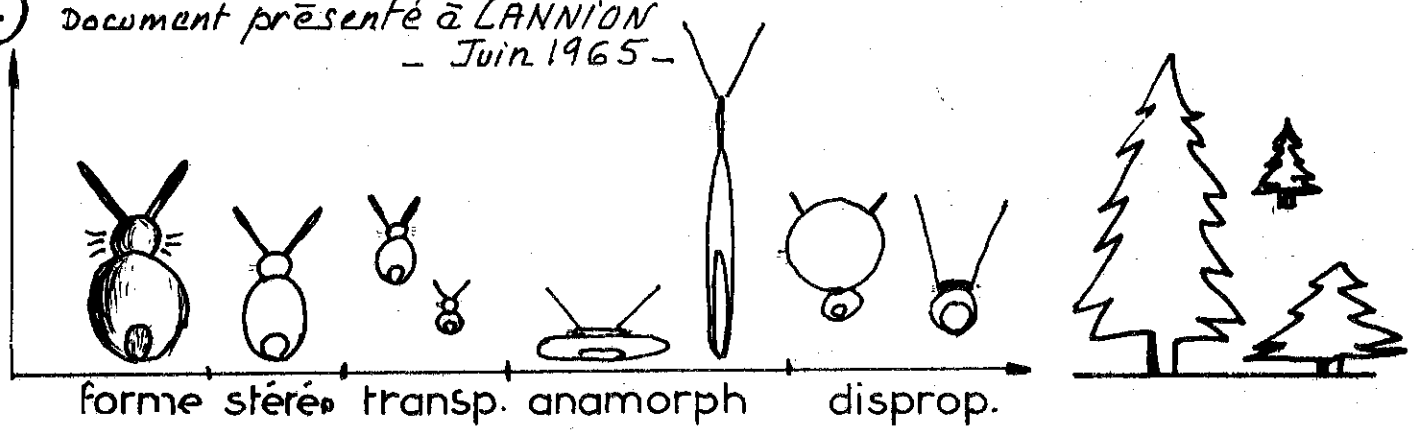
...../

③



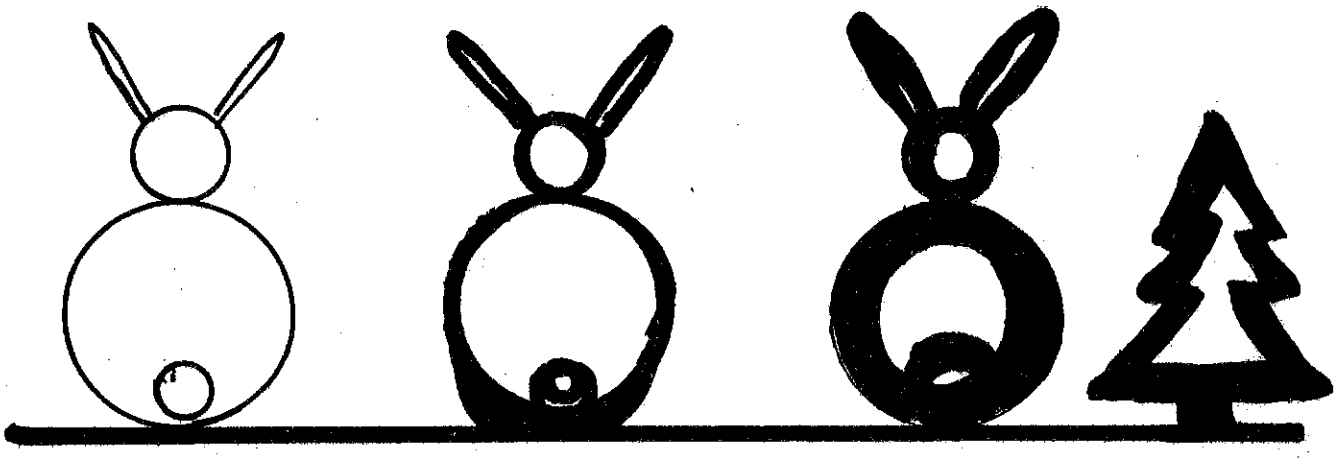
Une forme n'est pas égale à la somme de ses parties.... C'est une totalité dont la reconnaissance dépend de la manière dont les parties sont raccordées entre elles et du choix des points de jonction.

④ Document présenté à LANNION - Juin 1965 -



Une forme est transposable et anamorphosable très largement avant qu'elle ne cesse d'être reconnue. Même disproportionnée, on ne confondra jamais la forme d'un lapin avec celle d'un sapin.

⑤



Le niveau ne joue pratiquement aucun rôle dans la reconnaissance des formes: que les tratis soient plus gros, plus flous, avec des rapports de niveaux (des grosseurs des traits différents), on ne confondra jamais un lapin avec un sapin.....

passer de l'une à l'autre de ces positions. Chaque position est définie selon les cas par les dispositions des joues, des mâchoires, du voile du palais etc..., organes ayant eux-mêmes leur champ de liberté propre. Finalement un mot articulé normalement est une combinatoire de mouvements dont chacun est limité par les autres. Un mot est fondamentalement un mouvement (conception dynamique de la parole), et un mouvement ne peut être décrit que par des éléments de mouvement. Ces éléments sont définis par les points limites entre lesquels se déplacent les organes. Un mot comme PARIS, du point de vue acoustique, n'est donc pas la somme des phonèmes P, A, R, I ; on le vérifie facilement en collant bout à bout lesdits "phonèmes" enregistrés sur bande magnétique; on n'obtient jamais ainsi le mot "Paris" ....

Par contre un mouvement élémentaire de l'appareil phonatoire est représenté par le passage d'un phonème (c'est-à-dire d'un ordre de positions) à un autre.... d'où l'on repart pour le suivant etc.., Donc PARIS peut acoustiquement se décomposer en "diphonèmes" qui sont les éléments insécables de la parole. Ainsi le mot "Paris" se décompose en 3 diphonèmes : soit PA, AR, RI. Chacun de ces trois diphonèmes, que nous appelons PHONATOMES parce qu'ils représentent l'atome parlé, se raccorde nécessairement avec le précédent et le suivant. La forme globale du mot constitue, puisque les points de jonction sont alors déterminés, une superforme composée de 3 phonatomes. Le nombre des phonatomes possibles en français est bien facile à préciser; si on admet 30 phonèmes par exemple, les phonatomes, associations deux à deux des phonèmes, seront au nombre de 30 x 30 c'est-à-dire 900.

Etant donné la structure et la musculature de l'appareil phonatoire, certains éléments de mouvement sont plus facile à faire. Il n'est donc pas étonnant que l'homme ait utilisé d'abord ceux-ci, en vertu de la loi du moindre effort. Par contre d'autres mouvements sont difficiles ou impossibles, sauf à imaginer un long entrainement. Bref, il n'est peut-être pas utile d'étudier les 900 phonatomes dans leur intégralité, si certains d'entre eux ne sont jamais utilisés ! Aussi avons-nous pensé dès le début faire une "enquête" pour avoir une idée du taux d'occurence des diphonèmes. Avec Melle CASTELLENGO, j'ai donc analysé des textes littéraires, en français et en allemand, "pointant" les diphonèmes sur une matrice carrée de 30 x 30, dans laquelle on pouvait localiser tous les diphonèmes possibles. Les premiers résultats montrèrent qu'effectivement certains phonatomes étaient très fréquents et d'autres rares, d'autres inexistants.... Une petite statistique "manuelle" fut faite. Mais ce genre de travail est long et fastidieux Il se trouvait que J.S. LIENARD était à l'époque en rapport avec M. TEIL, au C.N.A.M., qui disposait alors d'une petite calculatrice (CAB 500). Faire un travail sur les taux d'occurence des phonatomes, est une idée qui nous est venue rapidement ! .... et ce travail fut débuté dès lors, parallèlement à la suite de nos recherches.

L'intérêt de l'opération est évident. Si effectivement la parole est faite de 900 phonatomes associés de diverses manières, il suffit de tirer 900 sonagrammes représentant l'ensemble de ces phonatomes. Ces phonatomes, enregistrés surbande, en parole chuchotée-filtrée, fournissent alors 900 graphismes élémentaires; 900 digrammes phonétiques, qui se raccordent nécessairement tous entre eux si l'on admet ce découpage de la parole. Ces digrammes constituent donc proprement les signes élémentaires d'une "sténographie" acoustique, signes que l'on peut assembler pour en "dessiner" des superformes, des mots ou des phrases, de façon fluide, simulant au mieux la parole chuchotée normale.

Dès cette époque démarrèrent les travaux d'établissement du "dictionnaire" des éléments phonétiques, ou, mieux, des "digrammes phonétiques" puisqu'il s'agit ici de graphismes.

A cette époque fut organisé un colloque sur la parole, par le Centre National des Etudes de Télécommunication (CNET) à LANNION, en Juin 1965. Ce fut la "première" de l'ICOPHONE, et nous y avons publié un opuscule ronéotypé, distribué à tous les participants, puis à d'autres personnes. Cette publication fut une manière de manifeste, résumant nos idées sur la parole considérée comme Gestalt, donnant nos premiers résultats en taux d'occurence des phonatomes, et apportant à la fois des sonagrammes de nos premières phrases synthétisées ainsi que les échantillons sonores correspondants. Voici le document de base présenté à LANNION (fig.6) en même temps d'ailleurs que la figure 4.

...../

Notre méthode de synthèse s'avérait à la fois valable et économique. Le point noir était l'appareillage, l'ICOPHONE I, vraiment inutilisable pour des recherches systématiques. M. SIESTRUNCK décida dès lors de nous faire construire une machine plus fonctionnelle : ce fut l'ICOPHONE II.

III. L'HISTOIRE DE L'ICOPHONE II

Pour faire la synthèse d'une phrase de 2,4 secondes avec l'ICOPHONE I, il fallait évidemment attendre que l'opération de relecture de l'ICOGRAMME par bandes successives de 45 Hz soit terminée, ce qui demandait un temps notable (5 minutes). L'idée maîtresse ayant présidé à l'élaboration du cahier de charges de l'ICOPHONE II (figure 7) était l'instantanéité de la relecture. Ceci impliquait l'utilisation non pas d'une cellule unique, mais d'un grand nombre de cellules travaillant simultanément. Arbitrairement il fut décidé d'utiliser une barrette comportant 44 photo-diodes alignées dont chacune piloterait un générateur de sinusoïdes autonome. Les 44 générateurs étaient ensuite mélangés, et le signal, convenablement amplifié, écouté directement et instantanément sur haut-parleur. On pouvait ainsi "lire" une image graphique quelconque, y compris des ICOGRAMMES de parole chuchotée, et les transformer en "images" sonores. Le support est une feuille de mylar transparente, sur laquelle on dessine les formes voulues à l'aide d'un pinceau et d'une encre noire spéciale. Chaque générateur est réglable de façon autonome en fréquence, autour de la valeur affichée, et en intensité. On peut par exemple accorder avec précision (compteur-électronique) les générateurs de 100 en 100 Hz, réalisant ainsi un son harmonique de fondamental 100 Hz. En fait, on "quantifie" le dessin contenu de l'icogramme par tranches de 100 Hz....

Le but était alors de simuler la parole chuchotée. On eut donc l'idée de réaliser plutôt un "bruit" qu'un spectre de raies harmoniques ! Ce but est atteint en désaccordant systématiquement les cellules pour les étager "en gros" de 100 en 100 Hz entre 100 et 4400 Hz environ. De plus, un système particulier permet de faire fluctuer aléatoirement chaque fréquence choisie, dans des marges réglables, autour de la valeur affichée. Ainsi la parole chuchotée synthétisée prenait des allures plus naturelles, car la voix chuchotée naturelle n'est pas stable non plus !

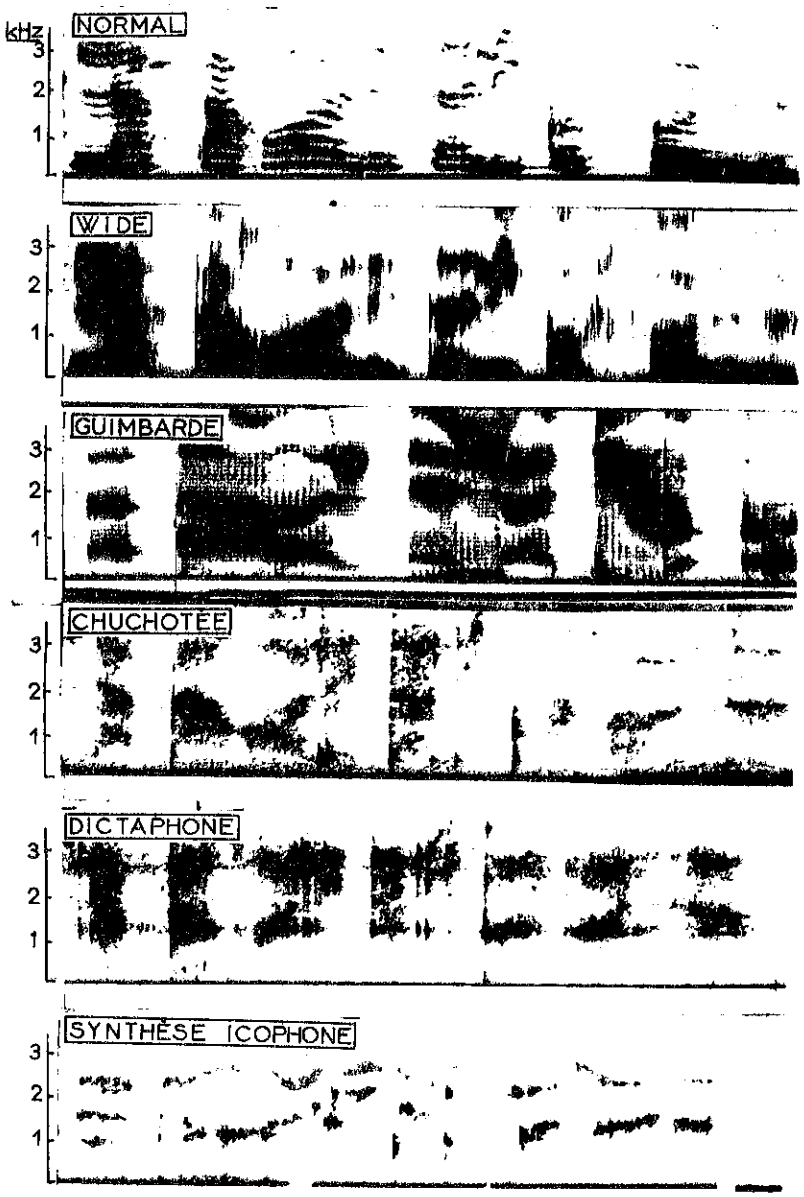
L'expérience montre rapidement que la formule était bonne. La bande était entraînée soit manuellement, soit avec un moteur particulier permettant de faire varier la vitesse de défilement de façon très régulière entre 0 et 50 cm/seconde environ. Ce procédé permet évidemment de réaliser un clivage entre hauteur et vitesse de défilement du mylar. Contrairement à ce qui se passe dans le magnétophone où le changement de vitesse produit un changement de hauteur, on peut ici défiler plus ou moins vite sans modifier celle-ci.

Cet appareil terminé en 1967, fonctionne toujours de façon parfaite après plus de 4 ans d'usage..Ce sera un outil puissant de recherche, produisant une parole synthétique dont la qualité est sans commune mesure avec celle de l'ICOPHONE I. Les premiers résultats furent présentés au Colloque sur la parole organisé par le GALF à Grenoble, en Juin 1967. Des phrases synthétisées à partir de copies de sonagrammes (le dictionnaire des éléments phonétiques n'étant pas assez avancé) furent présentées lors de cette réunion; ce furent d'ailleurs les seuls exemples de synthèse de parole présentés alors, et dont les auditeurs purent apprécier la qualité d'intelligibilité. A cette occasion, le laboratoire fit 4 communications : (bib.7) l'une sur les idées générales relatives à la structure sémantique de la parole (LEIPP), l'autre sur la description de l'appareillage ICOPHONE II (SAPALY); puis sur les problèmes posés par la synthèse avec cet appareil (M. CASTELLENGO); enfin ce fut la communication de J.S. LIENARD sur les problèmes du "Dictionnaire" des éléments phonétiques.

Après le colloque de Grenoble, Melle CASTELLENGO s'attaqua à la "rédaction" de textes en ICOGRAMMES ayant une certaine envergure (La malle sanglante de la gare de Lyon, réalisé par copie de sonagrammes; la Cigale et la fourmi par LEIPP etc...).

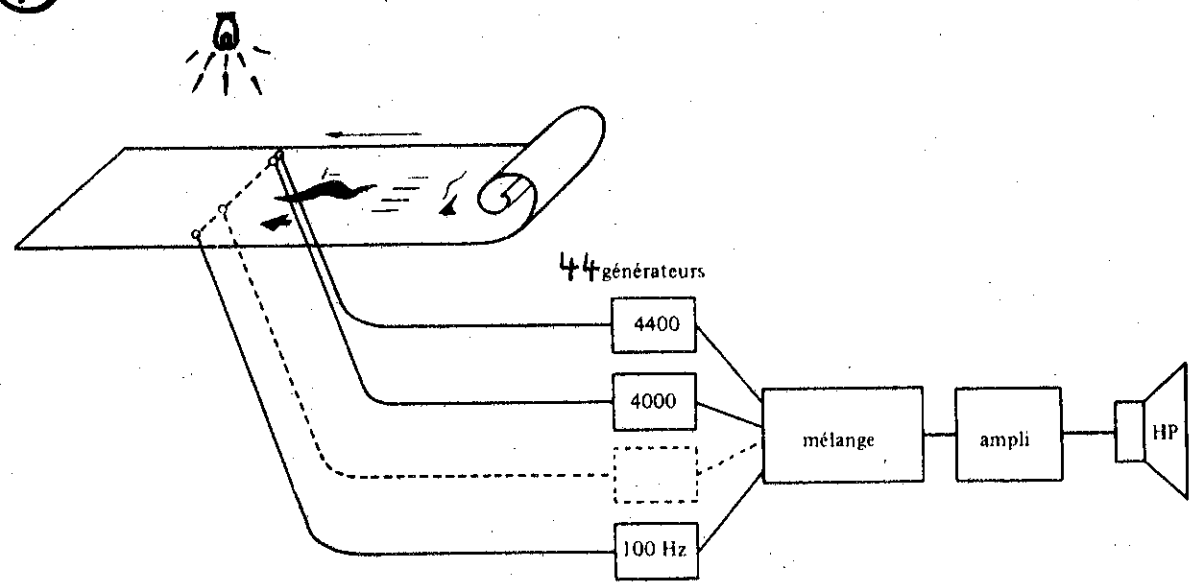
...../

⑥



Voici une diapositive présentée à LANNION, en Juin 1965. On y voit successivement la phrase "la parole est à l'Icophone" en sonagramme de parole normale, en sonagramme à bande d'analyse large (wide, bande d'analyse 300 Hz, au lieu de 45 Hz en analyse normale), en "parole synthétique à la guimbarde, en parole chuchotée, en enregistrement de mauvais dictaphone, puis enfin la synthèse à l'Icophone I. On vérifie bien que toutes ces "images" ont en commun des graphismes identiques dessinés par les formants, et que l'on retrouve dans les 6 analyses, plus ou moins déformées, anamorphosées, altérées, transposées, mais toujours reconnaissables. Or ce sont ces "formes sémantiques" qui véhiculent l'information conditionnant l'intelligibilité de la parole.

⑦



Voici l'ICOPHONE II.

Une feuille de mylar porte le dessin à l'encore noire spéciale représentant le squelette sémantique d'une phrase.Une ampoule éclaire le dessin par le haut et 44 cellules "lisent" ce dessin pendant que la bande défile de droite à gauche.Chaque cellule déclanche un générateur de sinusoïdes;les 44 générateurs sont ensuite mélangés et on entend le signal global à la sortie sur haut parleur.

Les générateurs sont "accordables" en fréquence,et en intensité de façon autonome:on peut donc réaliser à volonté un spectre de raies harmoniques rigoureux par accord au compteur électronique, ou bien un spectre de partiels.Chaque générateur comporte en outre un système "aléatoire" permettant de moduler légèrement la hateur de chaque cellule de façon aléatoire l'une par rapport à l'autre;ainsi on simule assez bien la réalité de la parole chuchotée qui est contituée de bruits...Si l'aléatoire est déconnecté et si les cellules sont accordées strictement par 100 Hz par exemple, on obtient une voix "voisée",recto-tone.Si on passe l'enregistrement d'une phrase oud'un discours ainsi réalisé sur un magnétophone à vitesse continuement variable,on peut simuler l'intonation par "effet de pleurage".

Entre temps, la réalisation des sonagrammes de diphonèmes chuchotés était assez avancée pour que l'on puisse passer à la "synthèse vraie" à partir de digrammes. Melle CASTELLENGO, dont le "dictionnaire" était suffisamment au point, réalisé alors des phrases-tests par assemblage de digrammes, restées célèbres .... (Le petit chat fait sa toilette) et écrivit même des textes beaucoup plus longs, dont certains sont bien connus de nos visiteurs au laboratoire (Bonjour Mesdames, Bonjour Mesdemoiselles, Bonjour Messieurs! Je m'appelle ICOPHONE. J'ai appris par coeur 400 éléments phonétiques à partir desquels je peux tout dire... etc... etc... Je parle français mais aussi d'autres langues ... Hello baby ! How do you do ? Bitte schön, danke schön, und so weiter...'

Dès Juin 1965 une réunion du GAM, en Juin 1966, avait permis à M. LEIPP de faire le point sur nos résultats. Le titre de cette réunion était : "Information sémantique et parole; essai d'une Gestaltttheorie" (Bulletin n° 22). L'essentiel de nos recherches fut annoncé au Congrès International d'Acoustique de BUDAPEST (Octobre 1967. A cette époque le problème de l'anamorphose des formants et du spectre de raies était clairement posé, comme on peut le vérifier dans les comptes-rendus de ce congrès (Le contenu informatif de la parole; par E. LEIPP). J.S. LIENARD avait présenté à ce congrès la machine parlante de KEMPELEN qu'il venait de reconstituer (Bulletin GAM n° 34 Mars 1968; bib. 8). Cette machine nous apprit beaucoup de choses, en particulier que KEMPELEN avait empiriquement très bien compris nombre de problèmes difficiles, celui de la suggestion liée à l'intelligibilité en particulier....

La communication de BUDAPEST sur la structure informative de la parole posait le problème de la quantification en fréquence de la parole par le spectre de raies, ainsi que celui de la hauteur. L'intelligibilité est de toute évidence liée à la hauteur, c'est-à-dire à l'écartement des raies du spectre harmonique de la parole. La question de l'anamorphose formantique, pressentie à LANNION, fut explicitée et les schémas, reproduits dans les Comptes-Rendus de ce congrès, montrent dans quelle voie nous nous engagions dès lors (problème de la phonation, de l'intelligibilité, de l'intonation, etc...).

Il s'avérait entre temps nécessaire de mieux comprendre les mécanismes de production de la parole par l'appareil phonatoire humain. Une série d'études préalables aboutit à une réunion du GAM, en décembre 1967, où ces mécanismes, en particulier, ceux de la production des spectres de raies, c'est-à-dire de l'intonation, furent clairement posés (M. LEIPP, Bulletin GAM N° 32; Mécanique et acoustique de l'appareil phonatoire).

Le problème de l'intonation était posé. On pouvait bien, avec l'ICOPHONE II, réaliser une voix "voisée" en recto-tono (hauteur constante) en accordant les générateurs strictement en raies harmoniques. Le procédé fut encore amélioré par Melle CASTELLENGO, en ajoutant systématiquement le "grave" en dessous du premier formant par noircissement de cette région lors du dessin des icogrammes. Mais il fallait une hauteur variable ! Un premier appareil fut construit par SAPALY, en 1968 dont le principe était simple. Si on quantifie temporellement un signal acoustique quelconque en intercalant des silences périodiquement, toutes les 5, 10, 50 millisecondes, on obtient une "enveloppe" du signal telle qu'on perçoit une hauteur "musicale", respectivement de 200, 100 et 20 Hz.... On fabriqua donc un "hachoir électronique" qui hachait avec une cadence réglable à volonté en cours de lecture les icogrammes, produisant à volonté une sorte de "hauteur synthétique" et permettant de "mettre de l'intonation dans la voix recto-tono, monocorde, de l'ICOPHONE II accordé en spectre harmonique. Le fonctionnement de cet appareil était imparfait, et l'idée fut (provisoirement pensons-nous) abandonnée, au profit d'une autre, susceptible de réaliser l'intonation par d'autres voies. On décida de jouer sur la possibilité de moduler systématiquement et parallèlement les raies du spectre harmonique délivré par l'ICOPHONE, en agissant simultanément sur les générateurs dont on fait monter ou baisser ensemble les signaux de façon à conserver entre eux des intervalles de fréquence égaux. Dès 1968 des essais sur table avaient été faits, et le système fonctionnait correctement avec 4 voies. On pouvait envisager dès lors la construction d'un ICOPHONE permettant la modulation systématique des hauteurs, c'est-à-dire l'intonation.

...../

Entre temps eut lieu le Contrès International d'Acoustique de TOKIO (Août 1968) où le Laboratoire présenta le résultat de ses travaux sur "la synthèse de la parole à partir de digrammes phonétiques" (LEIPP, Melle CASTELLENGO, LIENARD J.S.). Il était annoncé une collaboration avec le CENTRE DE CALCUL ANALOGIQUE du CNRS (C.C.A.) et fait état des premiers résultats obtenus en utilisant un ordinateur (IBM 1130). Ces résultats étaient tout à fait comparables à ceux que donnait l'ICOPHONE II, pour la simple raison que l'ICOPHONE III construit par le Laboratoire de Mécanique (SAPALY) était strictement identique à l'ICOPHONE II du point de vue résultats, sauf qu'au lieu d'être excité par des impulsions provenant en tout ou rien des cellules photo-sensibles, il était piloté par l'ordinateur . La mise en mémoire des digrammes phonétiques avait bien entendu demandé beaucoup de temps ainsi que l'établissement des programmes. Les premiers résultats étaient en tout cas encourageants. Il fut possible d'en arriver là grâce à M. le Professeur MALAVARD, Directeur du C.C.A. à la suite d'une réunion, le 2 décembre 1966 nous obtenions les moyens d'accès à l'ordinateur du C.C.A., et ceci détermina la réalisation de l'ICOPHONE III.

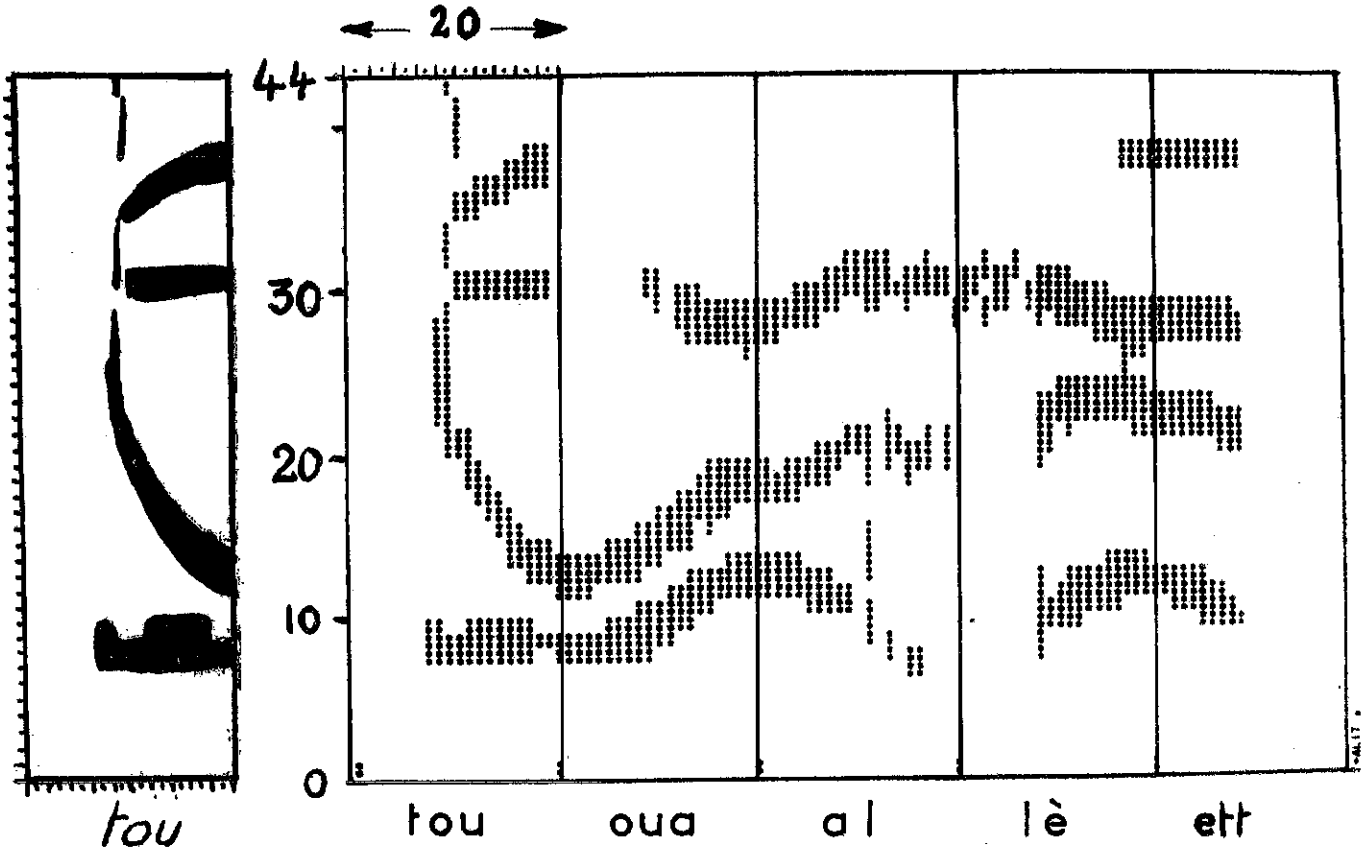
IV - L'HISTOIRE DE L'ICOPHONE III

Comme il vient d'être précisé, cet ICOPHONE est tout à fait [illegible] l'ICOPHONE II, pilotage à part. Il présente cependant sur ce dernier un avantage énorme. Au lieu de dessiner au pinceau les éléments phonétiques à la main, d'assembler les digrammes pour en faire mots et phrases, on met les digrammes, préalablement quantifiés et numérisés, en mémoire. Pour ceux qui ne sont pas familiarisés avec ces mots, rappelons que le principe de la mise en machine des digrammes est simple. On prend le dessin manuel d'un digramme et on divise le rectangle qui l'encadre en un certain nombre de "quantas" acoustiques, de "grains de son" (fig.8). Dans le sens de la hauteur (fréquences, en ordonnée) le problème est résolu automatiquement, puisqu'il y a 44 cellules lectrices qui "quantifient" donc l'échelle des fréquences en 44 parties égales. Dans le sens de l'abscisse (durée) on découpe des "quantas temporels" de 5 milliseconde par exemple. Comme nous avons normalisé le digramme à 100 ms, celui-ci a donc 20 "quantas temporels". Nous avons choisi 5 ms parce que l'oreille est capable d'identifier un grain de son dès qu'il a cette durée. On a donc finalement un quadrillage de 44 x 20 = 880 "grains de son" par digramme phonétique (dont chacun dure 100 ms). (figure 8).

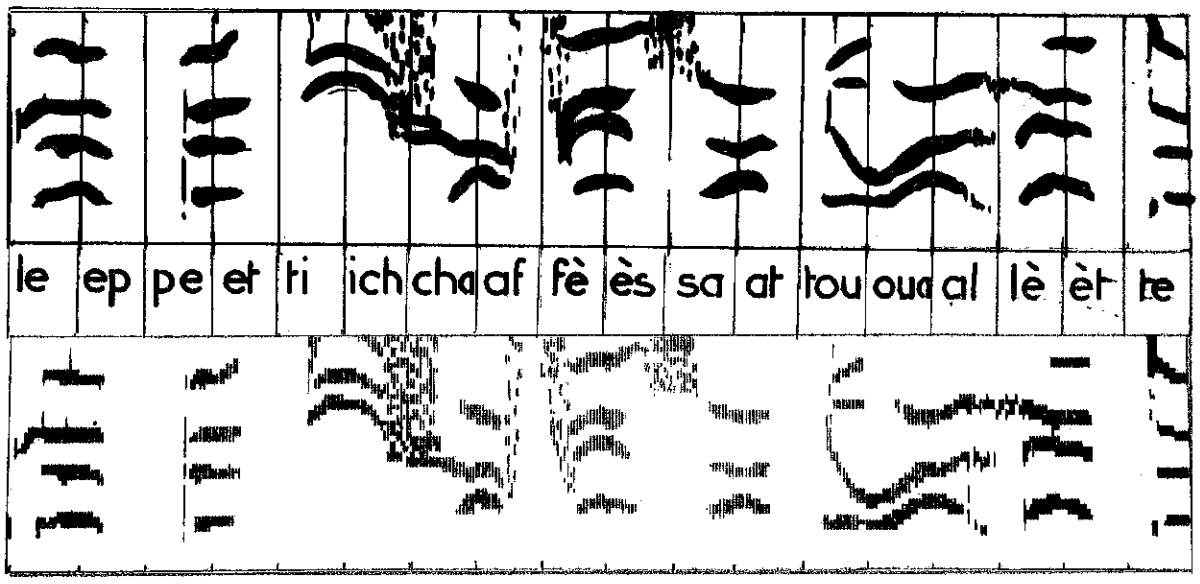
Un digramme phonétique avec ce quadrillage du type "mots croisés", peut donc facilement être décrit par des séries de nombres. Une case ne peut être que noire ou blanche. Chaque fois qu'une case est noire, on peut l'identifier par deux nombres : l'un en horizontal, l'autre en vertical. Si on peut décrire une forme de digramme par des nombres, on peut la mettre en mémoire. Il suffit d'attribuer à ce digramme une "adresse" (par exemple un "numéro d'appel" entre 1 et 900 s'il y a 900 phonatomes. Une touche du clavier de l'ordinateur peut dès lors provoquer l'appel, amener le digramme phonétique (ou une série de digrammes accolés) sous formes d'impulsions. Celles-ci déclanchent les générateurs de l'ICOPHONE, qui excitent le haut-parleur. On entend alors le diphonème ou le mot.

Il va sans dire que cette opération suppose un gros travail de programmation et de mise en mémoire préalable : après la quantification et numérisation, il fallait entrer manuellement chaque point en mémoire (une "entrée" graphique plus expéditive est en cours d'essai....). Avec 44 quantas fréquentiels et 20 quantas temporels de 5 ms cela fait 880 points par phonatome et 742 000 points pour l'ensemble des digrammes phonétiques.... Mais l'opération était possible : elle fut faite (QUINIO - TEIL), et à notre réunion G.A.M. sur l'Intelligibilité (bib. 12) le 8 Novembre 1968, furent présentés les premiers échantillons de parole réalisée à l'ordinateur. Le fait de pouvoir réaliser de la parole synthétique fluide par frappe sur le clavier de l'ordinateur, à la vitesse de la frappe normale en machine à écrire était intéressant : la synthèse de parole chuchotée ou de parole voisée recto-tono devenait donc quasi-instantanée. On [illegible]

...../



A gauche,un "élément phonétique" ,un "digramme phonétique" (tou) représentant un élément de la parole,un "phonatome".Le "dictionnaire" comporte 900 éléments de ce genre.On peut "quantifier" et numériser ces éléments (20 x 44 "cases"...),donc les entrer en mémoire d'ordinateur,chaque digramme étant alors "appelé" par une touche du clavier:on parle en frappant un texte sur le clavier....



En haut,la phrase "le petit chat fait sa toilette" dessinée en icogramme à partir des digrammes phonétiques (Dictionnaire Melle Castellengo);en bas,la même phrase quantifiée,telle qu'elle sort de l'imprimante de l'ordinateur.

effectivement un texte sur le clavier, et avec un décalage de quelques secondes, on entend le texte, en parole fluide, continue, pratiquement en temps réel. Grâce à l'équipe du C.C.A., la méthode de synthèse devint rapidement opérationnelle.... Par la suite, un programme de frappe littérale fut mis au point (on ne frappe plus "Monsieur" en phonétique, c'est-à-dire "me es si ieu", mais on écrit le mot normalement avec ses 8 lettres.

Entre temps (1969), TEIL soutint au C.N.A.M. son diplôme d'ingénieur sur le travail de mise en machine qu'il avait fait en synthèse de parole. Un contrat D.R.M.E. vint providentiellement nous apporter une aide, nous permettant de développer les recherches au C.C.A.

L'intérêt de l'ordinateur, pour nous, au laboratoire, était évident. Des trois "dictionnaires" de digrammes phonétiques que nous avions réalisés au laboratoire (Melle CASTELLENGO, J.S. LIENARD et LEIPP), un seul était suffisamment au point pour être utilisable au C.C.A. : celui de Melle CASTELLENGO, qui avait d'ailleurs besoin de nombreuses corrections encore. Avec l'ordinateur, ces corrections deviennent beaucoup plus expéditives, car à mesure que l'on frappe un texte au C.C.A., l'ICOGRAMME numérique s'inscrit sur un écran de visualisation, et l'on peut alors y faire des corrections assez aisément grâce à un "crayon magique" permettant d'effacer et de remettre tel ou tel détail dans les icogrammes. La correction du dictionnaire ne se fit cependant que peu à peu.

Entre temps se posèrent bien entendu les problèmes d'intelligibilité et de tests d'intelligibilité de cette parole synthétique. Où en étions-nous de ce point de vue ? Une étude générale sur les problèmes de l'intelligibilité fut entreprise et aboutit à une réunion GAM, en novembre 1968 (L'intelligibilité de la parole : par E. LEIPP; Bulletin N° 37 et bib. 13). Un certain nombre de points importants apparurent, en particulier celui de la prévisibilité et de son rôle capital en intelligibilité. Il fut fait état à cette occasion de recherches systématiques que nous avions faites auparavant sur la question de l'émergence d'un signal sur le bruit de fond, (bib.14) ainsi que des recherches que nous avions faites sur l'intelligibilité dans des amphithéatres de la Faculté des Sciences, et qui avaient fait l'objet de plusieurs rapports au Doyen de la Faculté (M. ZAMANSKI). Nous avions alors été conduits, pour obtenir des résultats significatifs, à utiliser des textes à haute imprévisibilité, comprenant des mots "sonnant" français mais n'existant pas dans le dictionnaire (RABELAIS, MICHAUX etc...) et aussi des textes de mots français mais associés en textes obscurs où la prévisibilité des mots était quasi nulle (MALLARME). Il fut montré que tout test, les logatomes par exemple, où la prévisibilité des phonatomes est importante, n'a plus de signification. L'idée se fit dès lors de réaliser des mots synthétiques à l'ordinateur, en partant des taux d'occurence étudiés au début. Ces mots sonneraient "français" par définition, mais n'existeraient pas dans le dictionnaire français. La prévisibilité étant très faible, les tests seraient alors beaucoup plus "méchants" qu'avec des logatomes, mais ils auraient une signification. Nous avons entre temps fait quelques essais dans ce sens; lorsqu'on trouve 90 % d'intelligibilité avec des logatomes usuels, on n'a plus guère que 60 ou 65 % avec ces mots, mais il suffit de multiplier par un coefficient donné (qui reste à déterminer) pour retrouver le coefficient d'intelligibilité dans le langage courant.

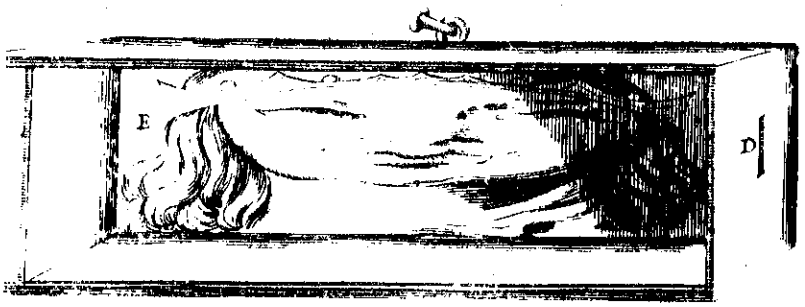
L'intérêt de tels tests en d'autres domaines (tests avec sourds) est évident : on utilise une voix normalisée (celle de l'ICOPHONE) et on supprime la prévisibilité. C'est la seule condition pour tester vraiment ce qu'un sourd partiel entend vraiment. Tout cela est en cours de développement, et l'on trouvera d'autres détails plus loin dans les autres communications de ce colloque. Lors de ces recherches sur l'intelligibilité, un problème important allait cependant émerger, et qui mérite ici des développements : celui des anamorphoses.

...../

⑨

ANAMORPHOTIQUE _1642_

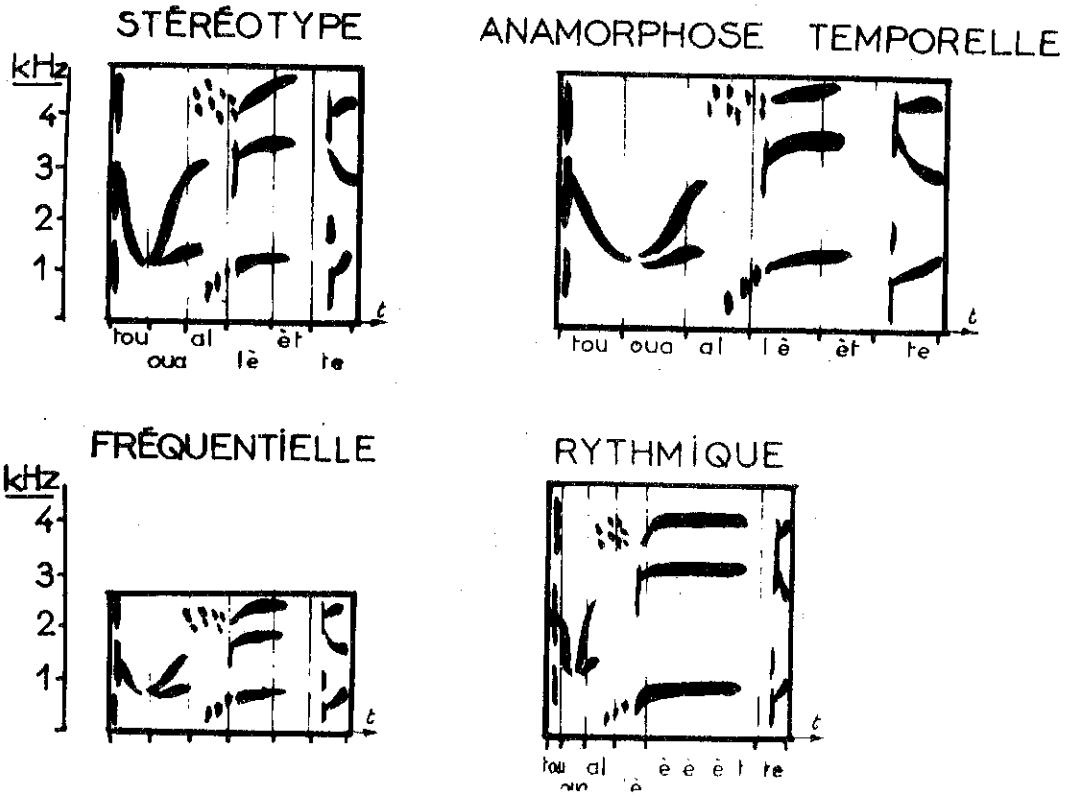
a



d'après Jurgis BALTRUSAITIS . "Anamorphoses" Olivier PERRIN Ed.

Les artistes peintres ont réalisé des anamorphoses par divers procédés depuis longtemps....Dans cette "boite à anamorphoses", on regarde par la fente D et on voit l'image normale de la jeune fille.

b



Les anamorphoses de la parole sont de trois sortes: temporelle, fréquentielle ou rythmique. On peut ramener chacune de ces formes anamorphosées du mot "toilette" dans le "cadre" d'un stéréotype donné. C'est un problème de mathématiques, donc soluble à l'ordinateur.

V. L'HISTOIRE DES ANAMORPHOSES

Le problème théorique des anamorphoses, c'est-à-dire des transformations d'une forme en une autre, a préoccupé les chercheurs depuis bien longtemps, mais à peu près exclusivement dans le domaine visuel, où les choses sont plus simples parce qu'on peut les toucher du doigt... En acoustique il a fallu attendre la visualisation des sons, et plus spécialement le sonagraphe ....

Au moment de la Renaissance, et longtemps après, de nombreux artistes, peintres, graveurs s'adonnèrent à la mode des anamorphoses et l'on connait à ce sujet de très nombreuses gravures et tableaux qui ne laissent aucun doute de ce point de vue. On savait parfaitement anamorphoser des formes en utilisant soit des procédés matériels (fils et perspective) soit des procédés géométriques (carrelage déformé) soit tout simplement les mathématiques. Nous avons montré quelques images typiques à la réunion du GAM. C'est par exemple ce portrait de jeune femme anamorphosé par BETTINI en 1642 (fig.9a), et dont on retrouvait l'allure réelle en regardant l'image de côté. Nous avons montré ce curieux "portrait des Ambassadeurs" peint en 1533 par HOLBEIN et où, entre le portrait grandeur nature de deux ambassadeurs français en Angleterre, flotte un objet de forme indéfinissable.... sauf si on se place tout près du mur où se trouve la peinture et si on regarde le tableau sous un angle rasant. L'objet devient alors un crâne, et le tableau a une signification philosophique : il montre la vanité des choses de ce monde lorsqu'on les regarde sous un certain angle ....

On fit aussi, dès cette époque, des "anamorphoses cylindriques". Les formes dessinées, généralement impossibles à identifier lorsqu'on les regarde normalement, deviennent instantanément reconnaissables si on les observe par réflexion sur un miroir cylindrique. De même on réalisa des anamorphoses coniques, et tout le monde connait finalement les "glaces déformantes" devant lesquelles s'amusent les enfants au Jardin d'Acclimatation ou au Musée Grévin... Une chose est certaine, c'est que toutes ces anamorphoses peuvent être simulées par l'ordinateur pour peu qu'on en connaisse le "secret" mathématique, y compris les déformations compliquées des glaces. Par le mécanisme des anamorphoses les formes peuvent être étirées en long, conservant par ailleurs toutes leurs proportions; ou en large .... Une glace de courbure complexe donne évidemment une anamorphose "rythmique", dans le sens d'anamorphose des proportions... Dans tous les cas on peut soit simuler les déformations, soit ramener les images déformées à leur état normal par le calcul, donc par l'ordinateur.

Lorsqu'on considère le problème de la parole, telle qu'elle est articulée normalement par divers locuteurs, on est frappé par le fait que chacun "dessine" ses formes sémantiques à sa façon, tout comme c'est le cas pour notre lapin (fig.4, LANNION). Nous nous sommes alors demandé comment notre système auditif pouvait s'y retrouver, car dès lors on ne peut plus définir une forme sémantique de parole par des fréquences ou des durées absolues. Les réflexions que nous avions été amenés à faire au fur et à mesure de l'avancement de nos travaux, nous avaient conduits à concevoir, dans le système auditif, un dispositif anamorphoseur, permettant à volonté de transformer les formes sémantiques réalisées par les divers locuteurs afin de les ramener dans un cadre fixe, ayant des dimensions standardiées. Dès lors il devenait facile de reconnaitre les formes sémantiques de n'importe quel locuteur, par corrélation, en les comparant, avec une référence apprise et stockée dans notre mémoire humaine.

Nous avons donné à plusieurs reprises le schéma de ce que nous pensons être un modèle fonctionnel cohérent du système auditif, en particulier dans le Bulletin n° 37 du GAM, relatif à l'intelligibilité. Lors de cette réunion, nous avions d'ailleurs abordé systématiquement le problème de l'anamorphose de la parole chez des locuteurs variés. On trouvera tous détails dans la communication publiée dans la REVUE D'ACOUSTIQUE parue récemment (bib.12), ainsi que dans la communication au Congrès d'acoustique de BUDAPEST (bib.8).

...../

Le problème peut être résumé ainsi. Les proportions de l'appareil phonatoire varient d'un individu à l'autre ainsi que les programmes de mouvement appris pour parler. La parole comporte trois types d'anamorphose, schématisées par la figure 9b. Tout se rapporte ici à un stéréotype défini arbitrairement, par exemple à celui du mot "toilette" réalisé dans notre parole normalisée à l'ICOPHONE. En ne considérant que les formes sémantiques à l'exclusion du spectre de raies, on peut observer :

- une anamorphose temporelle. On parle plus lentement : les digrammes phonétiques normaux sont tous allongés dans le sens horizontal de façon parfaitement similaire. On reconnait toujours la forme du mot "toilette"....

- une anamorphose fréquentielle. La forme est étirée (ou comprimée, comme c'est le cas ici) dans le sens de la hauteur; mais elle reste parfaitement reconnaissable et se distingue de toute autre forme.

- l'anamorphose rythmique. C'est la plus compliquée : les digrammes phonétiques ont à présent des longueurs, des durées inégales; le rythme interne du mot est modifié.

Si l'on fait intervenir à présent la hauteur de la voix, c'est-à-dire le spectre de raies, on peut évidemment quantifier les formes ci-dessus par des raies plus ou moins écartées : c'est l'anamorphose de l'intonation, de la hauteur de la voix.

Voici (figure 10) des figures qui résument la communication de Budapest (bib.8 - 1967).

On prend ici la phrase "mais oui" dont la forme sémantique est une sorte de double "v". Cette forme peut être étirée dans tous les sens, par exemple allongée vers le haut, tout en étant modifiées par une anamorphose rythmique légère : c'est une voix de femme ou d'enfant par exemple. Ceci représente bien entendu de la parole chuchotée... En parole normale, la forme sémantique sera quantifiée soit par un spectre aigu (écarts larges entre harmonique) soit par un spectre grave (raies serrées), soit par un spectre qui change de hauteur en cours d'élocution. Nous avions montré ailleurs que, plus le spectre est serré, plus la forme est précise, et plus le mot est intelligible (plus la forme est facile à reconnaitre). Cette constatations s'applique tout particulièrement au chant lyrique, sur lequel nous avions fait une étude détaillée (bib.13).

Pourquoi tout cet intérêt pour les anamorphoses ? Eh bien, tout simplement parce que la reconnaissance automatique de la parole d'un locuteur quelconque ne sera résolue que dans la mesure où l'on aura maitrisé le problème des anamorphoses. Très conscients de ce point de vue, nous avions provoqué une réunion "secrète" avec le Professeur SIESTRUNCK et le Professeur MALAVAR ainsi que les intéressés du C.C.A. dès le 10 février 1969. Ce problème y fut posé et les bases d'une étude systématique sur les anamorphoses furent exposées et définies.

Les anamorphoses peuvent être réalisées avec l'ICOPHONE II, en analogique, et en utilisant diverses manipulations. On peut produire, par simple dessin sur mylar, les anamorphoses temorelles, fréquentielles et rythmiques de parole chuchotée; on peut de même réaliser des anamorphoses de hauteur, en écartant plus ou moins les spectres par réglage individuel de la fréquence des générateurs de l'ICOPHONE; cela permet d'avoir aisément une voix à l'octave avec les mêmes formes sémantiques, une voix un ton au-dessus etc... On peut encore faire de l'anamorphose rythmique en enregistrant, comme l'a fait Melle CASTELLENGO, de la parole recto-tono réalisée à l'ICOPHONE II, puis en faisant défiler la bande avec des vitesses variables et fluctuantes, réalisant ainsi une sorte de pleurage artificiel qui simule plus ou moins l'intonation.

Mais toutes ces opérations sont peu praticables en analogique sans une perte de temps considérable. L'idée s'est donc rapidement faite que les anamorphoses seraient avantageusement réalisées à l'ordinateur. Il suffirait de réaliser les programmes d'étirement en hauteur et en largeur des formes des digrammes phonétiques, puis de réaliser de même une anamorphose rythmique par programme adéquat, modifiant les rapports de durée à l'intérieur des phonatomes. Pour l'anamorphose du spectre de raies, l'ICOPHONE III (périphérique de l'IBM 1130) est bien entendu inutilisable : il faut reconstruire un

.../...

autre ICOPHONE permettant la modulation de hauteur par programme réglant l'écartement des raies. Les essais préalables avaient été faits au laboratoire dès 1968, comme il a été précisé plus haut; il suffira de passer aux actes... Cet ICOPHONE IV fut donc réalisé, comme les précédents, au Laboratoire de Mécanique (SAPALY) et fonctionne depuis quelques semaines, ce qui a permis de fabriquer d'ores et déjà des textes avec intonation, et même avec chanson populaire (A la claire fontaine...) chantée par l'ordinateur avec son ICOPHONE IV, comme ont pu s'en assurer les personnes qui ont assisté à notre colloque.

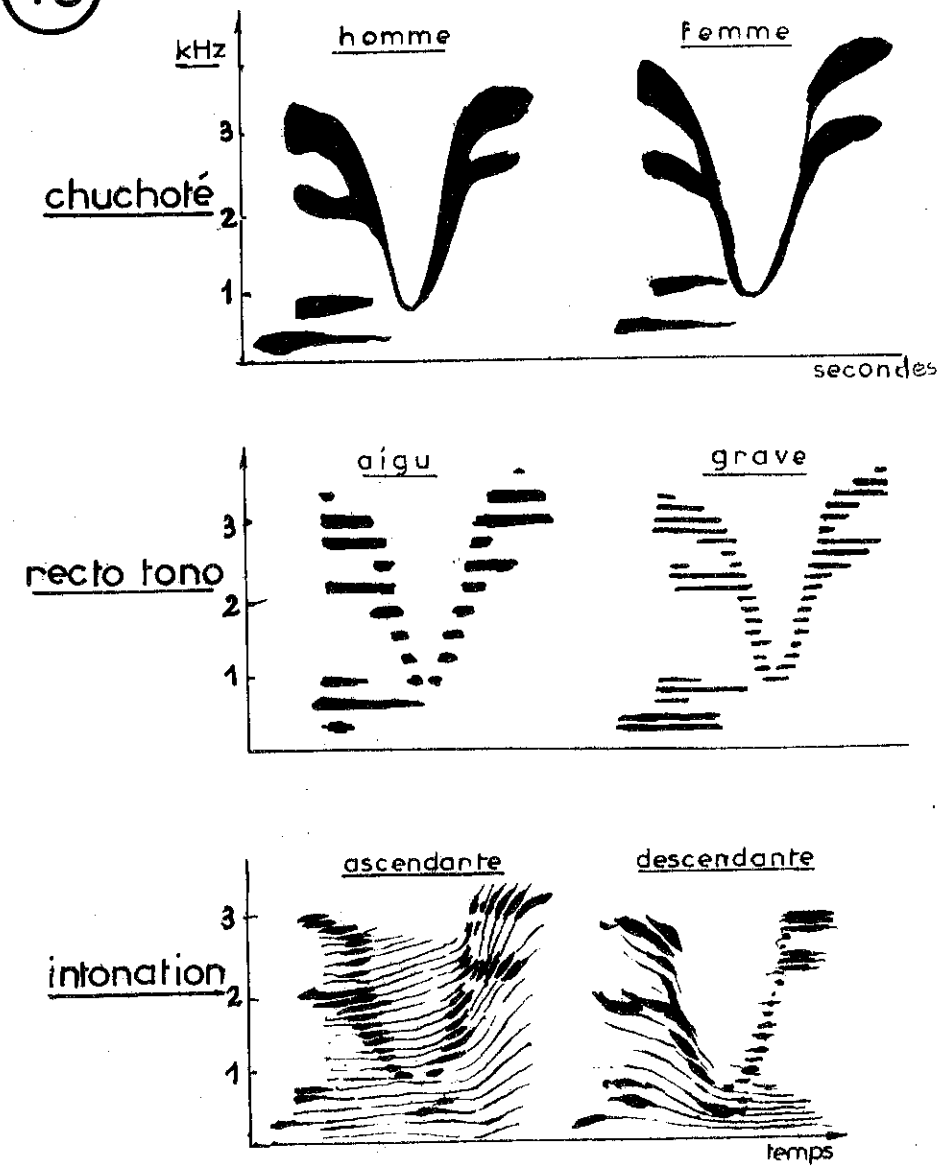
L'ICOPHONE IV est encore un appareil trop récent pour qu'on puisse écrire son histoire; mais les exemples sonores réalisés par Melle CASTELLENGO et présentés lors de notre réunion étaient pensons-nous, suffisamment démonstratifs pour que nous puissions compter sur des résultats ultérieurs.....

VI. PROSPECTIVE ET CONCLUSIONS

Les recherches que nous avons faites jusqu'à présent sur la synthèse de la parole, et plus spécialement celles qui concernent les anamorphoses ont un but lointain... Ce but, que beaucoup poursuivent actuellement, est celui de la reconnaissance automatique de la parole d'un locuteur quelconque, préalable impératif pour attaquer le problème du "dialogue" avec l'ordinateur. Il nous reste à faire de nombreuses recherches, par exemple sur l'intonation et sur les procédés à utiliser pour anamorphoser la parole d'un locuteur quelconque de manière à faire cadrer les phonatomes produits par celui-ci dans le cadre normalisé de la parole synthétique que nous fabriquons. Dès lors, la reconnaissance de la parole ne posera plus que des problèmes de mise en mémoire des mots et phrases types que l'on veut reconnaitre, et la reconnaissance se fera par "effet de masque" ou, si on préfère, par des opérations réalisables désormais en machine (corrélation et autocorrélation sur des formes fréquence-temps). Des travaux sont en cours, qui sont susceptibles d'aboutir à des résultats intéressants dans la mesure où nous aurons les moyens de travail et le temps suffisant, les uns et les autres, pour compléter nos connaissances et pour acquérir les données analogiques qui nous manquent et qui sont indispensables pour traiter le problème en machine. Comme Melle CASTELLENGO et M. J.S. LIENARD vont nous exposer plus loin les recherches qu'ils ont faites en jonction avec le C.C.A. et que les informaticiens du C.C.A. eux-mêmes (MM. TEIL - MLOUKA - CALINET - BRAURE) ainsi que M. SAPALY vont nous résumer les recherches et les réalisations qu'ils poursuivent, je pense inutile d'insister davantage sur toutes ces questions. Mon but, ici, était surtout de mettre en lumière les idées directrices et les faits qui ont conditionné nos cheminements respectifs dans le domaine de la parole et qui ont abouti aux résultats que j'ai exposés. Peut-être aurons nous réussi à convaincre les incrédules que l'apport de non spécialistes peut être parfois non négligeable. Comme en synthèse, nous espérons d'ailleurs pouvoir le montrer dans le domaine de la reconnaissance automatique de la parole d'un locuteur quelconque pour un vocabulaire illimité et en toutes langues. C'est un but ambitieux, mais comme chacun sait, "il n'est pas nécessaire d'espérer pour entreprendre, ni de réussir pour persévérer"....

E. LEIPP

10 Février 1971



Voici, résumées en une figure, les résultats annoncés dans la communication au Congrès International de BUDAPEST (LEIPP, bib 8):

Les anamorphoses de la parole concernent deux aspects :

-l'anamorphose des formes sémantiques, dessinées par les formants. Ainsi la voix d'enfent ou de femme est généralement placée plus "haut"(les formes sémantiques sont étirées plus ou moins vers le haut.

- l'anamorphose du spectre de raies. La même voix(d'homme, ici) peut être "quantifiée" par un spectre de raies écartées (voix aiguë en recto-tono) ou par un spectre de raies resserrées (voix grave, recto tono). On notera que la forme apparait d'autant mieux que le spectre est plus serré (voix graves). Lorsque la hauteur évolue (écartement variable des raies), on réalise de l'intonation.

BIBLIOGRAPHIE

réduite à nos publications concernant la parole :

1°) LEIPP (E) et MOLES (A) - L'emploi du sonagraphe dans la détermination de la qualité d'un instrument de musique. Communication au Congrès International d'Acoustique, Lausanne 1957.
in : Annales des Télécommunications T. 14, 5-6 (1959) p. 135.

2°) LEIPP (E) et MOLES (A) - Méthode objective d'appréciation des qualités d'un instrument de musique.
Comptes-Rendus du Congrès International d'Acoustique de STUTTGART.
Proceedings of 3. ICA. Elsevier, Amsterdam - T2 (1961) p. 752-755.

3°) LEIPP (E) - La cavité buccale, paramètre sensible des spectres rayonnés par les instruments à vent.
C.R. 4° ICA, Copenhague (1962) p. 51 sequ.

4°) LEIPP (E) - La guimbarde.
Revue du Son n° 126 (1963) CHIRON, Paris.

5°) LEIPP (E) - Etude acoustique de la guimbarde
Acustica Vol. 13, n° 6 (1963) p. 382 sequ.

6a) LEIPP(E) - Un vocoder mécanique, la guimbarde.
Annales Telecom. T. 18 n° 5-6 (1963) p. 82 sequ.

6b) LEIPP (E) et John WRIGHT - La guimbarde.
Bulletin n° 25 du G.A.M. (Janvier 1967).

7°) LEIPP (E) - Structure physique et contenu sémantique de la parole.
Exposé au Colloque GALF sur la parole, Grenoble (juin 1967)
Revue d'Acoustique n° 3-4, Paris 1968.

CASTELLENGO (M) - La synthèse de la parole à l'Icophone.
Les problèmes de la perception d'une voix synthétique. Grenoble 1967. In Revue d'acoustique n° 3-4.

SAPALY (J) - Principe de l'appareillage électro-optique ICOPHONE II.
(Grenoble 1967). in Revue d'acoustique 3-4.

LIENARD (J.S.) - Le rôle des éléments phonétiques dans la synthèse de la parole et leur importance en linguistique quantitative .
(Grenoble 1957) in Revue d'Acoustique n° 3-4.

8°) LEIPP (E) - Le contenu informatif de la parole.
C.R. du Congrès international d'Acoustique, (1967) - BUDAPEST.

9°) LIENARD (J.S.) - La machine parlante de KEMPELEN.
Bulletin GAM n° 34, mars 1968. Ed. interne Faculté des Sciences de Paris.

10°) LEIPP (E) - Mécanique et acoustique de l'appareil phonatoire.
Bulletin GAM n° 32 (décembre 1967).
in Revue d'acoustique (Paris), n° 5 - p. 11 sequ.

11°) LEIPP (E) - CASTELLENGO (M) - LIENARD (J.S.)-
La synthèse de la parole à partir de digrammes phonétiques.
C.R. 6 - ICA n° C 5-6 Tokio (1968).

12°) LEIPP (E) - L'intelligibilité de la parole
Bulletin GAM n° 37 (Novembre 1968)
in Revue d'Acoustique (Paris), n° 12 (1970) p. 343 sequ.

...../

13°) LEIPP (E) - CASTELLENGO (M).
L'intelligibilité de la parole dans le chant.
Conférences des journées d'étude du Festival International du Son.
Chiron, Paris (1969), p. 15 sequ.

14°) LEIPP (E) - Le problème de la perception des signaux acoustiques par effet de contraste.
Annales Télécommunications - Tome 20, n° 5-6 (1965) p. 103 séqu.

J. S. LIÉNARD

APPLICATION DE L'ORDINATEUR A LA SYNTHESE, LA RECONNAISSANCE ET L'ETUDE STATISTIQUE DE LA PAROLE

JANVIER 1971

N° = 53

GAM

BULLETIN DU GROUPE d'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES . TOUR 66 . PLACE JUSSIEU. PARIS 5e

J.S. LIENARD - GAM Janvier 1971

PAROLE ET ORDINATEUR

APPLICATION DE L'ORDINATEUR A LA SYNTHESE, LA RECONNAISSANCE ET L'ETUDE STATISTIQUE DE LA PAROLE

I - INTRODUCTION

Cet exposé vise à regrouper l'ensemble des problèmes qui se trouvent au confluent de l'informatique et de l'étude de la parole au Laboratoire d'Acoustique. Certains de ces problèmes sont traités de manière plus approfondie dans les autres exposés, dont les références sont données au cours du texte.

Après un bref rappel des idées de base du Laboratoire dans le domaine de la parole, on trouvera quelques notions élémentaires sur le fonctionnement des ordinateurs. Le principe de commande numérique de l'Icophone sera expliqué ensuite, et quelques expériences sur l'anamorphose fréquentielle de la parole seront décrites.

Le problème de la reconnaissance automatique sera posé, tel que nous pouvons l'envisager aujourd'hui, et l'on évoquera enfin un aspect insolite de la recherche qui, partant de simples comptages de phonatomes, aboutit actuellement à la fabrication par l'ordinateur de mots et de textes aléatoires.

II - STRUCTURE ACOUSTIQUE DE LA PAROLE

Les idées développées au Laboratoire sur ce sujet ont été largement diffusées (voir en particulier l'exposé de E. LEIPP et les références bibliographiques n° 1 à 6).

Les voici sous une forme condensée et quelque peu dogmatique :

a) La parole est composée de formes, au sens de la Gestalttheorie, dans lesquelles est inscrit le sens du message (information sémantique).

b) Les formes peuvent être mises en évidence par l'analyse temps-fréquence (sonagramme), et sont particulièrement nettes en voix chuchotée.

c) Elles peuvent être schématisées en "tout ou rien", sans tenir compte de l'amplitude relative des composantes. Le dessin ainsi obtenu est appelé "squelette sémantique" de la phrase. Relu sur un synthétiseur rudimentaire comme l'Icophone, il donne lieu à une parole parfaitement intelligible.

d) Si l'on veut synthétiser la parole sans analyse préalable on ne peut pas utiliser des éléments tels que les phonèmes, dont la plupart n'ont pas d'existence isolée. Par contre on peut utiliser des "diphonèmes" ou "phonatomes", correspondant aux mouvements élémentaires de l'appareil phonatoire.

e) Un "dictionnaire" de 500 à 600 schémas de phonatomes suffit pour fabriquer une parole parfaitement intelligible, par un assemblage comparable à celui des dominos.

La recherche en étant arrivée à ce point, l'amélioration du dictionnaire et l'exploitation de la méthode se révélèrent être des tâches gigantesques, à cause de la lenteur des manipulations : la copie sur bande transparente des phonatomes d'une courte phrase demandait plusieurs dizaines de minutes. C'est pourquoi le procédé a été confié à l'ordinateur.

...../

III - QU'EST-CE QU'UN ORDINATEUR ?

Contrairement à une mythologie largement répandue, l'ordinateur n'est qu'une machine, un outil, le prolongement perfectionné du boulier chinois. C'est une machine électronique capable d'effectuer à très grande vitesse (quelques centaines de milliers par secondes) des opérations arithmétiques et logiques, et surtout capable de mémoriser des nombres (fig.

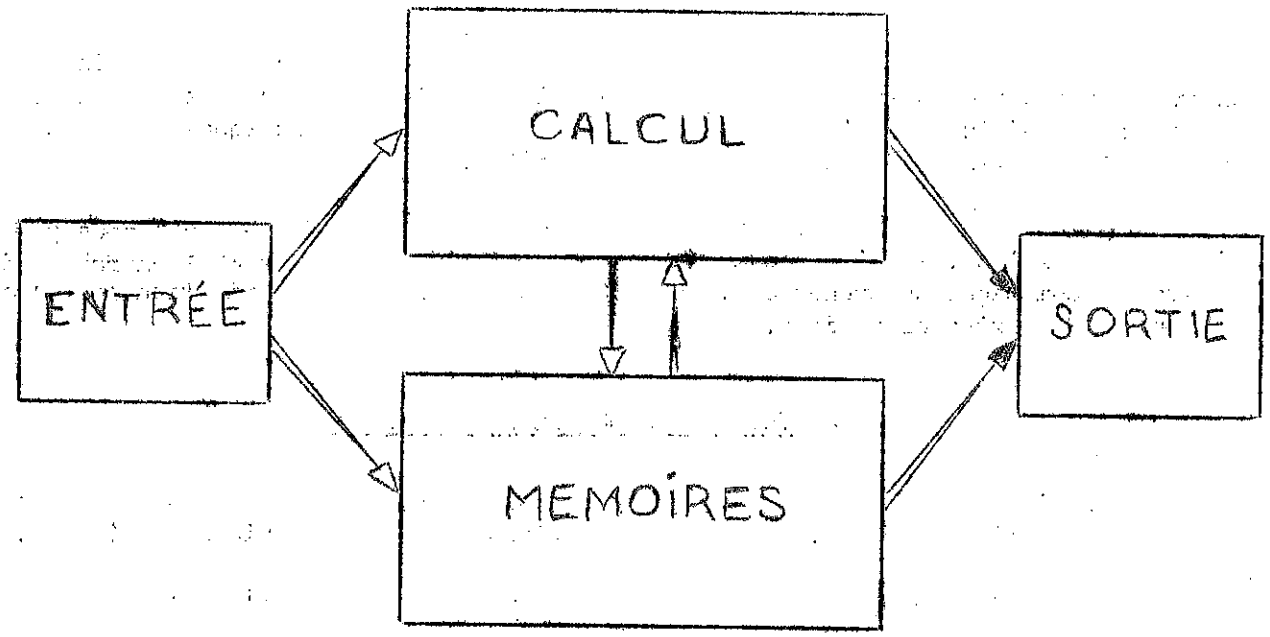


Figure 1

1°) Représentation binaire des nombres.

La notation binaire permet d'utiliser des dispositifs fonctionnant en "tout ou rien", c'est-à-dire avec sécurité et rapidité.

Tout nombre, exprimé habituellement en notation décimale, c'est-à-dire suivant les puissances de 10 (1; 10; 100; 1000 etc..), peut être représenté par sa notation binaire, suivant les puissances de 2. (1; 2; 4; 8; 16; etc...).

Par exemple le nombre 103 qui en notation décimale signifie :
(1 x 100) + (0 x 10) + (3 x 1), sera représenté par le nombre binaire 1100111 qui signifie : (1 x 64) + (1 x 32) + (0 x 16) + (0 x 8) + (1 x 4) + (1 x 2) + (1 x 1).

2°) Organe de calcul.

L'ordinateur peut effectuer à partir de la notation binaire des opérations arith-

...../

méthiques très semblables à celles qu'effectuent les machines de bureau : addition, multiplication, etc... Mais il peut aussi prendre des décisions logiques : selon, par exemple, que le nombre A est plus grand ou plus petit que le nombre B, l'ordinateur peut décider s'il doit ou non continuer le calcul commencé. Ces opérations logiques permettent d'utiliser l'ordinateur à de tout autres tâches qu'au calcul proprement dit.

3°) Mémoires.

C'est dans la possibilité de mémoriser des informations que se trouve l'essentiel de la puissance de l'ordinateur. Un chiffre binaire, 0 ou 1, peut être mémorisé au moyen d'un dispositif "bistable", qui ne peut prendre que deux états. La plupart des mémoires actuelles utilisent des dispositifs à hystérésis magnétique. Un calcul, aussi complexe soit-il, peut alors être exécuté par étapes distinctes, en mémorisant les résultats intermédiaires; quel que soit le nombre de données et de résultats partiels, ceux-ci sont à la disposition de l'organe de calcul.

Lorsqu'on met une information en mémoire, il faut savoir la retrouver : il lui est donc affecté une adresse, explicite ou implicite. Mais suivant la technologie adoptée pour la mémoire, le temps d'accès à l'information peut varier notablement. On est ainsi amené, pour limiter les pertes de temps et les investissements inutiles, à utiliser deux types de mémoires :

- Une mémoire rapide , dont le temps d'accès est comparable à celui d'une opération élémentaire, et qui échange en permanence des informations avec le bloc de calcul. Elle fait partie intégrante de la machine et porte le nom de mémoire centrale. Malheureusement elle coûte très cher, et sa capacité est un facteur déterminant du prix de la machine.

- Une mémoire "de masse", beaucoup plus lente et moins coûteuse, dans laquelle on stocke les informations dont on n'a pas besoin en permanence. Elle est constituée essentiellement par des bandes ou des disques magnétiques.

4°) Entrées-sorties, ou organes périphériques.

L'utilisateur peut introduire des informations en machine, ou en recevoir, par de nombreux moyens : machine à écrire connectée à l'ordinateur, cartes perforées, ruban perforé, imprimante; écrans de télévision etc...

Néansmoins c'est là le défaut actuel de l'ordinateur : les cadences d'entrée-sortie sont limitées, soit par les organes mécaniques (bien qu'une imprimante rapide fournisse jusqu'à 2000 lignes par minute), soit par la capacité d'appréhension de l'utilisateur, qui se trouve souvent face à des monceaux de papier et des quantités invraisemblables de chiffres dont la plupart sont inutiles. Si les échanges d'informations avec l'homme ont une cadence limitée, il n'en va pas de même avec d'autres systèmes ou machines : les informations sont alors transmises à très grande vitesse par des "canaux". L'ordinateur peut alors contrôler en temps réel des processus industriels, des trajectoires de fusées, des organes périphériques de toutes natures.

Dans certaines applications de l'informatique, il serait extrêmement précieux de pouvoir dialoguer oralement avec l'ordinateur, ou de pouvoir donner oralement des ordres à une machine. On verra plus loin que l'Icophone peut être envisagé comme organe de sortie d'ordinateur, et que l'entrée serait une application essentielle de la reconnaissance automatique.

IV - PRINCIPE DE LA COMMANDE NUMERIQUE DE L'ICOPHONE

La commande numérique de l'Icophone a été réalisée au Centre de Calcul Analogique du C.N.R.S. (Professeur MALAVARD, Mr. G. RENARD), par J. QUINIO et D. TEIL.

...../

L'ordinateur du C.C.A. est un IBM 1130, pourvu à l'époque d'une faible capacité de mémoire centrale (8000 mots de 16 bits), et déjà occupé à de multiples tâches; il a pourtant supporté sans inconvénients le raccordement de l'Icophone III, auquel s'est ajouté récemment l'Icophone IV à modulation de hauteur.

Avant de passer à la synthèse proprement dite, par mise en oeuvre des phonatomes, il a fallu introduire ceux-ci dans la mémoire de l'ordinateur, qui ne connaît que des nombres : cette opération est nommée quantification.

1°) Quantification des phonatomes.

Nous avons vu que l'ordinateur manipulait des nombres binaires; les formes acoustiques sont dessinées en "tout ou rien" : le mariage devait être possible sans complication... théorique. En réalité l'entrée des 600 phonatomes en machine a été passablement compliquée par le manque d'appareillage adéquat. Après plusieurs codages, vérifications et retouches, les phonatomes schématisés étaient en mémoire, chacun sous forme de 880 nombres binaires représentant les 44 fréquences de l'Icophone et 20 "évènements" temporels (fig. 2).

2°) Mise en oeuvre (bibliographie : n° 7 à 10).

a) Notation phonétique.

Le message à synthétiser est introduit en machine sous la forme d'une suite de symboles phonétiques; mais il faut un seul symbole par son, sans quoi il serait impossible de savoir si A suivi de N, par exemple, se prononce AN (voyelle, comme dans "PARENT") ou ANN (comme dans "ANNE"). On a donc adjoint aux caractères alphabétiques du clavier certains caractères spéciaux, qui représentent les sons dont la notation alphabétique prêterait à ambiguité.

Voici ce code :

AN ... *	O (sol) .... O	é (été) .....)	yod (bail) ... Y
ON ... /	O (saule) .. +	è (air) ..... (	E final ..... £
IN et UN..1	OU ..... W	CH (hache) .. X	

On n'a pas jugé utile, dans une première approche, de distinguer plusieurs A (celui de "patte" et celui de "pâte") ni plusieurs E (celui de "peur" et celui de "peu"), ces distinctions n'ayant actuellement qu'une faible valeur sémantique. La même raison nous a conduit à utiliser un seul symbole pour les sons IN ("brin") et UN ("brun"). Notons ici que le message parlé est continu entre deux respirations ou ponctuations : en général les séparations entre mots n'ont aucune réalité phonétique. Les espaces introduits par commodité visuelle dans la suite de symboles phonétiques seront ignorés par le programme.

b) Recherche des adresses des phonatomes (fig.3 et 4).

Chaque phonatome représente la transition entre deux phonèmes. Il peut donc être repéré par deux symboles phonétiques consécutifs, choisis chacun dans un répertoire d'une trentaine. Les adresses des 900 phonatomes possibles sont consignées dans un tableau de 30 lignes et 30 colonnes appelé MATOR. Pour la suite phonétique FRAZFON)TIK , par exemple, l'adresse du phonatome RA sera trouvée à l'intersection de la ligne R et de la colonne A; l'adresse du phonatome suivant, AZ, sera trouvée à l'intersection de la ligne A et de la colonne Z, et ainsi de suite.

c) Recherche et édition des phonatomes.

Après consultation de la table d'adresses MATOR, le message est transformé en

...../

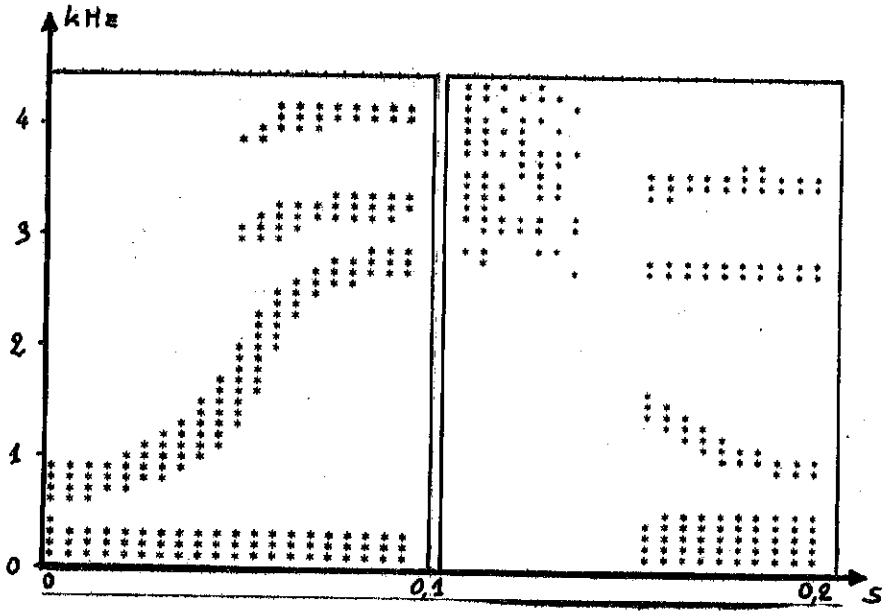


Fig 2 - Phonatomes "OUI" et "SO" discrétisés.

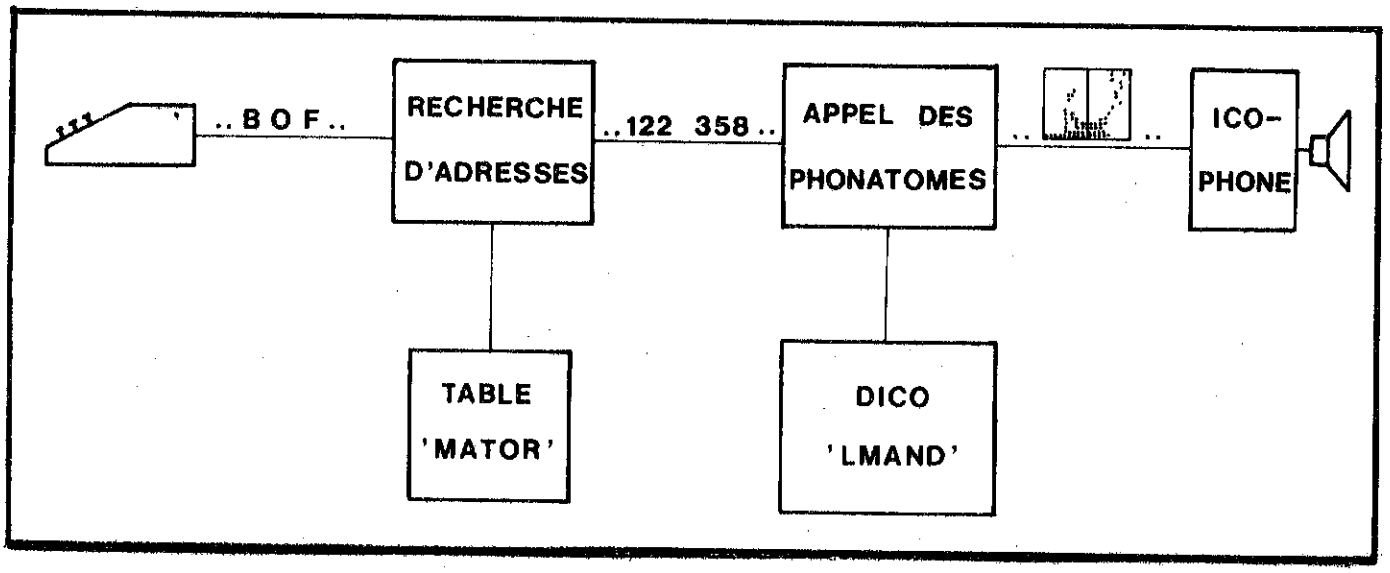


Fig 3 - Principe de la commande numérique

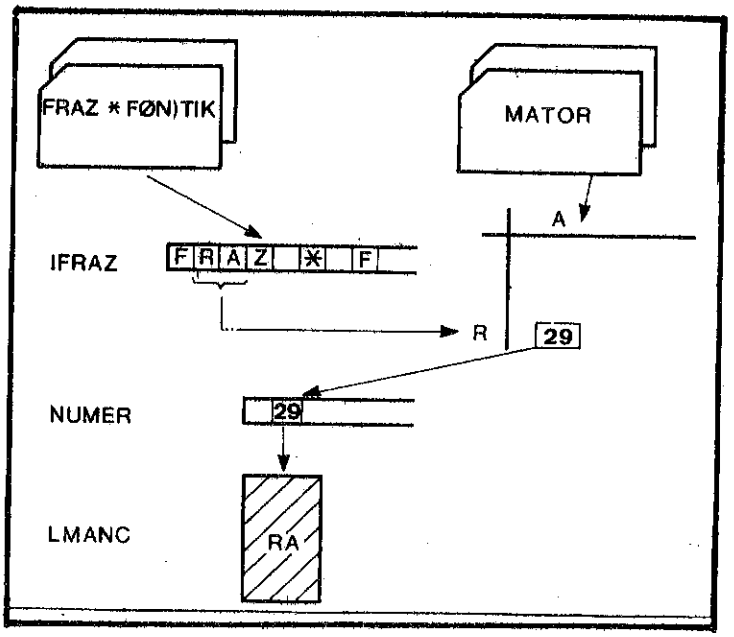


Fig 4 - Recherche de l'adresse d' un phonatome

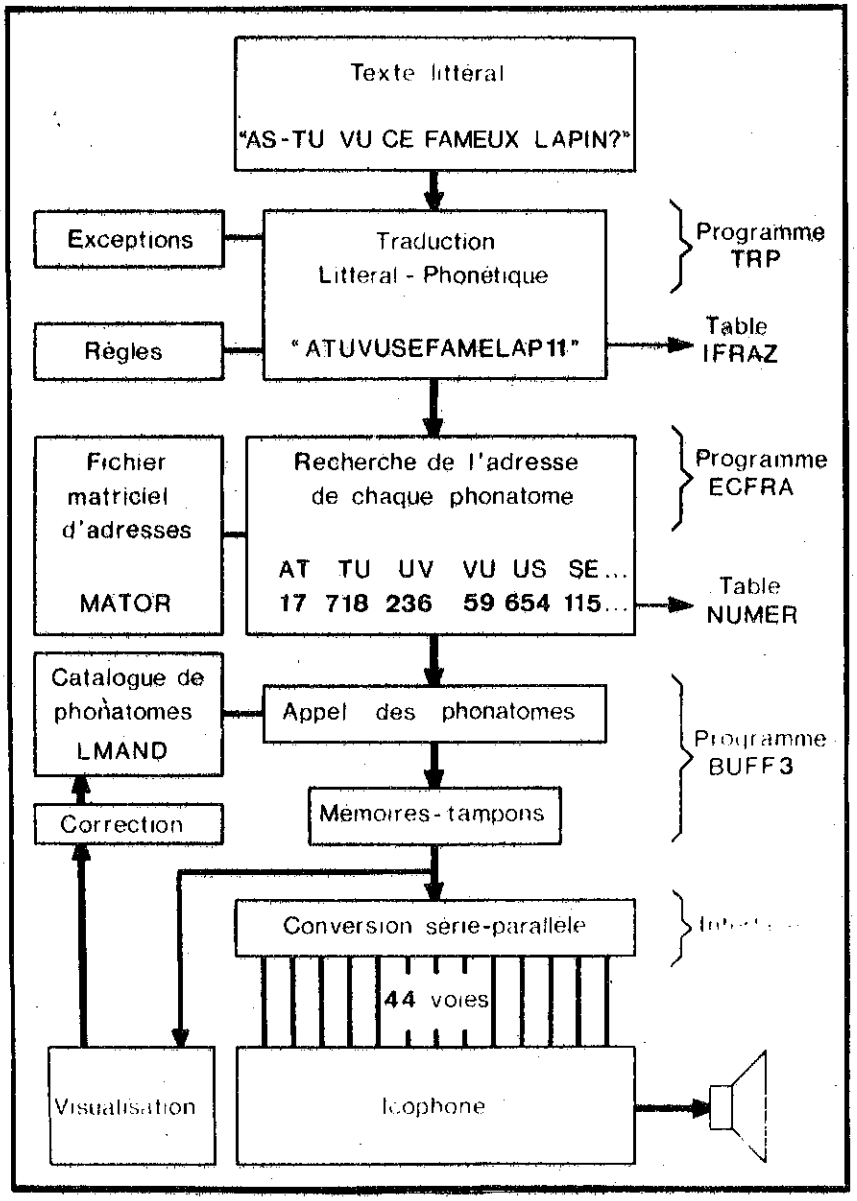


Fig 5 - Schéma d'ensemble de l'Icophone à commande numérique, avec la traduction phonétique et la visualisation.

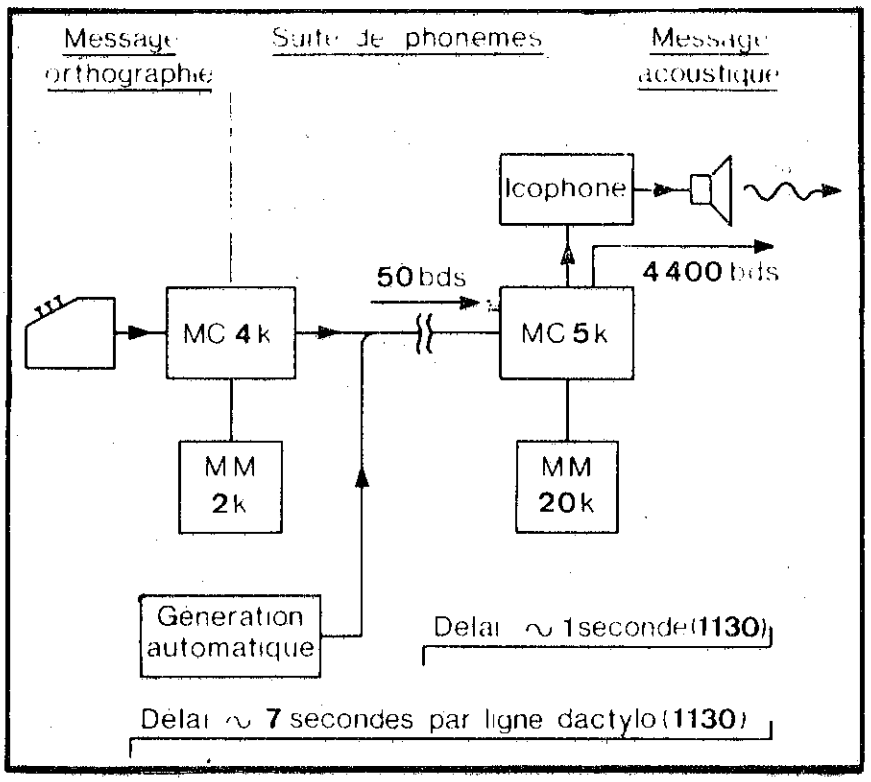


Fig 6 - L'Icophone considéré comme unité de réponse vocale, commandée par un message phonétique de faible débit sémantique (50 bauds).

MC : mémoire rapide

MM : mémoire de masse

une suite de nombres stockés en mémoire centrale, représentant les adresses sur disques des phonatomes à assembler. Le programme va lire sur disque les ensembles de 880 nombres binaires de chaque phonatome, et les transmettre à l'Icophone, par l'intermédiaire du canal et du sous-canal, dans l'ordre imposé par le message.

Pourtant une précaution s'impose lors de cette "édition", car le temps de recherche sur disque diffère fortement d'un phonatome à l'autre : on constitue une zone de mémoire "tampon", dans laquelle on entrepose le message sous sa forme binaire avant de l'éditer. Ici s'introduit donc un certain délai, de l'ordre d'une fraction de seconde, en même temps qu'une limitation de la longueur du message à synthétiser en une seule fois.

Le délai requis par la recherche sur disque a été fortement réduit en choisissant un ordre de rangement tel que les phonatomes les plus fréquents dans la langue française bénéficient des temps d'accès les plus courts. Quant à la limitation de capacité, elle a été supprimée dans le programme "temps réel" dont on trouvera le détail dans l'exposé de D. TEIL.

3°) Améliorations et développements.

a) Traduction phonétique.

La notation phonétique du message à synthétiser, parfaitement légitime si le message est transmis par un télétype ou élaboré par l'ordinateur lui-même, peut paraître contraignante si le message est élaboré par un homme, ou simplement lu par l'ordinateur sous forme orthographiée.

Un programme de traduction phonétique a donc été écrit par D. TEIL : le message orthographié est transformé en une suite phonétique selon les règles de la prononciation française, dans la mesure où les mots à prononciation exceptionnelle sont préalablement connus par la machine. On trouvera des détails sur ce programme dans l'exposé de D. TEIL.

Notons que l'Icophone à commande numérique, muni des programmes de temps réel et de traduction phonétique, a fait l'objet d'un brevet déposé par l'ANVAR (Agence Nationale de Valorisation de la Recherche , biblio.10).

b) Visualisation.

Pour améliorer la qualité de la voix synthétique, il était indispensable de visualiser les phonatomes, et de pouvoir les retoucher point par point tout en les écoutant. Ces possibilités, qui font de l'Icophone un outil de recherche phonétique très puissant, sont dûes à l'utilisation d'une console de visualisation (tube cathodique à mémoire, relié au calculateur par un sous-canal TITN), et surtout à un programme très complet autorisant n'importe quelle modification des phonatomes, avec écoute immédiate (voir les exposés de A. CALINET et de Melle CASTELLENGO).

Actuellement le fonctionnement de l'Icophone peut être schématisé selon la fig. 5, qui intègre les perfectionnements précédents.

c) L'Icophone comme unité de réponse vocale.

L'Icophone numérique peut être décrit selon un point de vue différent de celui que nous avions jusqu'à présent (fig.6). En effet, globalement, l'Icophone reçoit un message composé d'une suite de symboles phonétiques, message dont le débit informationnel est égal au débit sémantique de la parole courante, c'est-à-dire environ 50 éléments binaires par seconde (50 bauds). Il restitue ce même message sous forme acoustique avec un délai de l'ordre de la seconde, qui pourrait être ré-

...../

duit par l'utilisation d'une technologie appropriée.

En définitive, l'Icophone peut être considéré comme un périphérique d'ordinateur, commandé par un message phonétique. Celui-ci peut provenir d'une ligne de transmission peu coûteuse car de faible débit, ou être élaboré par le calculateur en réponse à une question formulée dans un certain code (par exemple au moyen d'un cadran téléphonique). La réponse est parfaitement intelligible et ne connaît aucune limitation de vocabulaire ou de durée. Le message de commande peut même être orthographié, et dans cette optique l'Icophone constitue le complément idéal de la machine à lire.

d) L'Icophone IV.

La mise au point du dictionnaire sur l'Icophone III a permis de donner à la voix synthétique une intelligibilité quasi-totale. Il est possible maintenant d'aborder l'étude des caractères esthétiques de la voix, à commencer par l'intonation. C'est dans ce but qu'a été conçu l'Icophone IV, dont les 44 oscillateurs peuvent être modulés en fréquence tout en restant harmoniques (voir les exposés de J.SAPALY et de Mlle CASTELLENGO). Mais en même temps que la fréquence fondamentale, l'Icophone IV, bien programmé, permet de dilater en temps et en fréquence les formes sémantiques de la parole. C'est un appareil anamorphoseur, et nous allons détailler quelque peu cet aspect.

V - L'ANAMORPHOSE

Dans la mesure où les formes sémantiques de la parole relèvent de la Gestalt-Theorie, elles peuvent subir certaines déformations (ou anamorphoses) sans perdre leur caractère de totalité. Cette invariance sémantique est évidente dans le domaine temporel, puisque la même phrase peut être prononcée avec des rythmes et des vitesses variés, sans pour autant perdre sa signification. Ceci du moins dans certaines limites : la parole ralentie ou accélérée d'un facteur 5 sans modification fréquentielle (l'Icophone optique permet de telles manipulations) perd à peu près totalement son intelligibilité; mais ce facteur 5, c'est-à-dire 25 entre extrêmes, est considérable, et ne se rencontre jamais dans la parole normale.

La possibilité d'anamorphose fréquentielle est moins couramment admise : la doctrine classique affirmait jusqu'à ces dernières années que chaque voyelle ou consonne stable était caractérisée par les fréquences exactes de ses premiers formants, (la notion de formant est d'ailleurs fort discutable), et ceci pour tout locuteur, homme, femme ou enfant. Quant aux consonnes transitoires, elles étaient caractérisées dans une optique tout aussi absolue par la fréquence précise du point de convergence des formants dans leur évolution au cours du temps (théorie du "locus").

L'Icophone IV permet d'anamorphoser les formes sémantiques dans leur ensemble, sans se préoccuper de la notion de formant. Nous allons voir de plus près l'anamorphose fréquentielle en synthèse, puis en analyse.

a) Anamorphose de la voix synthétique.

Il faut bien distinguer deux types de hauteurs : la hauteur de la voix, d'une part, c'est-à-dire l'écartement du spectre de raies, et la hauteur du squelette sémantique d'autre part, qui s'apparente à une composante du timbre et ne peut être que repérée par rapport à une voix de référence.

Si nous prenons pour référence la voix de l'Icophone III, avec un fondamental de 100 Hz et le dictionnaire actuel, parfaitement défini et reproductible, nous pouvons définir pour l'anamorphose sur l'Icophone IV un "coefficient d'intonation ", Ci, rapport de la fréquence du fondamental à la fréquence 100 Hz de l'Icophone III, et deux coef-

...../

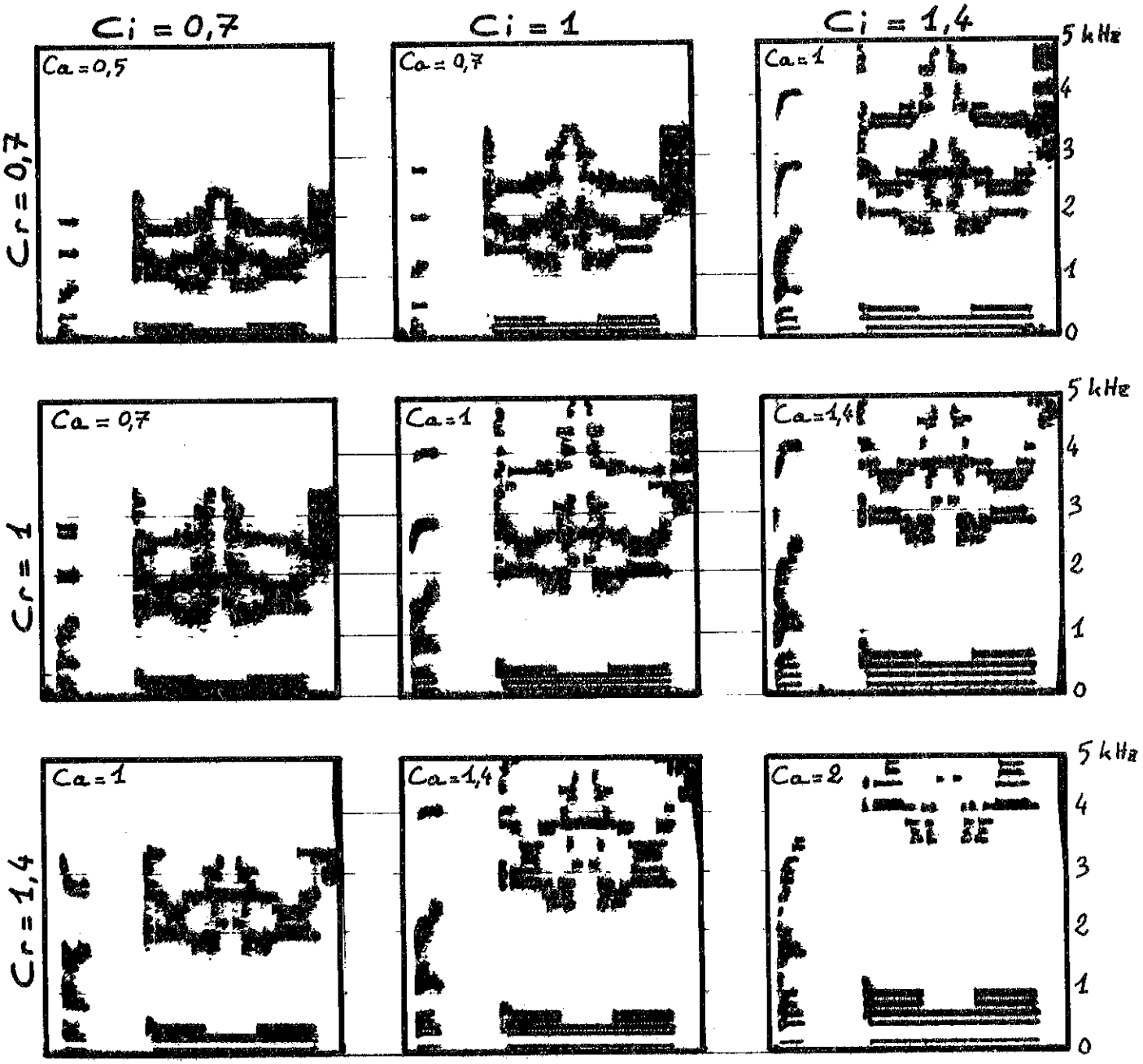


Fig 7 - Diverses anamorphoses de la séquence "ATUVUS"..., synthétisée par l'Icophone IV, à coefficients d'intonation Ci et d'anamorphose relative Cr variables. Le coefficient d'anamorphose absolue Ca est indiqué dans chaque cas.

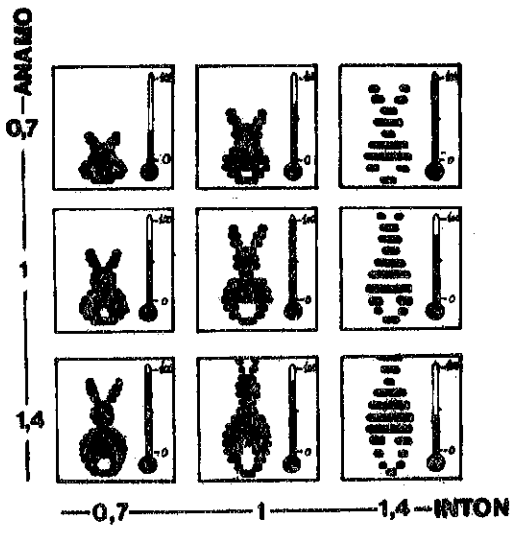


Fig 8 - Analogie graphique avec l'expérience précédente. La température indiquée par le thermomètre représente l'intelligibilité de la voix synthétique (sur des mots aléatoires), ramenée à 100 pour le cas Ca = Cr = Ci = 1.

ficients d'anamorphose. Le premier, coefficient d'anamorphose relatif Cr , est le coefficient par lequel sont multipliées les fréquences du squelette sémantique, le fondamental restant à 100 Hz. C'est ce coefficient Cr qui est pris en compte par le programme ANAMO (voir l'exposé de A. CALINET); le second, coefficient d'anamorphose absolu Ca , est le produit des précédents :

$$Ca = Cr \times Ci$$

En bref, Ci se rapporte à l'Icophone IV, Cr se rapporte au programme ANAMO , et Ca est le coefficient de l'anamorphose résultante.

La figure 7 présente le résultat acoustique du jeu des coefficients Ci et Cr , qui prennent chacun les trois valeurs 0,7 ; 1 et 1, 4 . La phrase synthétique utilisée pour l'expérience était " As-tu vu ce fameux lapin ", et seul le début " ATUVUS " est représenté, pour ne pas surcharger la figure. Le long de la diagonale principale (ici Sud-Ouest, Nord-Est), le coefficient Ca est égal à 1 : le squelette sémantique est inchangé par rapport à la voix de référence; seule change la hauteur du spectre de raies. Celle-ci varie d'une octave le long de la seconde diagonale (N.O - S.E.); donc le coefficient Ca passe de 0,5 à 2, ce qui paraît énorme si l'on pense que la phonétique cherche quelquefois à définir au Hz près les fréquences des formants : celles-ci varient ici dans un rapport 4 (2 octaves). Horizontalement Ci varie seul, pour les valeurs fixées de Cr, et vice versa dans la direction verticale.

L'expérience est schématisée selon une analogie visuelle dans la fig.8, disposée comme la précédente. On a en outre indiqué dans chaque cas par la hauteur de mercure d'un petit thermomètre le résultat de tests d'intelligibilité effectués avec la voix anamorphosée. Ces tests, qui ne prétendent pas à la précision statistique, ont été effectués avec des mots aléatoires (voir § VII), notés phonétiquement par 7 auditeurs variés. Le fait frappant a été le parallélisme des résultats,rendant inutile un test à plus grande échelle, puisque l'on ne cherchait qu'une estimation comparative.

On a pris comme base une intelligibilité de 100 % dans le cas Ca = Cr = Ci = 1 (case centrale de la figure). En réalité le taux d'intelligibilité était alors de 80 %. Mais ce taux, estimé sur des mots aléatoires, correspond effectivement à 100 % sur des mots courants.

Très sommairement, on peut constater que le pourcentage de compréhension varie peu (perte de l'ordre de 20 %) tant que l'anamorphose absolue reste dans les limites de 0,7 à 1,4 et qu'elle n'est pratiquement pas fonction du type de voix synthétique.

On peut d'ailleurs supposer que la perte observée dans les cas extrêmes provient plus de la mutilation des formes sémantiques dûe au programme ANAMO que d'une anamorphose excessive. En effet dans les anamorphoses de coefficient relatif inférieur à 1 une partie de l'information est perdue, puisque comprimer les formes sémantiques revient à diminuer la finesse de résolution fréquentielle. Par ailleurs, dans les anamorphoses de coefficient relatif supérieur à 1, les fréquences élevées sont perdues : une anamorphose de coefficient relatif Cr = 1,5 par exemple opère la transposition de la fréquence n° 30 vers une fréquence n° 45, qui n'est pas prévue sur l'Icophone IV. Ces anamorphoses s'accompagnent donc d'un effet de filtrage passe-bas.

Il n'est pas interdit de penser que la nature a bien résolu ce problème puisque, en général, aux voix les plus graves correspondent des cavités buccales de grand volume, donc des formes sémantiques relativement graves; mais ceci n'est aucunement une loi : une basse parlant en voix de fausset peut être parfaitement intelligible.

Le programme ANAMO permet de faire une expérience curieuse : l'anamorphose progressive d'une phrase, avec un coefficient Ca variant entre 0,7 et 1,4 (fig.9). L'auditeur éprouve une sensation très nette d'intonation (ici ascendante) alors que le spectre de raies reste rigoureusement fixe (Ci = 1). En réalité il s'agit seulement d'un changement de timbre de la voix; ce phénomène est analogue à l'intonation

...../

de la voix chuchotée, déjà remarquée (bib.4) mais difficile à mettre en évidence. On verra plus loin (fig.11) un exemple d'anamorphose en voix chuchotée, qui correspond également à une sensation d'intonation.

b) L'anamorphose dans les voix réelles.

On a regroupé sur la fig. 10 plusieurs sonagrammes de la même séquence parlée ("ASTUVUS"...), prononcée par des hommes et des femmes. Si l'on fait abstraction des différences de timbre, d'intonation et de rythme, il reste en commun la forme sémantique. L'anamorphose qui met en correspondance la forme sémantique de chacun avec une forme de référence (par exemple, celle isue de l'Icophone III) n'est pas aussi régulière que celles du paragraphe précédent, qui se ramenaient à un coefficient unique pour toutes les fréquences. Par ailleurs les formes sont floues; ce flou ne provient pas seulement de l'imperfection du sonagraphe, mais surtout de l'imperfection (toute relative ! ...) de nos organes phonatoires.

Notons en passant que la notion de "forme floue", qui nous paraît essentielle dans la structure acoustique de la parole, est à opposer à la notion d'"empreintes vocales" : l'exploitation commerciale qui en est faite est basée sur un parallélisme purement verbal avec la notion d'empreinte digitale. La fig. 11 vient à l'appui de cette remarque : le même mot, prononcé par le même locuteur (ici en voix chuchotée) avec deux nuances différentes, donne lieu à deux sonagrammes passablement différents, qui ne correspondent pas l'un à l'autre dans une anamorphose fréquentielle simple.

Les considérations précédentes posent en réalité le problème de la reconnaissance automatique de la parole, dont nous allons maintenant examiner la formulation selon les vues du Laboratoire.

VI - LA RECONNAISSANCE AUTOMATIQUE

Il existe depuis vingt ans des systèmes capables de reconnaître automatiquement certains sons de la parole. Les sons les plus simples, c'est-à-dire les voyelles et les consonnes stables, sont en général reconnus sans trop de difficultés dans une parole articulée très lentement, ceci pour un nombre réduit de locuteurs. Le problème devient beaucoup plus difficile lorsqu'il s'agit de faire reconnaître à une machine la parole "fluide", c'est-à-dire continue ("connected speech"), comprenant tous les sons de la langue. Certains systèmes contournent la difficulté en limitant au départ le nombre de mots ou expressions utilisables par le locuteur, et en isolant ces mots ou expressions les uns des autres. Ce qui, par certains côtés, est assez proche de la perception humaine qui appréhende globalement la forme sémantique; cependant ces systèmes nécessiteraient le stockage d'un nombre infini de mots ou expressions pour être véritablement utilisables. Par ailleurs aucun système, à notre connaissance, ne fonctionne correctement pour un grand nombre de locuteurs, hommes, femmes ou enfants.

Le véritable problème est donc, à notre sens, de concevoir un système capable de transcrire phonétiquement la parole de n'importe quel locuteur, sans limitation sur le vocabulaire ou la vitesse d'élocution, à ceci près que, dans une première étape, tous les sons à reconnaître devront être présents dans le message. On ne demandera pas au système, par exemple, de reconnaître " B A D A U D " quand le locuteur aura prononcé effectivement " B A T E A U ".

L'adaptation au locuteur se fera d'après la prononciation d'un mot-clé ou d'une phrase conventionnelle, comparable au "Allo" téléphonique. D'après l'analyse de ce mot, la machine fera elle même les anamorphoses et pondérations ramenant la voix de ce locuteur particulier dans le cadre général.

...../

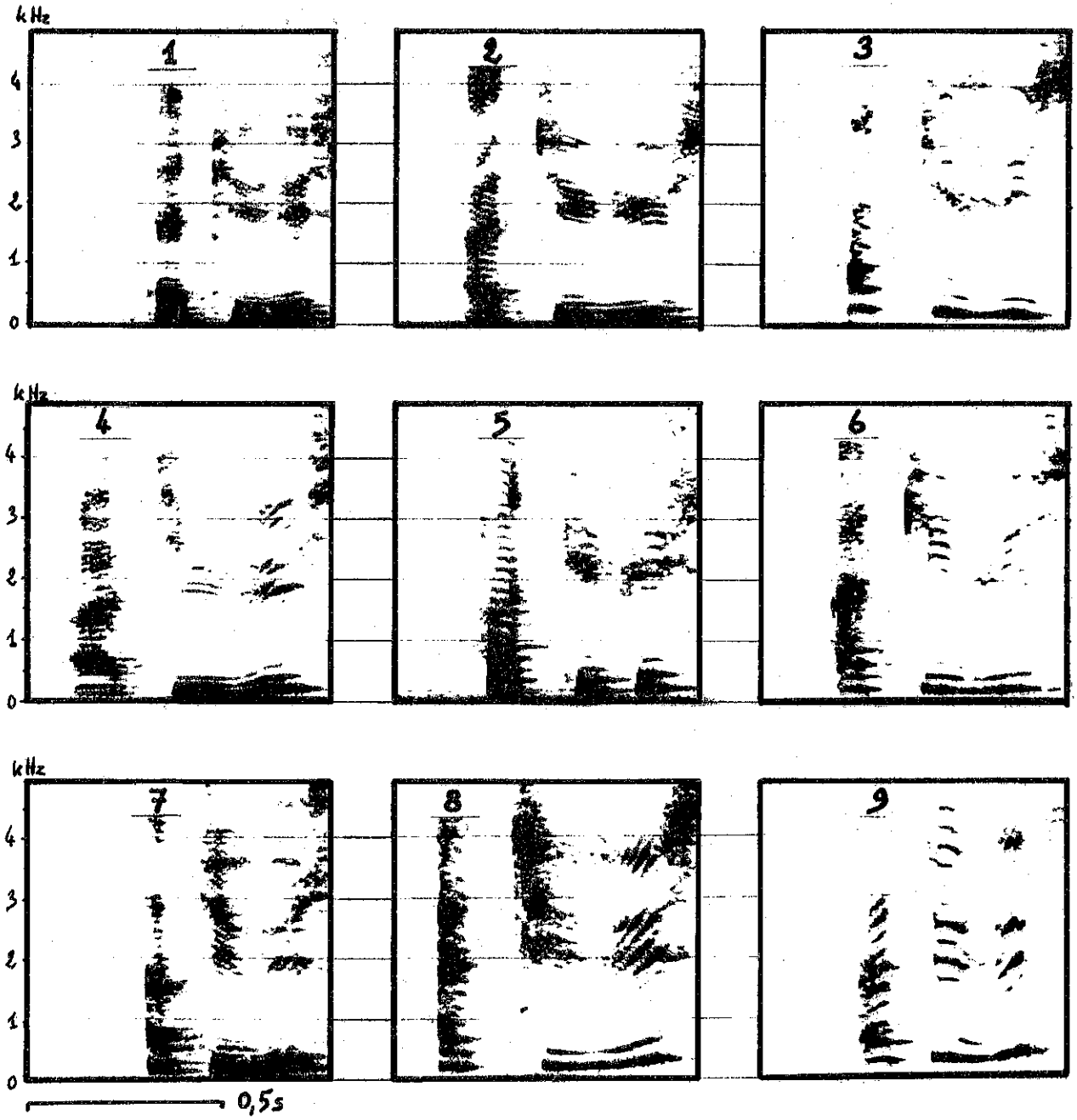


Fig 10 - L'anamorphose dans les voix réelles. La même séquence, "ATUVUS..", est prononcée par 9 locuteurs différents. Les sonagrammes n° 1, 2, 4, 7, sont relatifs à des voix d'hommes, les autres à des voix de femmes.

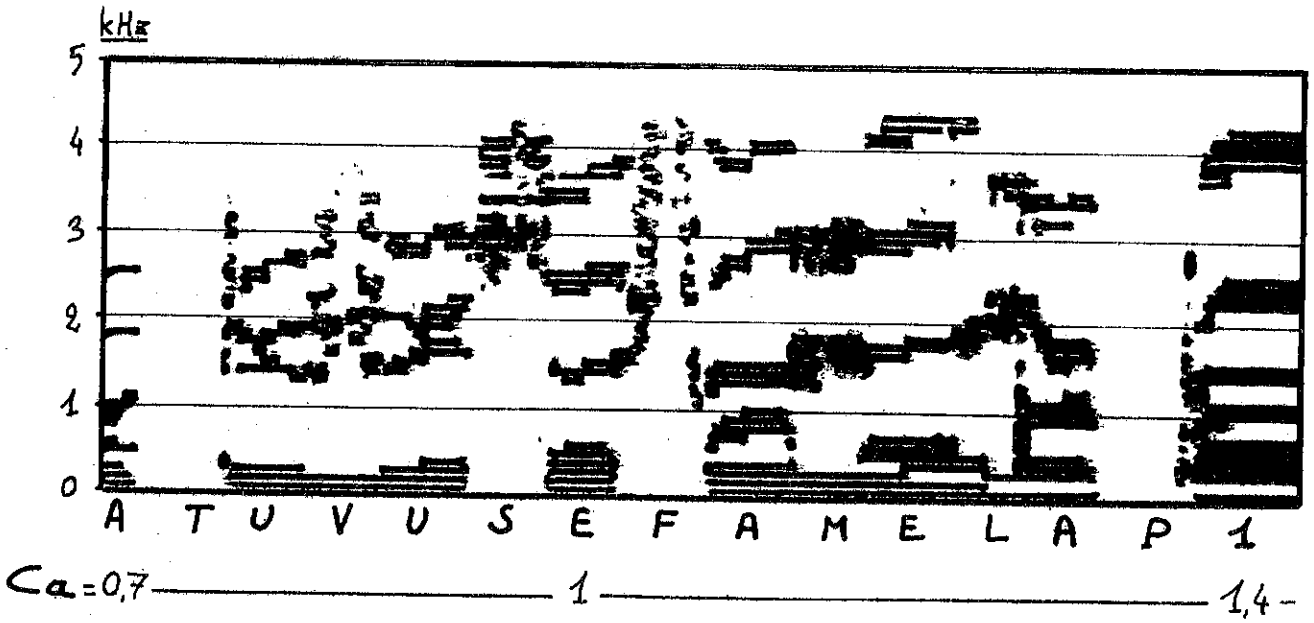


Fig 9 - Anamorphose progressive de la phrase synthétique :"As-tu vu ce fameux lapin". Le réseau de raies reste fixe en fréquence (Ci = 1) ; les coefficients d'anamorphose (Ca et Cr sont égaux dans ce cas) évoluent de 0,7 à 1,4 par bonds de 0,05. Il en résulte une très nette sensation d'intonation ascendante.

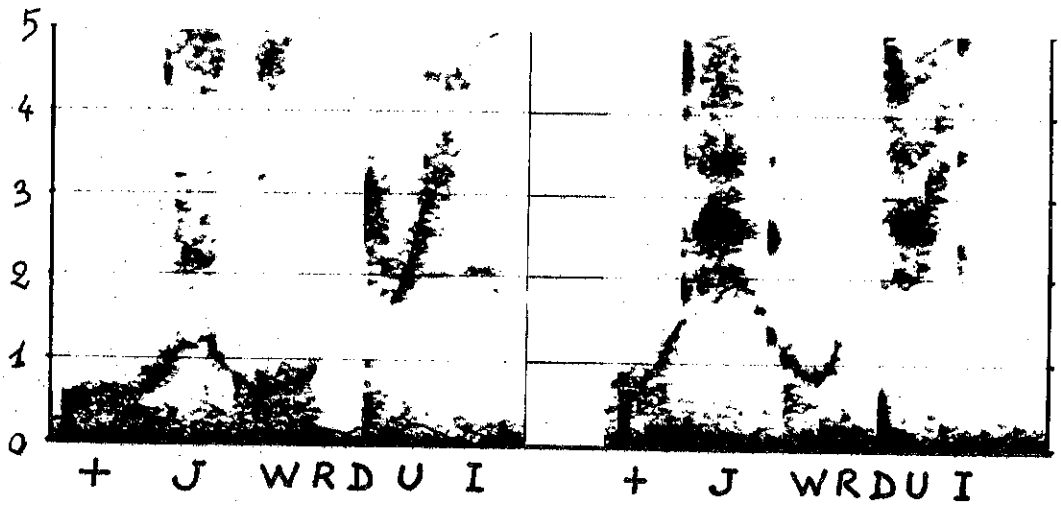


Fig 11 - Anamorphose fréquentielle d'une voix réelle. Le mot "Aujourd'hui" est ici prononcé par le même locuteur, en voix chuchotée, avec deux timbres de voix différents. Les formes sémantiques sont floues, l'anamorphose n'est pas régulière.

Posé de cette manière, le problème nous semble la suite logique de la synthèse : celle-ci nous a appris à isoler les éléments sémantiques de la parole, c'est-à-dire les phonatomes; nous en avons constitué un dictionnaire. Il s'agit maintenant de les reconnaître. On peut se demander si les phonatomes sont bien des signes parfaitement distincts les uns des autres : dans le cas contraire il serait inutile de chercher à les reconnaître. Pour répondre à cette question nous avons comparé certains phonatomes typiques du dictionnaire à tous les autres. Ces comparaisons, effectuées au moyen de l'ordinateur, avec un critère s'apparentant à la corrélation, ont donné des courbes du genre de celle de la figure 12. On peut en déduire que les phonatomes étudiés peuvent être distingués de leur plus proches voisins avec une bonne marge de sécurité. Naturellement, cette étude reste à faire pour des phonatomes issus de voix réelles.

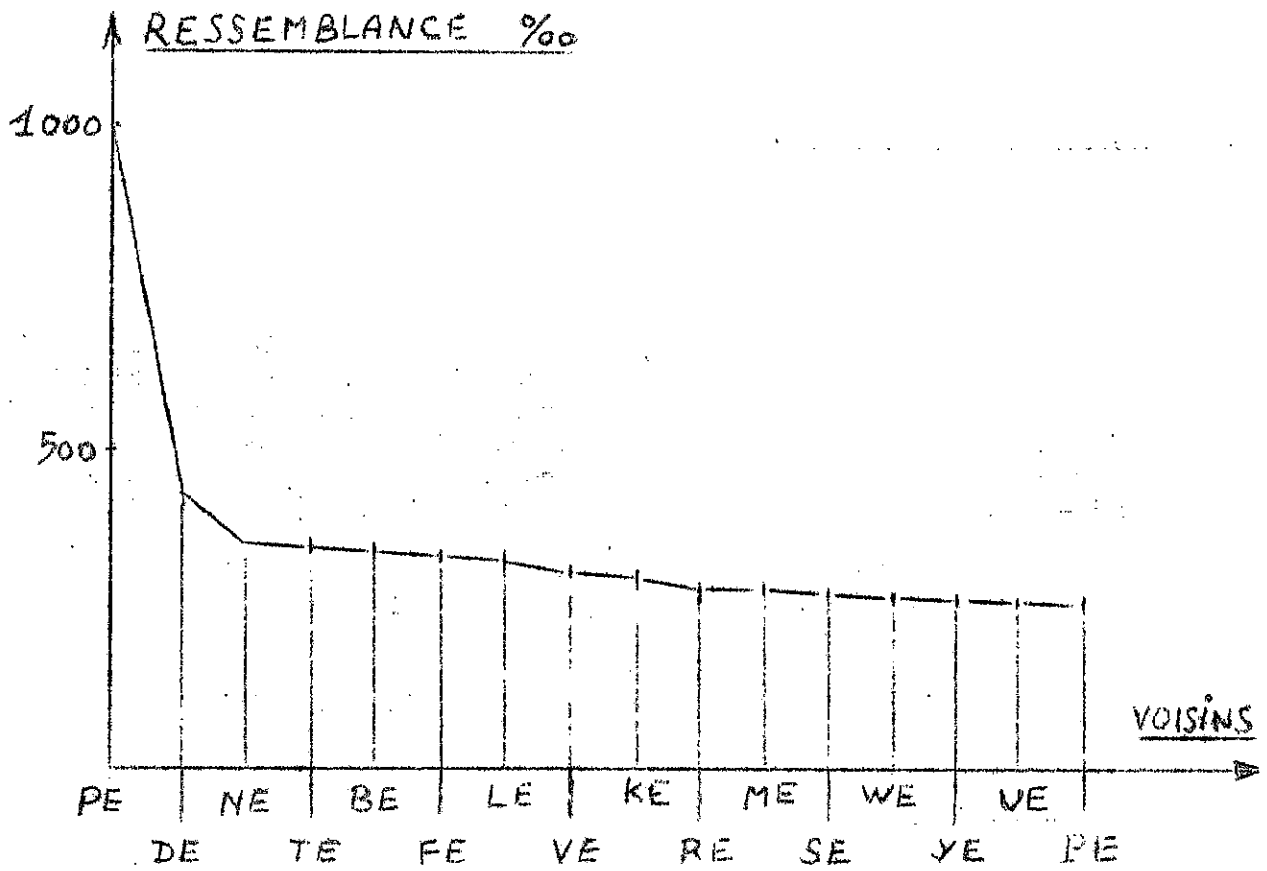


Figure 12

La toute première phase de la recherche, qui consistait à améliorer le dictionnaire, touche maintenant à sa fin. Elle a donné lieu à un contrat avec la Direction des Recherches et Moyens d'Essai, qui portait également sur le projet d'un analyseur adaptable à divers types de voix.

La seconde phase consiste à simuler sur ordinateur un processus de reconnaissance des phonatomes, sans perdre de vue les objectifs signalés plus haut. Après avoir résolu

...../

quelques difficultés d'ordre technique nous sommes maintenant capables d'effectuer en machine (IBM 360 du CIRCE) une analyse de la parole comparable à celle que nous donne le sonagraphe, ceci pour environ 30 minutes de parole enregistrée par des locuteurs variés (voir l'exposé de M. MLOUKA). Cette simulation a retenu l'attention du Comité de Recherche en Informatique, avec lequel un contrat est en cours de rédaction.

Parallèlement, nous envisageons la réalisation de l'analyseur et son raccordement à l'ordinateur du C.C.A. ; cette manipulation semble particulièrement prometteuse, car elle permettra non seulement d'introduire de la parole en machine sans délai, mais aussi d'utiliser l'Icophone et la visualisation comme moyens de contrôle de la reconnaissance. L'ensemble constituera un outil universel d'étude de la parole et des systèmes de transmission.

VII - LES MOTS ALEATOIRES

L'idée d'utiliser statistiquement la distribution des phonatomes dans le langage s'est fait jour dès que le Laboratoire s'est intéressé aux problèmes de la parole (voir l'exposé de E. LEIPP). On a imaginé plus récemment d'utiliser ces statistiques pour recréer des mots sans signification, nommés mots aléatoires, respectant les normes phonétiques du français. Nous allons examiner sommairement ces deux points.

1°) Distribution des phonatomes.

Nous ne reviendrons pas ici sur le détail des comptages, exposé par ailleurs (bibl. 3 et 4). Signalons seulement quelques conséquences de ce type de recherche :

a) en phonétique.

L'existence d'une norme phonétique du français a été clairement mise en évidence : certains phonatomes sont plus utilisés que d'autres, et les probabilités d'apparition des phonatomes ne peuvent pas être reliées simplement aux probabilités d'apparition des phonèmes. Une étude phonologique approfondie permettrait sans doute d'expliquer certains faits. Pourquoi, par exemple, le phonatome "TR" apparait-il 2,5 fois plus souvent que le phonatome " RT " (probabilités 0,010 et 004); de même pour " AR " (probabilité 0,012) et " RA " (probabilité 0,006). Pourquoi le phonatome " DE " est-il le plus courant (probabilité 0,018), et pourquoi le phonatome " VON " est-il très rare (probabilité inférieure à 0,001) ?

En bref il semble que les phonatomes soient des unités phonétiques significatives et que leur distribution caractérise les contraintes phonétiques de la langue.

b) en synthèse.

On a vu plus haut deux utilisations très concrètes des comptages : d'une part certains phonatomes pratiquement inutilisés dans la langue n'ont pas été schématisés dans le dictionnaire, qui fonctionne avec 600 phonatomes au lieu de 900 possibles théoriquement; d'autre part le classement par fréquence d'occurence décroissante a permis de réduire le temps moyen de recherche sur disque.

c) en reconnaissance.

Il ne faut pas s'attendre à ce qu'un système de reconnaissance fournisse un résultat sans équivoque sur chaque phonatome. Dans certains cas ambigus, le système devra choisir entre deux phonatomes donnant lieu au même taux de reconnaissance; on pourra alors utiliser les statistiques, c'est-à-dire utiliser le fait que le locuteur parle en français et par là subit les contraintes phonétiques de la langue. Mais l'amélioration à attendre de cette intervention ne peut être que statistique : il ne faudra l'utiliser qu'en dernier ressort car elle n'ajoute rien

...../

au sens du message à reconnaître.

2°) Fabrication des mots aléatoires.

Il est extrêmement difficile de juger objectivement l'intelligibilité de la parole en synthèse. L'opérateur s'adapte à la voix synthétique, de si mauvaise qualité soit-elle. Il faut donc recourir à des tests objectifs de compréhension avec des auditeurs non conditionnés (voir l'exposé de Melle CASTELLENGO). Mais là encore le choix des mots synthétisés est déterminant : si l'on choisit des listes de mots simples (la maison, le sapin, etc...) on ne teste pas réellement les éléments synthétiques mais plutôt la forme sémantique globale du mot, laquelle peut être passablement déformée et encore reconnaissable. On teste aussi le niveau des préoccupations de l'auditeur. Si celui-ci a faim, par exemple, il aura tendance à comprendre " POULET " au lieu de " BALAI ", et la confusion peut-être justifiée de B et P entraînera une confusion incompréhensible entre A et OU.

Les téléphonistes ont depuis longtemps mis au point des listes de mots sans signification appelés logatomes. Malheureusement les listes normalisées sont en espérant afin d'avoir une valeur internationale; cela n'avance guère notre problème : comment tester une voix synthétique française avec les sons d'une autre langue ?

Les mots créés de toutes pièces par certains poètes, comme Rabelais ou Michaux, sont de ce point de vue bien plus satisfaisants, car ils " sonnent français ". Ils sont néanmoins en nombre limité et se rattachent souvent à des racines connues.

La solution consiste à fabriquer des mots de manière aléatoire, en imposant cependant les contraintes phonétiques dégagées par la statistique précédente. On aboutit ainsi aux mots suivants, retranscrits dans une " pseudo-orthographe " française :

suçin
treute
maikan
jouzon
naleu, etc..

De temps en temps apparaissent des mots français, comme :

compas
fondu
taré
phonème, etc...

Certains mots, néanmoins, ne ressemblent pas du tout à du français :

féeuèni
neleursse
ouapapu
oeprka
uneksau
nointèa, etc..

Nous avons cherché à réduire la proportion des mots entrant dans cette dernière catégorie. Pour cela des comptages distincts ont été effectués sur les phonatomes du début, du milieu et de la fin des mots, ainsi que sur les phonatomes de liaison. La création des mots aléatoires se fait en choisissant au hasard des phonatomes dans les tableaux appropriés (programme MOAL de H. LUCOT). Les mots obtenus de cette manière ont en général une sonorité agréable, bien qu'insolite; en voici quelques exemples :

...../

larare
bousque
omance
massire
vesti
daspan
raitoi
filure
fonelle
baipil
nanchor etc...

L'ordinateur peut fournir quelques centaines de mots en une seconde de travail. Il y a là, sans nul doute, un immense réservoir de mots, qui devrait intéresser les poètes et les linguistes autant que les utilisateurs involontaires du sabir atlantique...

Un aspect particulièrement prometteur de cette recherche se trouve dans la possibilité d'étudier expérimentalement les contraintes phonétiques d'une langue : on peut introduire des contraintes à priori, au moment du choix aléatoire (les tableaux eux-mêmes représentent certaines de ces contraintes), et prendre comme indice de validité la proportion de mots français trouvés dans la liste-résultat. Ainsi la différentiation des tableaux de répartition selon le début, le corps ou la fin des mots a fourni une proportion de mots français bien plus grande que lors du premier essai, effectué avec une statistique globale : il est donc raisonnable d'admettre que les sons du début, du corps et de la fin des mots n'obéissent pas exactement aux mêmes lois statistiques. Ceci conduit à réfléchir à l'"environnement phonétique" d'un mot : pour résister au cours du temps, un mot doit pouvoir être prononcé facilement "en continu", quels que soient le précédent et le suivant, et cette contrainte de la langue est nécessairement inscrite dans les différences entre les tableaux précédents.

Ce n'est là qu'un exemple parmi d'autres : on pourrait aussi interdire les suites de trois consonnes dans le corps d'un mot, ou n'autoriser que certaines d'entre elles (STR, p. ex.); on pourrait encore modifier arbitrairement certaines valeurs des tableaux etc...

Nous avons utilisé des mots aléatoires au cours de tests d'intelligibilité sur la voix de l'Icophone. Il existe une bonne corrélation entre le nombre de phonèmes (ou de phonatomes) correctement notés par les auditeurs (sur une liste de 20 mots aléatoires) et le nombre de mots bien compris (sur une liste de 80 mots courants, normalisés par le CNET). Une étude est en cours sur cette question, mais il semble d'ores et déjà que les mots aléatoires soient susceptibles de remplacer avantageusement les logatomes; ils requièrent en outre des auditeurs moins qualifiés, et en plus petit nombre.

VIII - CONCLUSION

Sous sa forme actuelle, l'Icophone est avant tout un outil de recherche sur la parole. Grâce à lui, nous avons pu établir un répertoire des formes sémantiques élémentaires, et montrer que l'anamorphose laissait invariant dans une large mesure le sens du message parlé.

C'est à l'ordinateur que l'Icophone doit sa puissance et sa commodité d'emploi. Dans le domaine de l'informatique, l'Icophone numérique constitue un moyen efficace et universel de donner la parole à l'ordinateur. Il est d'ailleurs vraisemblable que la parole synthétique, toujours semblable à elle-même, deviendra rapidement plus intelligible que la parole humaine (A. MOLES), de même qu'un texte imprimé est plus lisible qu'un texte manuscrit car les caractères en sont normalisés.

D'autres perspectives sont ouvertes en phonétique. L'ordinateur, capable de compter des mots et des sons, et de dégager des lois statistiques, est également capable d'"imaginer" des mots et des textes dans lesquels le hasard est contrôlé, et dosé par l'expérimentateur. Les applications de ce processus ne manquent pas, et vont des tests d'intelligibilité à la poésie futuriste, en passant par l'élaboration d'un modèle phonétique des langues

...../

La reconnaissance automatique de la parole est un problème plus difficile que la synthèse, à cause de la grande variété des locuteurs et de la méconnaissance que nous avons de certains mécanismes perceptifs. Néanmoins, les mêmes idées de base doivent nous permettre de reconnaître la parole d'un locuteur quelconque, dans la mesure où tous les sons auront été effectivement prononcés. Si un tel système arrive à fonctionner avec des performances suffisantes, on peut penser que la parole, par l'intermédiaire du téléphone, deviendra à plus ou moins brève échéance le mode de communications privilégié entre l'homme et l'ordinateur ainsi domestiqué.

o

o o

BIBLIOGRAPHIE

1. E. LEIPP, Mlle CASTELLENGO, J.S. LIENARD.
" Parole et Gestalt-Theorie "
Colloque sur la parole organisé par le CNET à LANNION, juin 1966.
Ed. int. Fac. Sciences - PARIS.

2. E. LEIPP - Information sémantique et parole.
Bulletin du GAM N° 22, Juin 1966.

3. J.S. LIENARD - Le dictionnaire des éléments phonétiques et ses applications.
Bulletin du GAM, n° 22 bis, juin 1966.

4. E. LEIPP, Mlle CASTELLENGO, J.S. LIENARD, J. SAPALY
" Structure physique et contenu sémantique de la parole ".
Colloque sur la parole organisé par le GALF, Avril 1967.
Revue d'Acoustique N° 3 - 4, décembre 1968.

5. E. LEIPP - L'intelligibilité de la parole.
Bulletin du GAM N° 37, Novembre 1968. Publié également dans la revue d'Acoustique n° 12, 3° année, 1970, pp 343-360.

6. J.S. LIENARD - La synthèse de la parole, historique et réalisations actuelles.
Nucléus, t. 10, n° 3, 1969.
Paru également dans la Revue d'Acoustique, vol 3, 1970, pp. 204 - 213.

7. D. TEIL - Etude de génération synthétique de parole.
Thèse C.N.A.M., Paris, 1969.

8. J. QUINIO, D. TEIL - La synthèse de la parole par ordinateur à partir de digrammes phonétiques - Revue d'Acoustique, t. 3, n° 9, 1970.

9. J.S. LIENARD, D. TEIL - Les éléments phonétiques et la traduction automatique du message écrit en message parlé.
Automatisme n° 10, octobre 1970.

10. E. LEIPP, Mlle CASTELLENGO, J.S. LIENARD, J. QUINIO, D. TEIL.
Générateur synthétique de parole.
Brevet ANVAR n° 182 925, décembre 1968.

M. CASTELLENGO

ELABORATION DU DICTIONNAIRE DES PHONATOMES ET AMELIORATIONS APPORTEES A LA VOIX DE L'ICOPHONE

JANVIER 1971 N°=53

GAM

BULLETIN DU GROUPE d'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES . TOUR 66 . PLACE JUSSIEU . PARIS 5°

Mlle CASTELLENGO - GAM Janvier 1971

P L A N

1 - INTRODUCTION

2 - ELABORATION DU DICTIONNAIRE DES PHONATOMES

a) Synthèse globale

b) le premier dictionnaire en voix chuchotée

c) Normalisation des phonatomes en vue de l'entrée en ordinateur.

3 - AMELIORATION DU DICTIONNAIRE DES PHONATOMES

a) Corrections des formes sémantiques

b) Le deuxième dictionnaire de phonatomes : le voisement

c) Tests sur la voix de l'icophone et 3ème dictionnaire

4 - ESSAIS D'INTONATION

a) Hauteur harmonique et hauteur formantique

b) Premier essai d'intonation

c) L'intonation à l'Icophone IV

5 - AUTRES UTILISATIONS DE L'ICOPHONE IV

a) Le chant

b) La synthèse des chants d'oiseaux

6 - CONCLUSIONS

Mlle CASTELLENGO - GAM Janvier 1971

ELABORATION DU DICTIONNAIRE DES PHONATOMES
ET AMELIORATIONS APPORTEES A LA VOIX DE L'ICOPHONE

1 - INTRODUCTION

Comme M. LEIPP l'a exposé dans cette même étude, nos premiers essais en synthèse de parole avaient pour objectif de produire une voix, qui tout en restant intelligible soit la plus économique possible. Nous sommes donc partis de la voix chuchotée filtrée, éliminant ainsi l'information esthétique relative aux qualités acoustiques de la voix (timbre, intentions émotives etc...), pour ne garder que les formes sémantiques essentielles. Celles-ci ne sont pas immédiatement lisibles sur le sonagramme. Il faut les en extraire par approximations successives, et adapter le dessin à l'appareil de synthèse.

L'Icophone I ne pouvait permettre qu'une justification de nos idées. La parole ainsi produite n'était intelligible qu'aux auditeurs fortement suggestionnés... et les manipulations beaucoup trop longues, en raison du seuil ultra critique de la cellule photoélectrique lisant par réflexion : une demi-journée de travail pour 2 secondes de parole ....

Avec l'Icophone II nous avons pu entreprendre une étude sérieuse de la parole synthétique.

2 - ELABORATION DU DICTIONNAIRE DES PHONATOMES.

a) Synthèse globale .

Il n'a pas été inutile au début, de recopier des sonagrammes, l'extraction des formes n'étant pas toujours évidente. Nous avons ainsi appris à exploiter l'appareil au mieux et surtout à manier le pinceau.

Sur la figure 1, on voit l'analyse au sonagramme de la phrase " les p'tits oiseaux chantent c'est l'printemps." en voix normale (a) puis en voix chuchotée (b). En (c) on a reproduit le dessin fourni à l'Icophone. On remarque le dessin hachuré qui permettait d'obtenir un bruit simulant la voix chuchotée.

On constate aussi une anamorphose assez importante entre l'original sonagraphique et le dessin. Prenons par exemple les bandes formantiques du "I", elles sont situées vers 2500, 3200 et 4500 Hz pour la voix féminine analysée. (Le formant grave est quasi-inexistant en voix chuchotée et nous l'avons supprimé). Or il a fallu tout "rentrer" dans la gamme de fréquences de l'appareil qui coupe à 4400 Hz, y compris les bruits des S et des Z dont le spectre s'étend bien au delà. La copie des sonagrammes comportait donc une transposition à vue que nous avons systématisée par la suite en prenant les zones formantiques du "E" comme référence. Cette façon de faire confirmait nos idées sur la parole considérée comme forme acoustique au sens de la Gestalttheorie.

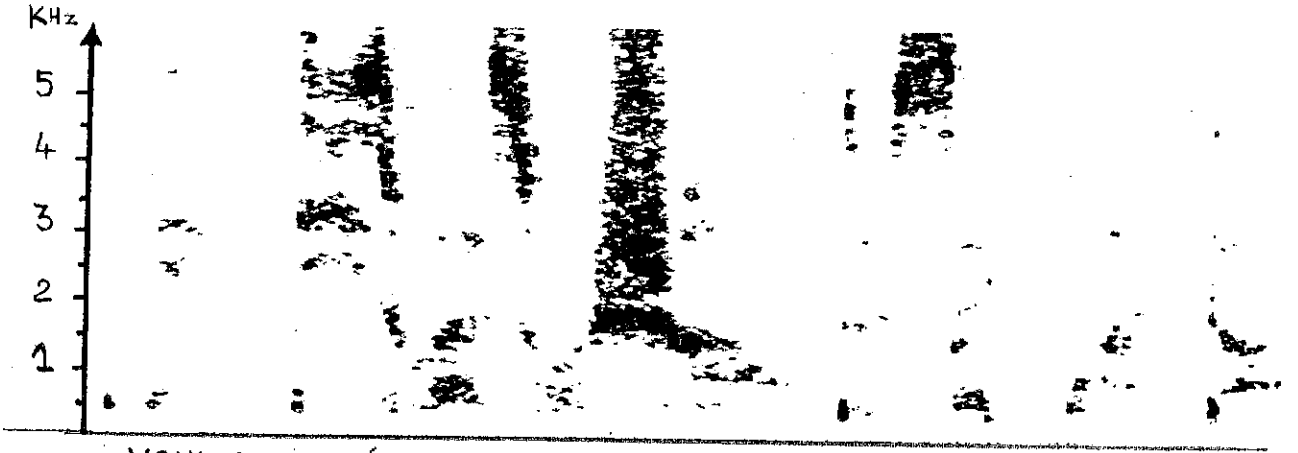
La parole ainsi obtenue était intelligible, seulement, à partir d'analyses en voix féminine on synthétisait une voix masculine. Le premier texte d'une certaine longueur (45 secondes) l'horrible histoire de la "malle sanglante de la Gare de Lyon", fait divers extrait du journal Le Monde d'Août 1967 a permis de faire quelques tests auditifs de plus longue durée et de mettre en évidence la part importante de l'habituation à ce timbre de voix si particulier.

Mais la recopie de sonagrammes n'est pas de la synthèse vraie, sans compter que l'analyse préalable prenait un temps non négligeable.

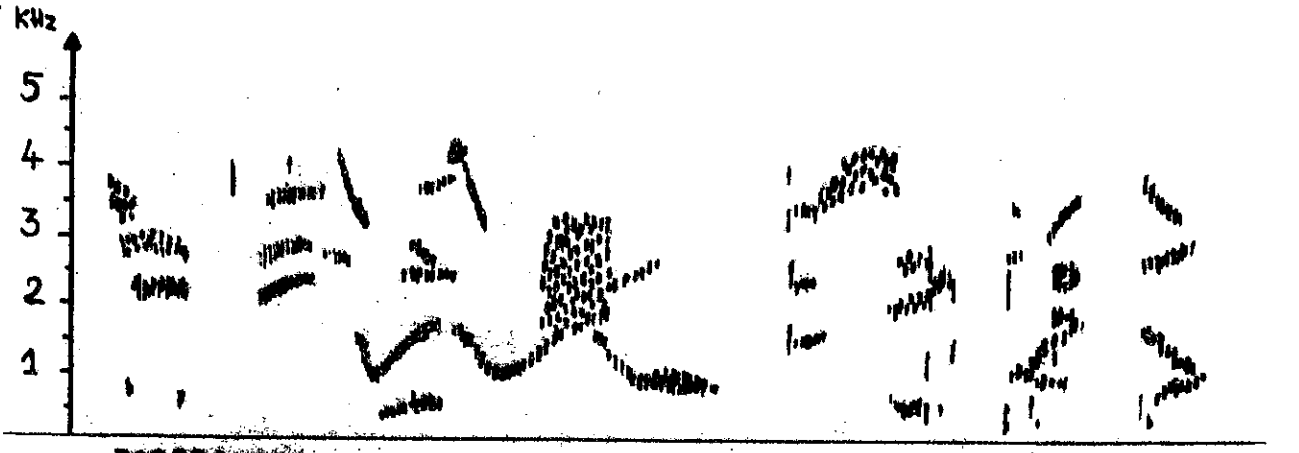
...../

M. CASTELLENGO - GAM Janvier 1971

VOIX NORMALE



VOIX CHUCHOTÉE



ICOGRAMME

LJ P'T I Z WA Z+ X * TE S C L' PR1 T*

Fig. 1

« les p'tits oiseaux chantent c'est l'printemps »

Sur la base des idées exposées au colloque de Lannion (Bib. 1) nous avons donc entrepris la schématisation d'un premier dictionnaire.

b) Le premier dictionnaire en voix chuchotée.

Conjointement avec la synthèse globale nous avions commencé l'étude d'un premier dictionnaire en voix chuchotée à partir duquel quelques mots français avaient été synthétisés et présentés au colloque de Grenoble : "ticket, bureau, saucisson etc... (Bib. 2)

Chacun de nous, c'est-à-dire, M. LEIPP, J.S. LIENARD et Mlle CASTELLENGO décida de faire son propre dictionnaire. Ultérieurement, nous ferons une comparaison des 3 réalisations.

Nous avons choisi 30 phonèmes dont les associations 2 à 2 sont les phonatomes, les éléments de transition entre 2 phonèmes. Il y a 12 voyelles :

A moyen, E moyen, I , Ô fermé, O ouvert,

U , OU , ON , AN , IN , É , È

6 consonnes plosives : P, B, T, D, K, Gu.

6 consonnes fricatives : S, CH, F, Z, J, V,

5 autres consonnes ou assimilées M, N, R, L, Y (le Y de fille) et

un E final.

Ce choix nous a paru satisfaisant pour synthétiser une parole courante sans ambiguité, la confusion des différents types de A, l'abandon du UN ne se sont pas révélés (gênants) pour l'intelligibilité. Des études préalables (Bib. 3) nous ont permis de connaître quels phonatomes étaient les plus fréquents dans le langage courant afin de les aborder en premier.

Le travail préparatoire consiste donc à tirer les analyses au sonagraphe des phonatomes que l'on veut étudier. Puis on les dessine sur la bande de mylar qui passera dans l'Icophone II. L'expérience nous a vite appris que la suggestion personnelle était notre plus grande ennemie. Pour rester un auditeur valable il faut user de véritables ruses à l'égard de soi-même, présenter les phonatomes en désordre, les repérer par des numéros arbitraires, et un mois plus tard on est bien surpris de ne pas reconnaître ce que l'on avait considéré comme excellent, ce qui explique souvent les réactions des visiteurs non prévenus.

Tout le travail de corrections se fait à l'oreille. Il est facile de modifier le dessin, de supprimer une ou plusieurs fréquences, en agissant sur le niveau de sortie des oscillateurs, mais on se fatigue vite d'écouter des phonatomes d'1/10ème de seconde : au bout de peu de temps on est incapable de juger. Il s'est avéré rapidement nécessaire de composer des textes pour tester les phonatomes, seulement, pour 1 minute de parole il fallait 3 jours de travail... L'ordinateur est heureusement intervenu pour nous relayer dans ce travail long et fastidieux.

c) Normalisation des phonatomes en vue de l'entrée en ordinateur.

Le phonatome est ramené à un cadre de 88 mm sur 20 mm pour une durée de 100 millisecondes. Comment opérer le découpage de façon que tous les phonatomes se raccordent ?

...../

M. CASTELLENGO - GAM janvier 1971

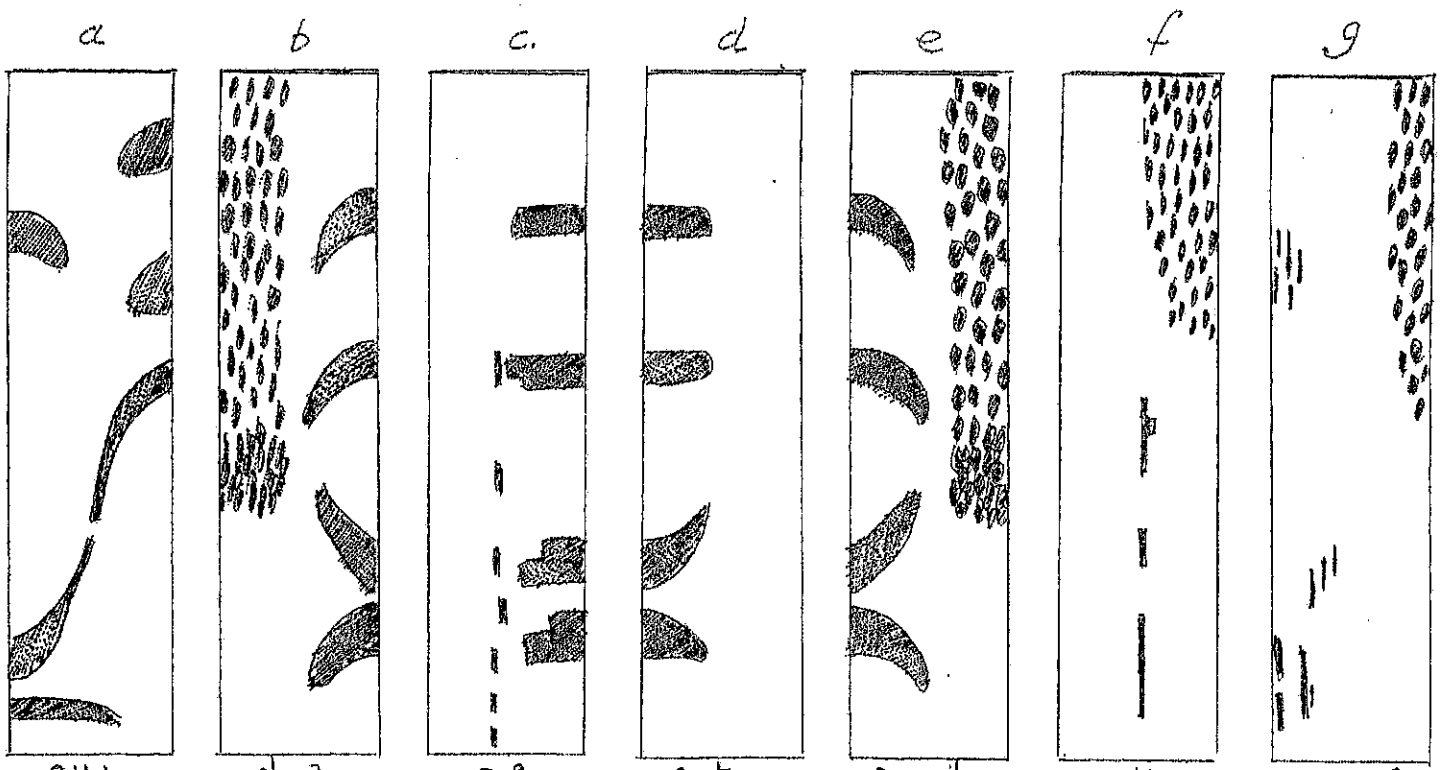
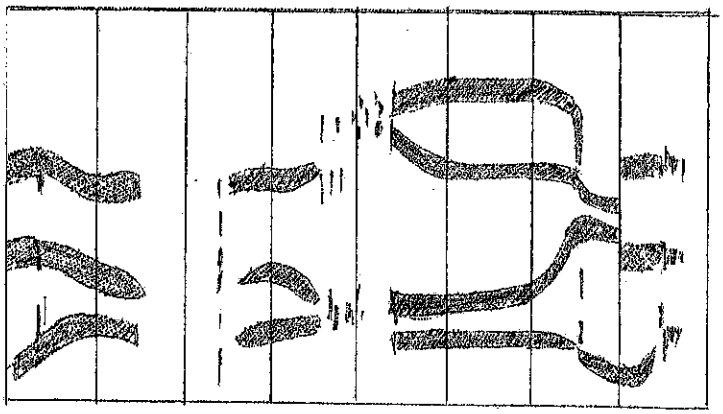
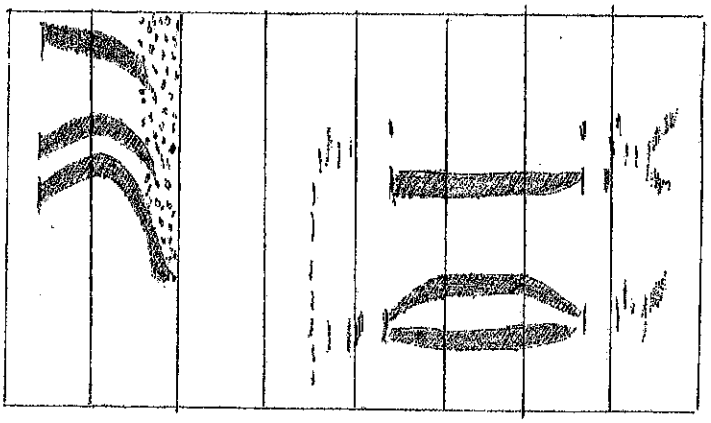


FIGURE 2 . *Exemples de phonatomes normalisés (code phonétique)*





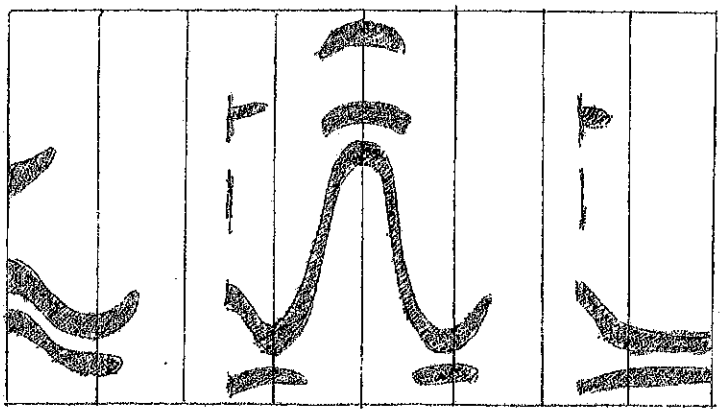


FIGURE 3

(Dictionnaire chuchoté)

- élément voyelle/voyelle - Le phonatome représente la transition entre la voyelle de départ et la voyelle d'arrivée. Ex. Fig. 2a OU/I ; les positions stables sont réduites au strict minimum et permettent le raccordement des phonatomes.

- élément consonne/voyelle - La première moitié du phonatome est réservée à la consonne et la deuxième aux transitions vers la voyelle d'arrivée. Ex. Fig. 2b CHA ; dans le cas des consonnes plosives on trouve d'abord le silence (50 ms), puis l'explosion qui dure de 2 à 15 ms, puis la transition vocalique Ex. PA fig. 2c.

- élément voyelle/consonne - Lorsque la consonne est une plosive on crée un phonatome comportant d'abord l'évolution de la voyelle avant l'occlusion, puis un silence d'environ 50 ms. Lors de l'enchaînement avec le phonatome suivant on reconstitue ainsi un silence de l'ordre de 100 ms. Ex. AT fig.2d. Dans tous les autres cas, on prend tout simplement l'élément consonne/voyelle correspondant que l'on lit à l'envers Ex. ACH Fig. 2e.

- élément consonne/consonne - Deux cas sont à considérer. Si la consonne finale n'est pas une plosive, l'élément est créé normalement. Ex. KS, Fig. 2f et RS , fig. 2g. Dans le cas contraire, le phonatome n'existe pas : il est remplacé par un demi-silence. Ainsi, pour synthétiser le mot ARDU on aura AR + 1/2 silence + DU . De même pour le mot OBTU on aura OB + 1/2 silence + TU.

  Il ne reste plus qu'à associer les phonatomes comme des "dominos" pour créer des textes quelconques, de durée illimitée, même en langue étrangère. Actuellement, pour la langue française, 625 éléments suffisent, dont 219 sont réversibles.

La fig. 3 montre un exemple de synthèse par phonatomes avec les mots

LA PAROLE

DIE SPRACHE

HOW DO YOU DO

Dès que le premier dictionnaire d'éléments phonétiques chuchotés a été introduit dans la mémoire de l'ordinateur, en Octobre 1968, nous avons pu composer toutes sortes de textes par simple frappe au pupitre de la machine : textes littéraires, scientifiques mots inventés de Rabelais. Un point important est immédiatement apparu : le rôle du rythme de la parole dans l'intelligibilité. Les phonatomes étant tous de durée identique, certaines séquences étaient incompréhensibles en raison de l'absence de rythme. Un programme simple a permis d'améliorer considérablement la qualité de la voix synthétisée. Il suffit de doubler la dernière voyelle rencontrée avant une respiration (virgule, point virgule ou point).

Tel quel, ce premier dictionnaire présentait encore bien des imperfections. Nous avons donc entrepris de l'améliorer. Ce travail, long et patient fut grandement facilité par les possibilités offertes par l'ordinateur.

3 - AMELIORATIONS DU DICTIONNAIRE DES PHONATOMES.

a) Corrections des formes sémantiques.

Dès AoÛt 1969 une unité de visualisation fut connectée à l'ordinateur. En affichant 3 phonatomes consécutifs sur l'écran on pouvait vérifier les raccordements, modifier le phonatome central si on le désirait et comparer à l'écoute ce même phonatome avant et après correction. Beaucoup d'éléments furent ain

...../

si retouchés par tatonnements, mais les résultats n'étaient pas encore satisfaisants.

Il était très difficile, malgré certaines distinctions de ne pas confondre les consonnes sourdes avec les sonores (S et Z, P et B etc...). D'autre part, bon nombre d'auditeurs reprochaient à la voix de l'Icophone son timbre particulier de voix chuchotée. Pour résoudre ces deux points il fut décidé de "voiser" les phonatomes.

b) Le voisement : deuxième dictionnaire de phonatomes.

Un premier essai sur l'Icophone II à commande optique, toujours en fonctionnement nous a permis de résoudre le problème simplement. Pour passer d'un phonatome chuchoté à un phonatome voisé on remplit plus ou moins le bas du dessin (fréquences de 100 à 500) pour les voyelles et les consonnes sonores.

Sur la figure 4 on peut comparer le même mot "LE BATEAU" dessiné à partir du dictionnaire chuchoté puis du dictionnaire voisé.

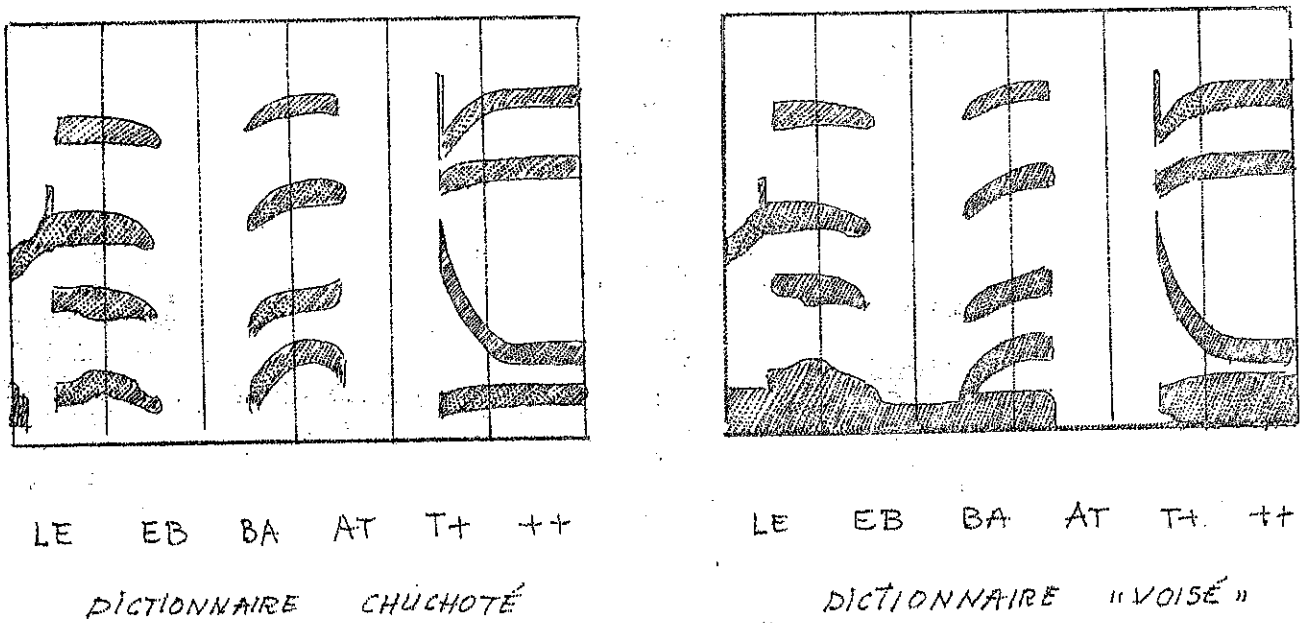


Figure 4

Il suffit maintenant d'accorder soigneusement les oscillateurs de l'Icophone. Toutes les fréquences étant harmoniques, la voix produite est monocorde, sur un fondamental de 100 Hz. Pour juger de la qualité de cette nouvelle voix et situer l'icophone parmi les autres réalisations en synthèse de parole, nous avons entrepris une série de tests avec des auditeurs "naifs".

c) Tests sur la voix de l'Icophone et 3ème dictionnaire.

Le choix s'est porté sur le personnel du laboratoire de Mécanique de Saint-Cyr l'Ecole. En dehors du fait que nous étions accueillis avec sympathie, nous

...../

trouvions sur place un auditoire assez varié : ouvriers mécaniciens, secrétaires, enseignants, étudiants etc... et suffisamment nombreux (28 personnes d'âge divers).

Le test consistait en dix listes de 100 mots français, masculins, disyllabiques,. Ces listes, utilisées en téléphonométrie, nous ont été aimablement communiquées par M. LORAND du C.N.E.T. Avant l'écoute du test, les auditeurs pouvaient entendre, dans le but de les habituer un peu à la voix synthétique, un court texte d'introduction (40 secondes) imprimé sur leur feuille de test.

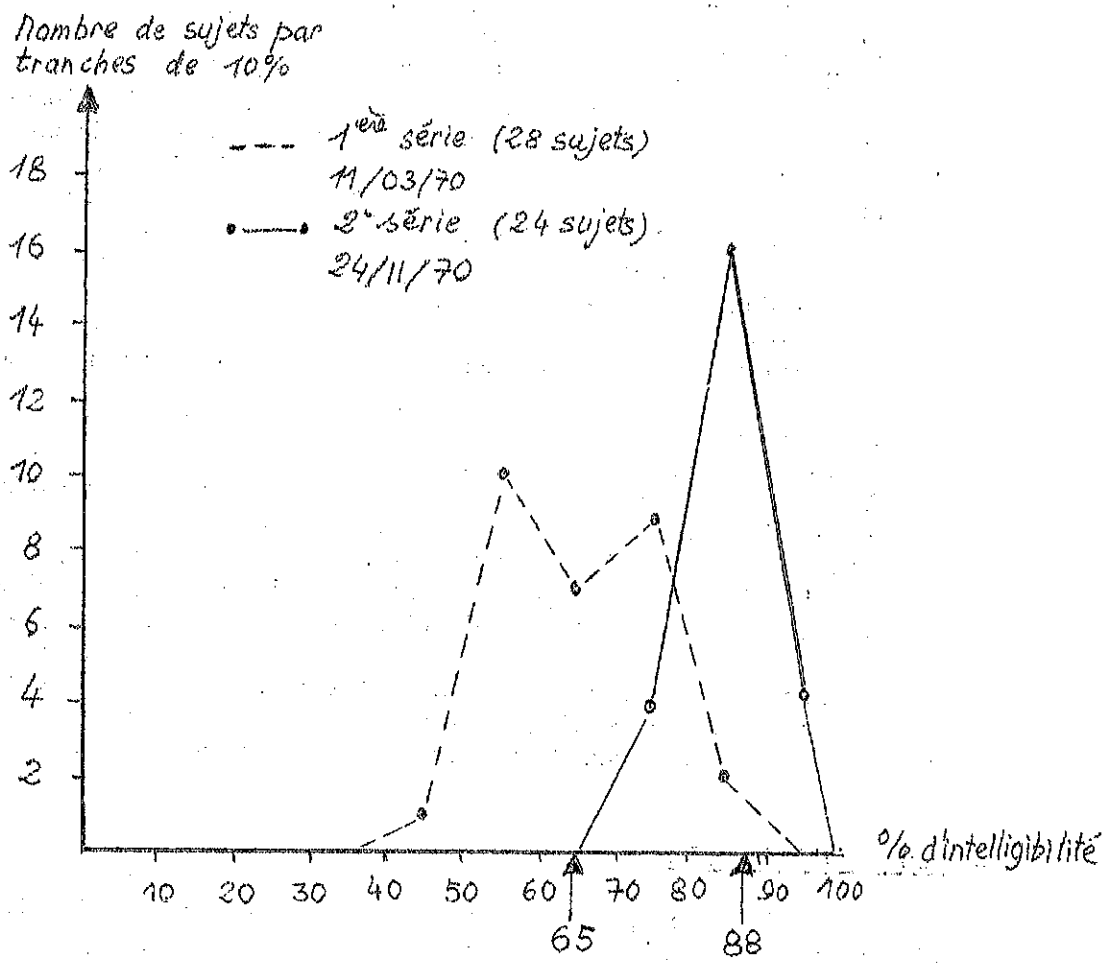


Figure 5

Les résultats (fig. 5a) s'ils n'étaient pas excellents nous ont beaucoup apporté pour la correction des phonatomes. Après un dépouillement soigneux et exhaustif nous avons constaté de nombreuses confusions entre les voyelles : nous étions nous-mêmes trop habitués à la voix de l'icophone pour nous en être rendu compte. Nous avons donc fait une étude comparative des rapports de fréquence des voyelles, chez divers locuteurs en vue de corriger la "grille des voyelles". Mais lorsqu'on change une voyelle, il faut modifier 36 éléments... et il y avait aussi beaucoup de formes à corriger, des éléments nouveaux à introduire. Pratiquement nous avons créé un 3ème dictionnaire de phonatomes qui est encore actuellement en mémoire. Ce fut un important travail, mais heureusement, nous avons bénéficié des importantes améliorations apportées au programme d'utilisation de la console de visualisation (cf exposé de M. CALINET, programme CREPH).

Huit mois plus tard, les résultats d'une deuxième série de tests toujours à Saint-Cyr, marquaient les progrès accomplis : voir Figure 5b. Les courbes comparatives portent sur les mots français. Nous avions aussi fait entendre des phrases courtes - toujours fournies par le CNET - pour lesquelles l'intelligibilité était quasi totale, et des mots "aléatoires". Ces mots, de 2 à 3 syllabes, créés par l'ordinateur sont dépourvus de sens, mais sonnent Français

...../

(cf exposé de J.S. LIENARD). On vérifie que les pourcentages de compréhension portant sur les phrases, les mots courants et les mots aléatoires sont en bonne corrélation avec ceux des courbes classiques admises en téléphonométrie (courbes de KRYTER, Bib. 4). On peut en conclure que l'intelligibilité de la voix synthétisée à l'icophone est, dès maintenant, au moins égale à celle du téléphone.

Pourtant, bien des points restent encore à améliorer : certaines confusions persistent entre L et N, M, R. Ces consonnes, particulièrement fluctuantes ne pourraient être bien définies qu'avec des "triphonèmes". Prenons par exemple l'élément RA. Il ne présente pas exactement la même forme sémantique dans l'enchaînement OURA et dans l'enchaînement IRA. Il faut donc trouver une forme moyenne qui convienne sans ambiguité, quel que soit l'enchaînement. Nous avons trouvé aussi quelques difficultés pour la synthèse du son GN (ex. Rognon) qu'il faudrait introduire comme une consonne distincte, et du son Y; dans ce dernier cas, nous touchons par surcroît aux limites de l'appareil dont la bande passante s'arrête à 4400 Hz. Il faut donc "suggérer" par le dessin ce qui se trouve au delà. Par la suite nous disposerons de trois plaquettes supplémentaires produisant des bandes de bruit réglables jusque vers 8000 Hz. Nous pourrons améliorer grandement la qualité des fricatives. Enfin, le fonctionnement des oscillateurs en tout ou rien limite une reproduction fidèle des voyelles nasales.

Mais notre propos n'était pas de faire une vois la plus fidèle possible. On pourrait y parvenir en améliorant les formes et en apportant des modifications à l'icophone. L'étude que nous poursuivons en collaboration avec Mme BOREL MAISONNY nous a permis de faire encore quelques progrès, mais ceux-ci sont de plus en plus difficiles à obtenir. Dans l'état actuel, il est plus économique de s'habituer à la façon de parler de l'icophone, qui d'ailleurs présente sur l'être humain un avantage de fidélité et de reproductibilité incontestable.

Il restait un dernier point à explorer. La voix monocorde est ennuyeuse et fatigante pour l'oreille; il serait plus agréable d'entendre une voix modulée en hauteur. Nous avons donc porté nos efforts dans ce sens.

4 - ESSAIS D'INTONATION.

A plusieurs reprises la remarque suivante nous a été faite sur la voix chuchotée synthétisée à l'icophone : " cette voix ne comporte pas de cordes vocales, pas de hauteur, et pourtant on entend une intonation". L'explication est bien simple.

a) Hauteur harmonique et hauteur formantique.

Soit un son riche en harmoniques de tous rangs. Sa hauteur est bien définie : c'est par exemple un sol2 de 200 Hz. A l'aide d'un résonateur variable, produisons un formant ascendant (Figure 6a). Nous entendons "quelque chose qui monte" alors que la hauteur proprement dite du son est restée fixe. On réalise facilement cette expérience avec une guimbarde. La hauteur fixe est fournie par la lame vibrante de l'instrument, le formant variable par la cavité buccale. Inversement, nous pouvons varier la hauteur du son harmonique et conserver un formant fixe (Figure 6b).

A l'écoute, selon que l'on porte son attention sur le son harmonique ou sur le formant, la sensation de hauteur peu différer largement (Bib. 5). Ces phénomènes, exploités en musique traditionnelle et surtout en musique expérimentale, vont nous permettre de comprendre ce qui se passe en parole.

Avec les cordes vocales, nous pouvons fabriquer un spectre de raies harmoniques de hauteur variable; c'est la hauteur de la voix en chant, l'intonation en parole (Bib. 6).

...../

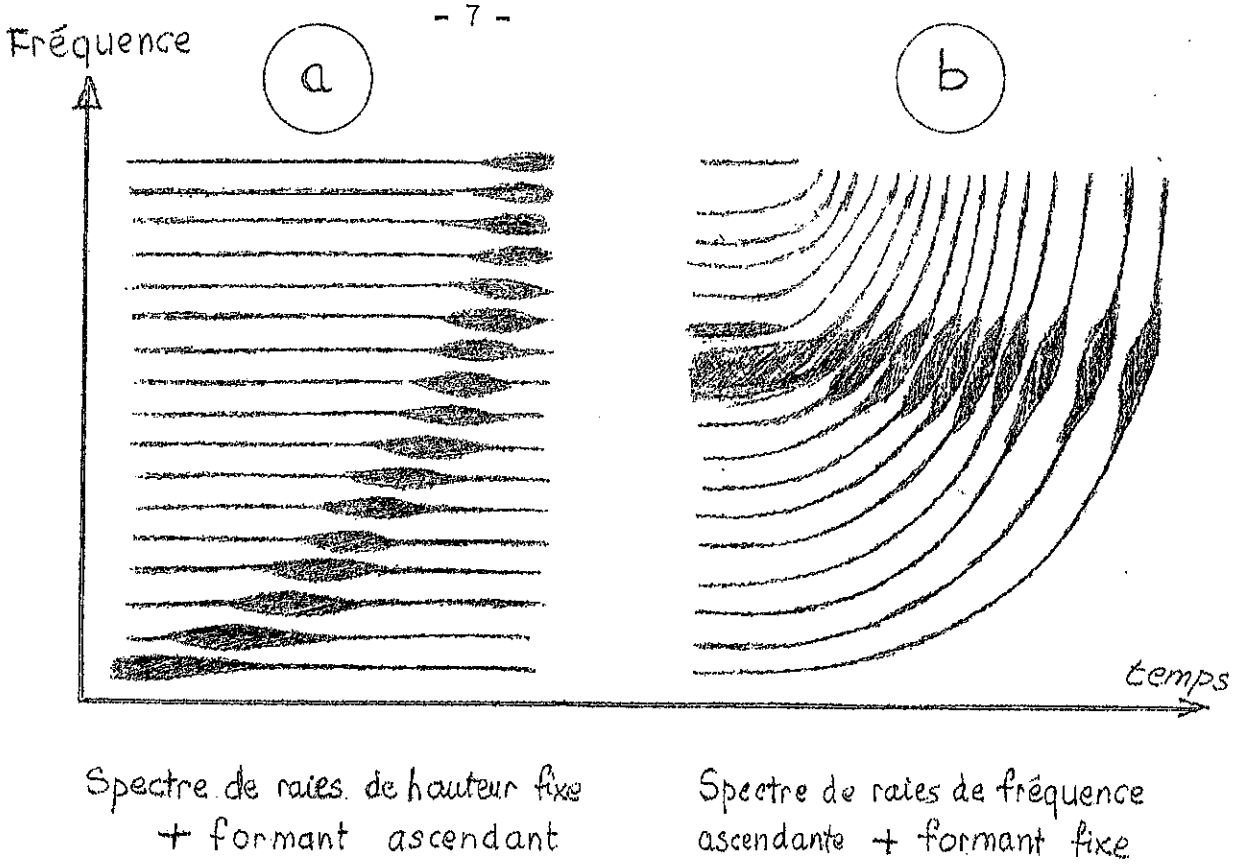


Figure 6

En réglant les cavités de l'appareil phonatoire nous produisons des zones formantiques plus ou moins compliquées; aux formes ainsi produites nous avons convenu d'associer une signification en parole.

Hauteur du son harmonique laryngé et hauteur des formants évoluant indépendamment l'une de l'autre, nous pouvons produire les signaux acoustiques suivants (figure 7).

Un mot, par exemple "oui", en voix recto-tono, avec intonation ascendante, avec intonation descendante.

Dans le cas particulier de la voix chuchotée, le spectre de raies est remplacé par un bruit d'écoulement produit au niveau des cordes vocales. On peut tout de même entendre une pseudo-intonation en portant son attention sur l'évolution mélodique du 2ème formant relativement plus intense et placé dans la zone sensible de l'oreille. Pour le mot oui, on entendrait toujours une intonation ascendante.

A l'Icophone III les raies harmoniques dessinent les formes sémantiques, elles en sont solidaires : il se pose donc un problème important, celui de l'anamorphose en fréquence des formes sémantiques. En effet, si nous voulons moduler les fréquences des 44 oscillateurs, les formes seront déplacées en conséquence.

...../

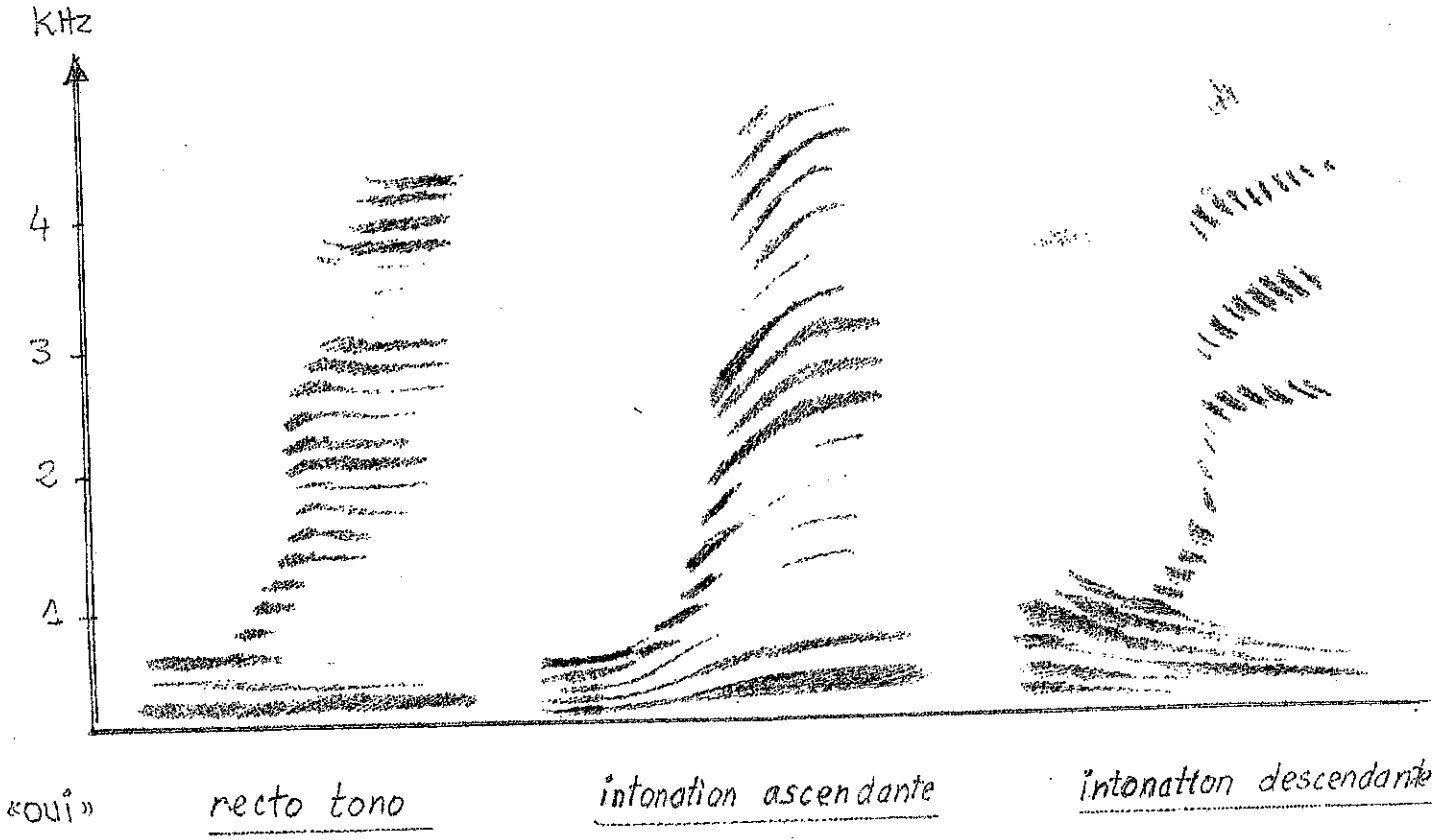


Figure 7

Prenons par exemple le mot "demain" (dem1) synthétisé à l'icophone monocorde (Figure 8a)

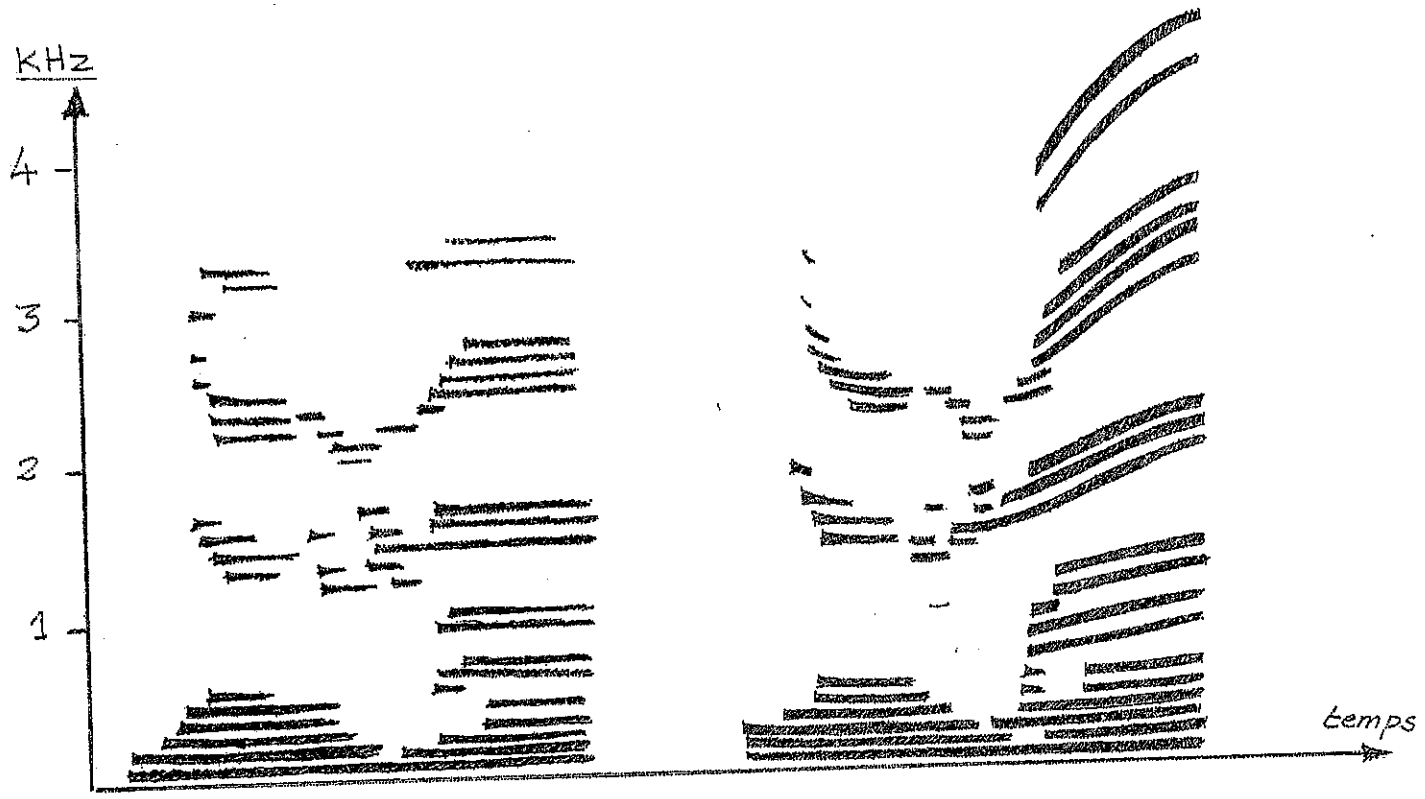


Figure 8

...../

Avec une intonation ascendante, les formes sont déformées, amplifiées. Mais selon les données de la Gestalttheorie, la reconnaissance ne doit pas en être affectée, si on ne dépasse pas certaines limites qu'il restait à définir. L'expérience était intéressante à tenter.

b) Premier essai d'intonation.

Lorsque tous les oscillateurs sont accordés, on entend une hauteur unique de 100 Hz (à peu près Sol 1). Le premier essai consistait à enregistrer une phrase synthétisée à l'icophone, et à la relire sur un magnétophone dont la vitesse variable pouvait être réglée manuellement. On entendait très nettement une intonation mais la voix ainsi produite était des plus curieuse. En effet, lorsque la vitesse de lecture du magnétophone ralenti, la fréquence baisse, mais simultanément la durée est allongée. De plus, les variations de hauteur sont anormalement lentes du fait de l'inertie des systèmes mécaniques. Par contre, nous avons pu vérifier que les anamorphoses accompagnant inévitablement les changements de hauteur n'étaient pas gênantes. Elles donnaient au plus un accent particulier à la voix. Les résultats encourageants nous ont donc convaincus de construire un Icophone IV avec lequel nous pourrions étudier l'intonation.

c) L'intonation à l'Icophone IV.

Dans cet appareil, (cf exposé de M. SAPALY), la fréquence centrale de chaque oscillateur peut être multipliée par un facteur quelconque. Comme il n'y a aucune inertie, on peut produire très rapidement de grandes variations de hauteur (2 octaves). La commande de la hauteur peut se faire manuellement au moyen d'un curseur (mais on retombe sur le problème de l'inertie) ou automatiquement à partir d'une courbe préalablement inscrite sur l'écran de visualition relié à l'ordinateur (programme ANAMO - cf M. CALINET). L'ensemble constitue un outil de choix pour l'étude de l'intonation et surtout de l'anamorphose de la parole (cf. exposé J.S. LIENARD).

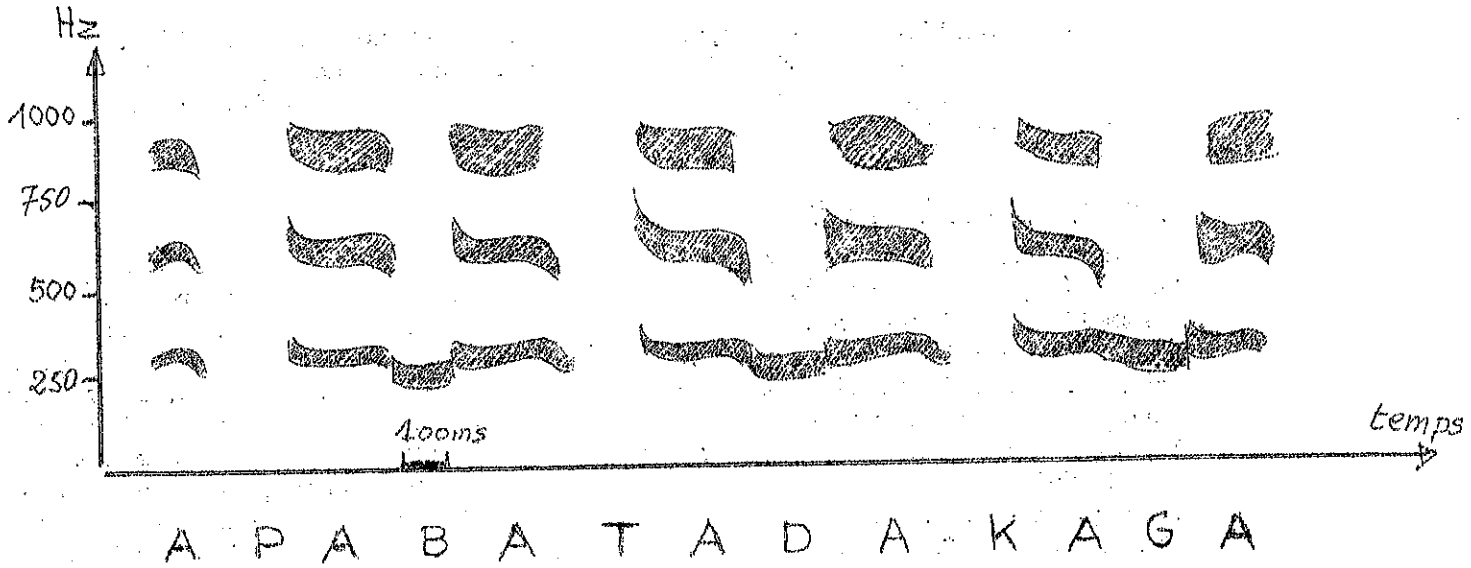
Les limites d'anamorphose acceptables sont variables selon les auditeurs et surtout selon le type d'expérience. Lorsqu'on se tient à un coefficient d'anamorphose donné on s'adapte très vite au type de voix ainsi produit. Mais dans le cas de l'anamorphose accompagnant l'intonation, les variations sont rapides et changent constamment de sens. Nous avons donc envisagé de faire une anamorphose automatique des formes, inverse de la variation de hauteur, ce qui revient à dissocier les formes sémantiques de la parole, du spectre de raies. Ainsi, lorsque la hauteur de la voix passe de 100 à 150 Hz, soit x 3/2, on réalise une anamorphose de 2/3. Cette opération se fait aisément à l'ordinateur, elle permet d'approcher de plus près la voix naturelle.

Dans le même but, nous commençons à introduire dans certains éléments phonétiques, des variations systématiques de l'intonation, qui apparaissent pendant la tenue des consonnes sonores et au moment de l'explosion des consonnes plosives. La figure 9 montre l'analyse au sonagraphe de quelques phonatomes prononcés à voix haute recto tono.

On voit que le mot "recto-tono" ne peut s'appliquer qu'aux parties tenues des voyelles. On observe de rapides variations de hauteur qui ne sont donc pas perçues en tant que telles, mais lorsqu'on peut les reproduire en synthèse, elles contribuent quelque peu à l'intelligibilité et surtout au naturel de la voix. Les cordes vocales sont assimilables à des anches doubles membraneuses (Bib.7). Au moment de l'occlusion d'un B par exemple, le débit aérien est brusquement stoppé par la fermeture des lèvres. La pression s'accroît légèrement dans la cavité buccale et le larynx, et entraîne une baisse de fréquence. Le phénomène inverse se produit à l'explosion.

L'étude de l'intonation à l'icophone est en cours, déjà les auditeurs du 30 Jan-

..../



On n'a représenté ici que les 3 premiers harmoniques de la voix (voix féminine, fondamental voisin de 280 Hz)
On lit très bien la ligne mélodique sur le fondamental ou sur l'harmonique 2, et on peut constater les très brèves variations de hauteur qui accompagnent l'émission des plosives.
L'ensemble est néanmoins perçu "recto-tono"

Figure 9

vier ont pu entendre une "idylle" entre 3 personnages de voix fort différentes. Les résultats ultérieurs donneront lieu à d'autres publications.

5. AUTRES UTILISATIONS DE L'ICOPHONE IV.

Les multiples possibilités de l'icophone IV couplé à l'ordinateur nous permettent d'aborder l'étude d'autres problèmes.

a) Le chant à l'Icophone.

Pour la réunion du 30 Janvier nous avons présenté "A la claire fontaine" chanté par l'icophone. En donnant aux variations de fréquence les rapports musicaux convenables, et en respectant le rythme de la chanson par l'allongement de la durée des voyelles on réalise une voix chantée tout à fait acceptable, mais bien mécanique !

Selon les suggestions de M. FÓNAGY il y aurait d'intéressantes études à faire sur les distinctions entre voix parlée et voix chantée; au moyen de la synthèse, on peut tester la valeur perceptive de brèves fluctuations de fréquence, systématiques ou non, à intervalle musical ou non; on peut étudier le rôle de l'anamorphose rythmique etc.... Les résultats apporteraient des données intéressantes en perception de la hauteur.

b) synthèse de chants d'oiseaux.

L'icophone est un synthétiseur sonore "universel". On peut faire de la parole, de la musique, mais aussi des bruits des cris d'animaux.

...../

En 1965, lors d'une réunion du GAM (Bib. 8) nous avions proposé une méthode de synthèse de chants d'oiseaux en utilisant un jazzo-flûte et des transpositions au magnétophone. Dès que nous avons disposé de l'icophone II, nous avons repris les essais à la demande de M. BRÉMOND de L'INRA, en particulier pour le chant de troglodyte. Pour couvrir la bande de fréquence nécessaire (au moins 8000 Hz) on synthétise à demi-vitesse et on pratique une transposition vers l'aigu, au magnétophone. Avec cette méthode de synthèse, il est facile de faire des opérations de retournement d'amputations, de transposition en fréquence sans changer le temps, enfin toutes sortes de dégradations du chant de l'oiseau, parfaitement contrôlées.

On peut aussi envisager un dictionnaire des "phonatomes" d'oiseaux. A titre d'exemple nous avons créé 20 éléments qui, combinés, répétés, permettent déjà de générer des formulettes intéressantes (cf. Fig. 10)

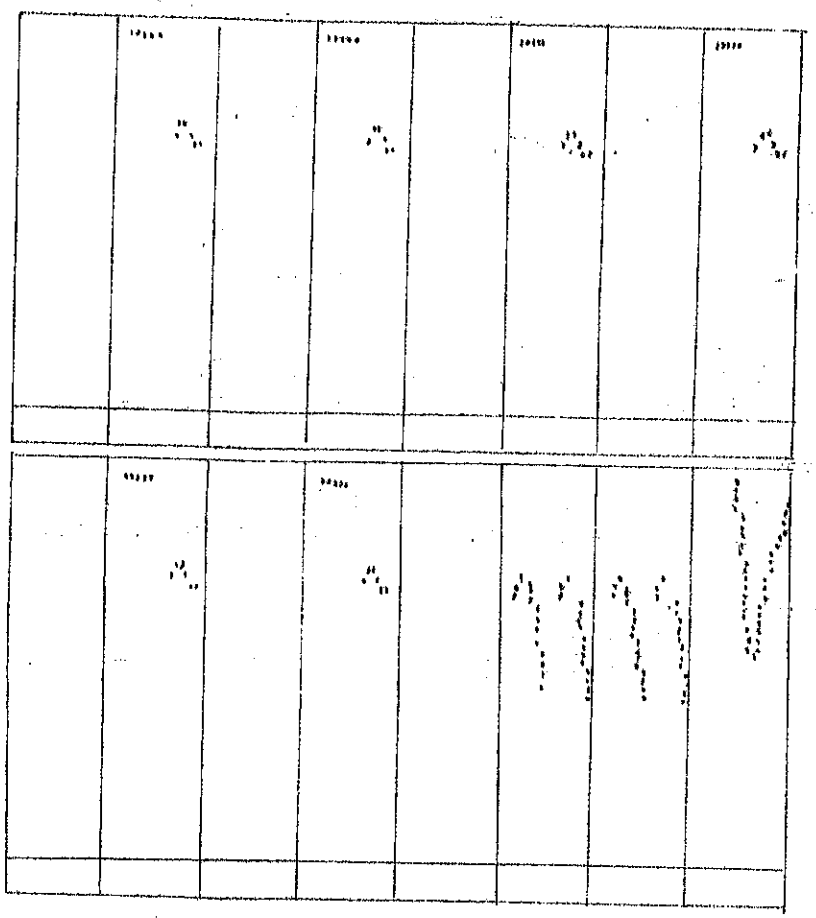


Figure 10

6. CONCLUSIONS

Grace à la construction et à l'utilisation des quatre Icophones nous avons pu vérifier par la synthèse le bien-fondé des hypothèses émises par M. LEIPP quant à la structure et à la perception des formes acoustiques de la parole. L'élaboration d'un dictionnaire de phonatomes utilisable a été un travail long et délicat qui a porté sur plus de 4 ans. Il reste à améliorer les deux autres dictionnaires pour permettre une comparaison des 3 réalisations dans le but de parfaire les invariants des phonatomes et d'étudier les anamorphoses en fréquence, d'un locuteur à l'autre.

Les résultats offrent dès maintenant une base intéressante pour entreprendre l'étude de la reconnaissance de la parole d'un locuteur quelconque.

...../

BIBLIOGRAPHIE :

1. E. LEIPP; M. CASTELLENGO; J.S. LIENARD
" parole et Gestalttheorie". Juin 1966 - Colloque GALF de LANNION.

2. M. CASTELLENGO
" Les problèmes de la perception d'une voix synthétique".
Avril 1967 - Revue d'Acoustique N°s 3-4 p.

3. J.S. LIENARD
" Le dictionnaire des éléments phonétiques et la linguistique quantitative"
Bulletin du GAM N° 22 bis - Ed. interne - Faculté des Sciences.

4. FLANAGAN
"Speech Analysis, Synthesis and Perception".
Springer Verlag Ed. P. 241.

5. E. LEIPP
" Un paradoxe en sensation de hauteur tonale".
Comptes-Rendus de 7ème ICA BUDAPEST 1971.

6. E. LEIPP; M. CASTELLENGO.
L'intelligibilité de la parole dans le chant".
Journées d'étude du Festival International du Son - Chiron Ed. 1969.

7. E. LEIPP
" Mécanique et acoustique de l'appareil phonatoire".
Bull. du GAM N° 32 - Ed.int. Faculté des Sciences. 1967.

8. M. CASTELLENGO.
" La musique des oiseaux".
Bull. du GAM N° 6 - Ed. int. Faculté des Sciences - 1965.

J. SAPALY

LES ICOPHONES

À COMMANDES OPTIQUE ET NUMÉRIQUE

JANVIER 1971 N° = 53

GAM

BULLETIN DU GROUPE D'ACOUSTIQUE MUSICALE

LABORATOIRE D'ACOUSTIQUE . FACULTÉ DES SCIENCES . TOUR 66 . PARIS 5°

LES ICOPHONES A COMMANDES OPTIQUE ET NUMERIQUE

J. SAPALY

Laboratoire de Mécanique Physique Expérimentale
de SAINT-CYR-L'ECOLE

Avant de parler de l'Icophone IV actuellement connecté à l'ordinateur IBM 1130 du Centre de Calcul Analogique d'Orsay, nous allons passer rapidement en revue les appareils qui l'ont précédé.

ICOPHONE I

L'Icophone I a été le premier montage de laboratoire destiné à vérifier les hypothèses sur l'utilisation des phonatomes en synthèse de parole.

Le principe de fonctionnement de l'Icophone I était le suivant (fig.1) : Un plateau circulaire est muni à sa périphérie d'un disque magnétique qui défile devant une tête d'enregistrement et une tête de lecture. Ce plateau est surmonté d'un cylindre de même axe et sur lequel est dessinée une suite de phonatomes dans le plan fréquence-temps. Pendant que l'ensemble fait un tour autour de son axe, une cellule photoélectrique explore une tranche du diagramme de hauteur $\Delta F = 45$ Hz et centrée sur la fréquence F : chaque fois qu'un point du dessin passe devant la cellule, celle-ci aiguille vers la tête d'enregistrement le signal fourni par un générateur dont la commande de fréquence est asservie à la position de la cellule le long de l'axe des fréquences de manière à fournir la fréquence F correspondante.

Pendant le tour suivant, la cellule explore la tranche suivante de hauteur F et de fréquence centrale $F + \Delta F$: chaque fois qu'un point du dessin excite la cellule, celle-ci aiguille vers la tête d'enregistrement le signal à la fréquence $F + \Delta F$ fourni par le générateur. Et ainsi de suite jusqu'à ce que le diagramme soit exploré sur toute sa hauteur.

On peut alors écouter la synthèse ainsi effectuée par enregistrements successifs cumulés.

ICOPHONE II

L'Icophone II a été conçu afin de lire en temps réel le diagramme fréquence-temps découpé en bandes successives de largeur 100 Hz depuis 100 Hz jusqu'à 4 400 Hz et de longue durée dans le temps.

Les phonatomes successifs sont dessinés sur une bande transparente qui provient d'une bobine débitrice de forte contenance et qui est entrainée par un système de défilement à vitesse continuement ajustable (fig.2).

La bande défile entre un tube lumineux et le lecteur optique constitué par une barrette de 44 photodiodes jointives dont chacune explore une plage élémentaire de largeur 100 Hz.

Lorsque la photodiode qui explore la plage de fréquence Centrale F est excitée par un point du diagramme, elle fait basculer un relais à mercure qui envoie vers le mélangeur le signal sinusoïdal fourni par l'oscillateur de fréquence F .

Les oscillateurs du type classique à réaction collecteur-base par self-capacité sont en fonctionnement permanent pour des raisons de stabilité en fréquence; les relais les connectent au mélangeur à l'appel des photodiodes correspondantes.

Chacun des oscillateurs peut être ajusté manuellement en fréquence et en amplitude. D'autre part chacun des oscillateurs peut être faiblement modulé en fréquence autour de sa fréquence centrale à l'aide d'un signal aléatoire provenant d'un générateur de signaux aléatoires à 44 voies indépendantes et réglables en amplitude.

C'est avec l'Icophone II qu'a été élaboré le premier dictionnaire des phonatomes par écoutes et retouches successives. La lecture de phonatomes dessinés bout à bout a permis d'obtenir une parole synthétique à débit continu mais dont la préparation demandait un délai considérable.

ICOPHONE III

L'objectif suivant a donc été de pouvoir synthétiser la parole pratiquement en temps réel, le stockage, la recherche et l'enchainement des phonatomes ne pouvant alors être assurés que par un ordinateur.

La manipulation envisagée était schématiquement la suivante :

- frappe sur clavier périphérique d'ordinateur
- traduction littérale-phonétique
- identification des phonatomes
- appel des phonatomes stockés en mémoire
- enchainement et mémorisation dans l'unité de liaison avec l'Icophone.

En ce qui concerne l'Icophone III, le sous-canal de gestion TITN devait lui envoyer toutes les 5 millisecondes :

- l'ordre à chacun des 44 oscillateurs de transmettre ou non son signal au mélangeur,
- l'ordre à l'amplificateur de sortie d'opérer à un niveau sonore choisi entre 8 possibles.

D'où le cahier des charges de l'Icophone III (fig.3) :

- 44 oscillateurs de fréquences centrales étagées de 100 Hz en 100 Hz depuis 100 Hz jusqu'à 4 400 Hz. La fréquence centrale et l'amplitude du signal de sortie sont réglables manuellement. Un signal aléatoire peut moduler faiblement l'oscillateur autour de sa fréquence centrale.

- 44 portes commandées par le sous-canal de gestion
- 1 générateur de signaux aléatoires à 44 sorties indépendantes
- 1 mélangeur
- 1 amplificateur dont un des 8 niveaux de sortie est appelé par le sous-canal.

Comme différences avec l'Icophone II notons que :

- les oscillateurs sont du type résistance-capacité à pont de Wien et sont compensés en température
- les portes ne sont plus des relais mais des transistors à effet de champ
- à l'aide du même cable, l'appareil peut être indifféremment branché sur le sous-canal de gestion de l'IBM 1130 ou sur le lecteur optique.

Ajoutons que le système de visualisation couplé à l'installation a permis l'étude et l'amélioration systématiques du dictionnaire des éléments phonétiques.

ICOPHONE IV

Pendant que se poursuivait le travail sur l'Icophone III, il est apparu que pour la suite de nos recherches, il était indispensable de vérifier certaines hypothèses portant sur l'anamorphose en fréquence des éléments phonétiques.

...../

Il fallait pour cela un Icophone dont la fréquence de chacun des oscillateurs puisse à volonté être multipliée par un même facteur que nous avons estimé devoir être compris entre 1/2 et 2, ce qui était impossible à réaliser avec l'Icophone III du fait de la nature de ses oscillateurs.

L'étude d'un nouvel Icophone étant décidée, nous en avons profité pour adjoindre aux oscillateurs des générateurs de bande de bruit afin d'améliorer la qualité de certains éléments phonétiques. Comme le sous-canal de gestion était utilisé pratiquement au maximum de sa capacité et que nous voulions pouvoir utiliser le même pour des questions d'ordre budgétaire nous avons adopté le compromis suivant pour l'Icophone IV (fig.4).

- 43 oscillateurs (au lieu de 44 sur l'Icophone III)
- 3 générateurs de bandes de bruit
- 1 système de mélange des signaux provenant des oscillateurs et des générateurs de bruit
- 1 amplificateur à 4 niveaux de sortie possibles (au lieu de 8 sur l'Icophone III)

On peut ajuster manuellement :

- la fréquence de chaque oscillateur à l'intérieur d'une gamme assez large pour que les gammes des oscillateurs successifs se recouvrent
  l'amplitude du signal de sortie de chaque oscillateur.
  la nature du signal de sortie de chaque oscillateur : sinusoïdal, carré ou triangulaire.
- la fréquence centrale et la largeur de bande de chaque générateur de bruit
  l'amplitude du signal de sortie de chacun de ces générateurs
- l'amplitude du signal de sortie de l'amplificateur entre zéro et sa valeur maximale.

On peut piloter électroniquement :

- en tout ou rien, l'amplitude du signal de sortie de chacun des oscillateurs et des générateurs de bruit entre zéro et la valeur affichée manuellement, les commandes pouvant provenir soit du sous-canal de gestion TITN soit d'un lecteur optique.

- simultanément les fréquences de tous les oscillateurs, c'est-à-dire que l'on peut anamorphoser le spectre de raies fourni par l'ensemble des oscillateurs.

- le niveau de sortie de l'amplificateur à l'une des 4 valeurs zéro, faible, moyen, maximum, le niveau maximal étant celui que l'on a affiché manuellement.

Eventuellement, chacun des oscillateurs peut être modulé de façon aléatoire autour de sa fréquence centrale à l'aide d'un générateur de signaux aléatoires indépendants les uns des autres.

L'ICOPHONE IV étant celui qui présente le plus de possibilités, nous allons donner quelques détails sur ses éléments les plus importants.

LES OSCILLATEURS

PRINCIPE DE FONCTIONNEMENT :

Nous avons choisi des oscillateurs du type convertisseur tension-fréquence à intégration, ce type de montage présentant l'avantage d'être très stable en fréquence, de fournir des signaux d'amplitude constante quelle que soit les variations de fréquence qu'on lui impose et enfin de pouvoir être aisément modulés en fréquence.

...../

Le convertisseur tension fréquence comprend (fig.5) un intégrateur, un comparateur de tension et un montage commutateur.

Si l'on applique une tension constante négative -V à l'entrée de l'intégrateur (fig.6), on obtient en sortie une tension qui partant de zéro, croît linéairement dans le temps. Lorsque cette tension atteint une valeur + E caractéristique du comparateur, celui-ci fait basculer le commutateur qui fournit alors à l'intégrateur la tension + V, d'où à la sortie de l'intégrateur une décroissance du signal avec la même linéarité qu'à la montée. La tension décroît jusqu'à une valeur - E caractéristique du comparateur, d'où basculement du commutateur qui fournit alors à l'intégrateur la tension -V et le cycle recommence.

La pente du signal à la sortie de l'intégrateur étant proportionnelle à V, on voit que la fréquence de l'oscillateur varie linéairement en fonction de V (fig.7).

On obtient donc en sortie de l'intégrateur un signal triangulaire d'amplitude constante et dont la fréquence dépend de la valeur de V, en sortie du comparateur un signal carré d'amplitude constante et de même fréquence, en sortie du commutateur un signal carré d'amplitude proportionnelle à V et de même fréquence.

Le signal de sortie sinusoïdal est élaboré à partir du signal triangulaire par l'intermédiaire d'un montage écrêteur à diodes (fig.8) qui permet d'obtenir une distorsion ne contenant que 2 % de l'harmonique 7 et 1 % de l'harmonique 9.

APPEL DU SIGNAL :

On peut annuler le signal de sortie en bloquant l'oscillateur, c'est-à-dire en court-circuitant électroniquement l'entrée et la sortie de l'intégrateur. On obtient à nouveau le signal en supprimant le court-circuit, l'oscillateur retrouvant exactement sa fréquence en un temps qui ne peut dépasser le quart de la période.

Le court-circuit est réalisé à l'aide d'un transistor à effet de champ dont on commande la résistance en tout ou rien à l'aide du signal fourni par le sous-canal de gestion ou par le lecteur optique.

PILOTAGE EN FREQUENCE :

- La caractéristique de fonctionnement du convertisseur tension-fréquence étant une droite passant par l'origine (fig.9), on voit qu'en réglant manuellement la tension V à une valeur fixe V_o, on fixe la fréquence centrale F_o de l'oscillateur.

- Si l'on désire moduler l'oscillateur en fréquence autour de la valeur F_o, il suffit de superposer à la tension fixe V_o une tension variable v(t) que la fréquence de l'oscillateur suivra fidèlement à l'intérieur des limites de la caractéristique de fonctionnement.

Les pentes des caractéristiques des oscillateurs étant proportionnelles à leurs fréquences centrales (fig.10) il est possible, à partir d'un unique signal de commande, de multiplier la fréquence centrale de chacun des oscillateurs par un même facteur qui, dans la réalisation actuelle, peut être choisi entre 1/2 et 2.

- L'unique signal v (t) qui module en fréquence l'ensemble des oscillateurs peut être engendré manuellement ou électroniquement.

En ce qui concerne la commande électronique à partir de l'ordinateur IBM 1130, nous disposons sur l'actuel sous-canal de gestion TITN de quatre circuits disponibles, chacun de ces circuits pouvant fournir l'un ou l'autre de deux niveaux de tension normalisés, la présence de l'un des niveaux pouvant être renouvelée toutes les 5 millisecondes.

...../

Un convertisseur fonctionnant en code 1-2-4-8 et attaqué par les quatre circuits de commande fournit un signal de sortie dont l'amplitude peut prendre l'une de 16 valeurs normalisées régulièrement étagées. On peut ainsi à partir de l'ordinateur élaborer un signal de commande de modulation dont le point représentatif dans le diagramme v (t) ne peut se déplacer que parallélement à l'un des axes (fig.11). On ajoute à la tension fournie par le convertisseur une contre-tension qui permet de pouvoir amener l'une des 16 valeurs au niveau de tension zéro, d'où possibilité de moduler de part et d'autre de sa fréquence centrale chacun des oscillateurs (fig.12). Le signal de commande ainsi constitué est appliqué à tous les oscillateurs après passage par un amplificateur dont la commande de gain permet d'élargir plus ou moins la plage de modulation en fréquence. Un filtre passe-bas permet de lisser la courbe v (t) obtenue par segments rectilignes horizontaux et verticaux.

GENERATEUR DE BANDE DE BRUIT

Chacun de ces générateurs est constitué par un générateur de bruit blanc classique suivi d'un filtre actif et d'un amplificateur (fig.13). Le générateur de bruit, en fonctionnement permanent, est commun aux trois systèmes.

La fréquence centrale et la bande passante de chaque filtre actif sont ajustables manuellement. Pour chaque amplificateur, l'amplitude de sortie est réglable manuellement tandis qu'elle peut être commandée électroniquement en tout ou rien par un signal provenant du sous-canal de gestion ou d'un lecteur optique.

AMPLIFICATEUR A NIVEAU COMMANDE

La liaison entre préamplificateur et amplificateur est réalisée à l'aide d'un pont de résistances qui peuvent être individuellement court-circuitées par des transistors à effet de champ commandés en tout ou rien par les signaux provenant du sous-canal de gestion ou d'un lecteur optique.

On a ainsi la possibilité de fixer le niveau global de sortie à l'une des 4 valeurs préréglées.

30 Janvier 1971

J. SAPALY.

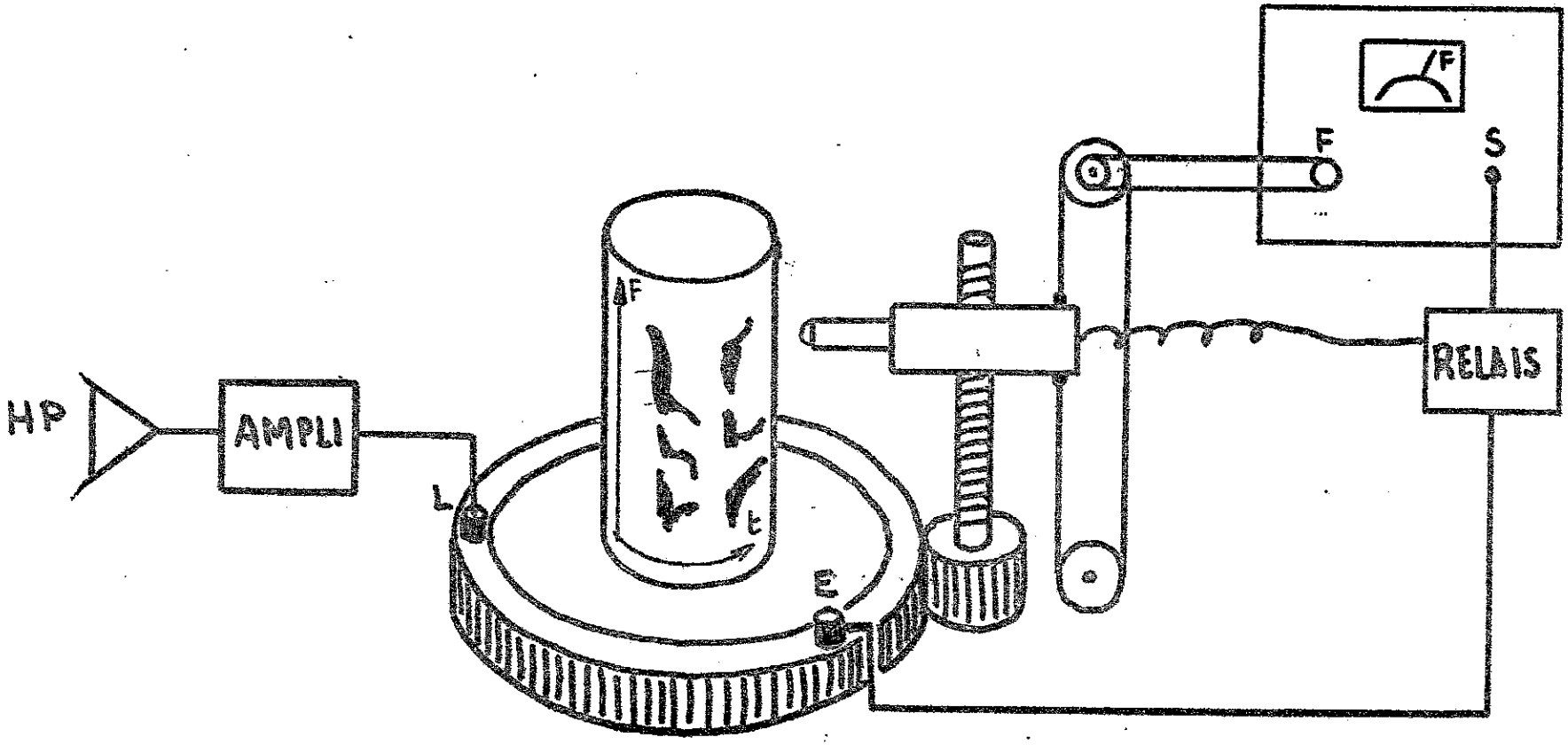


FIG 1 : SCHEMA de PRINCIPE
de l' ICOPHONE I

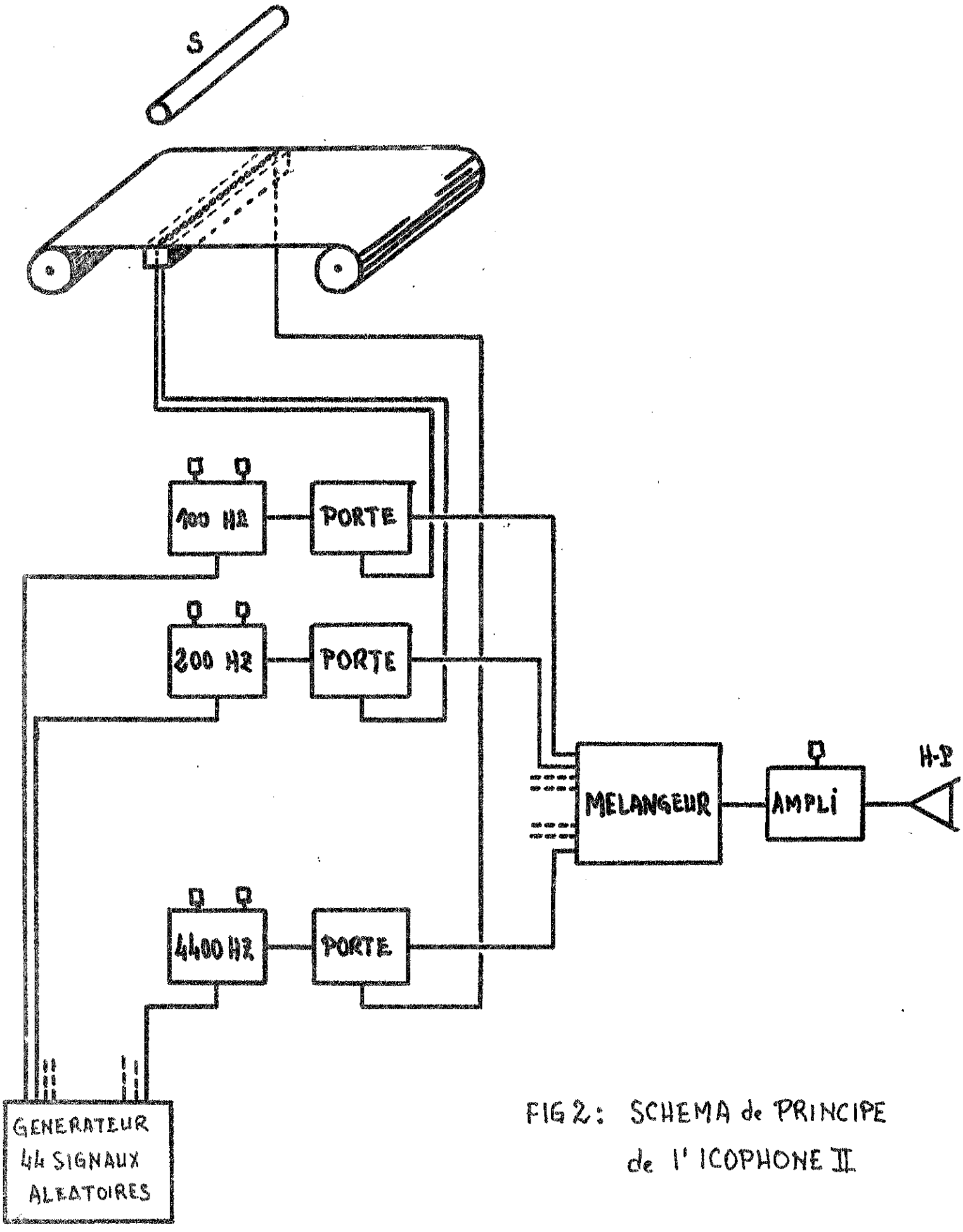


FIG 2: SCHEMA de PRINCIPE
de l' ICOPHONE II

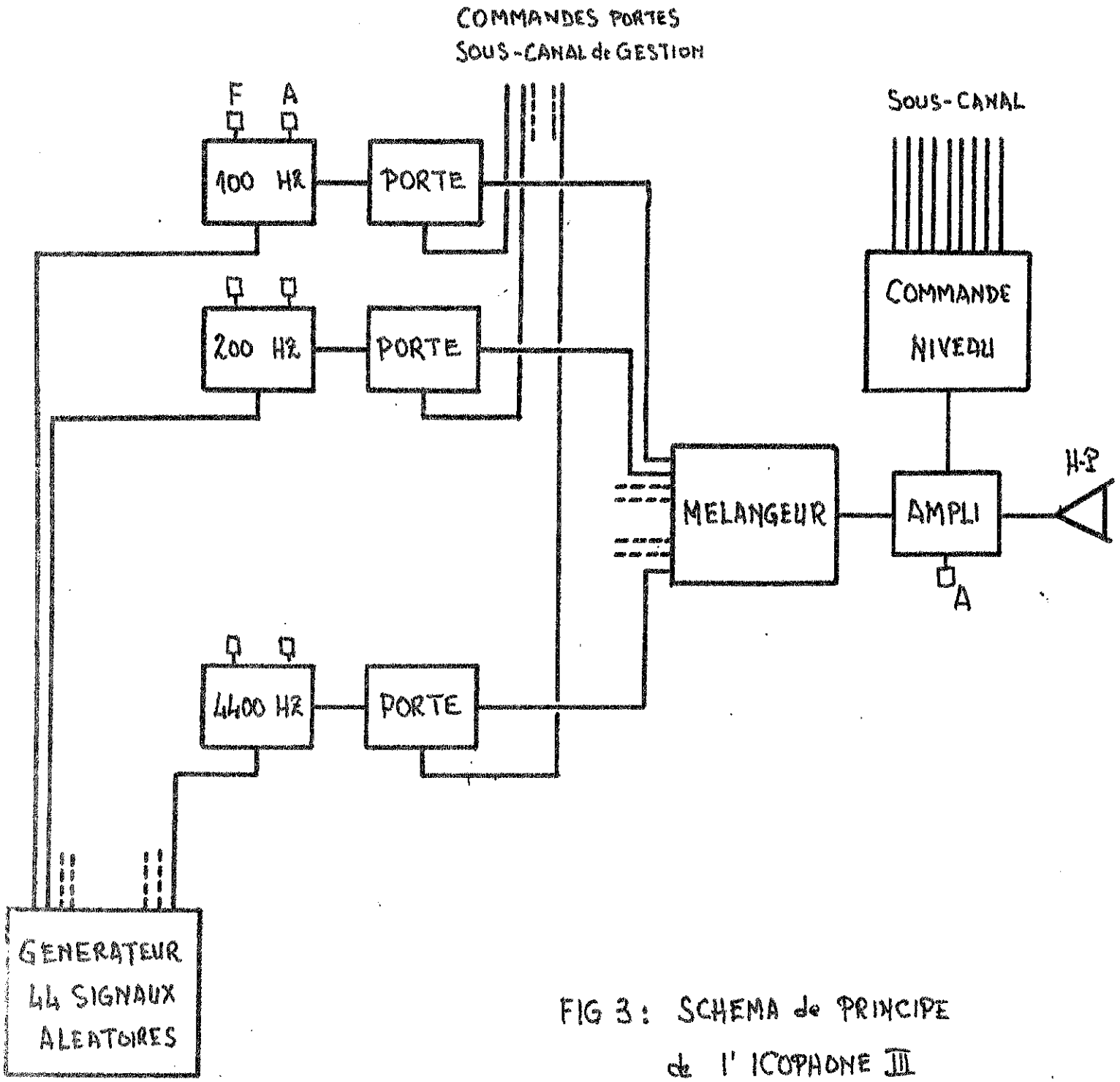


FIG 3 : SCHEMA de PRINCIPE
de l' ICOPHONE III

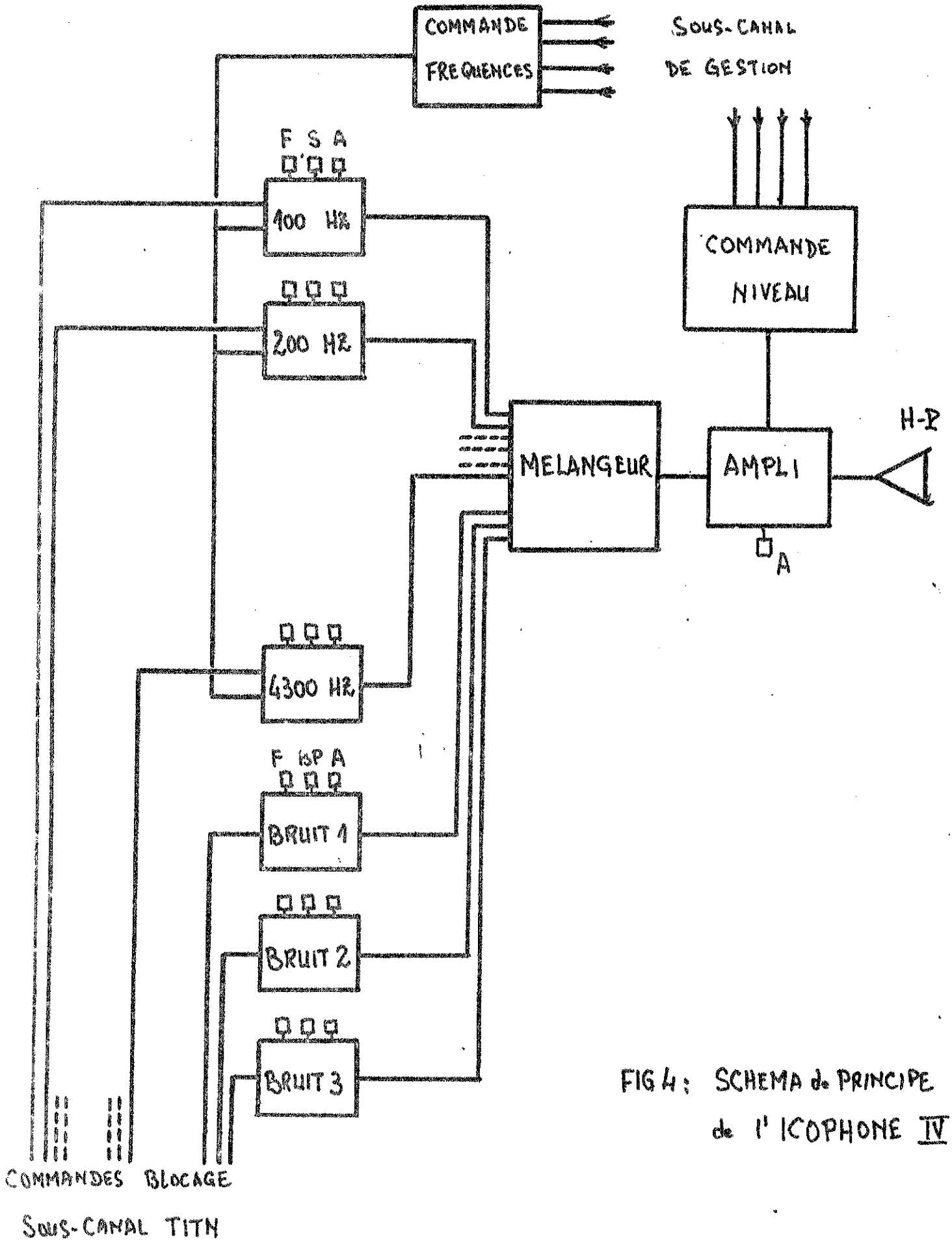


FIG 4 : SCHEMA de PRINCIPE de l'ICOPHONE IV

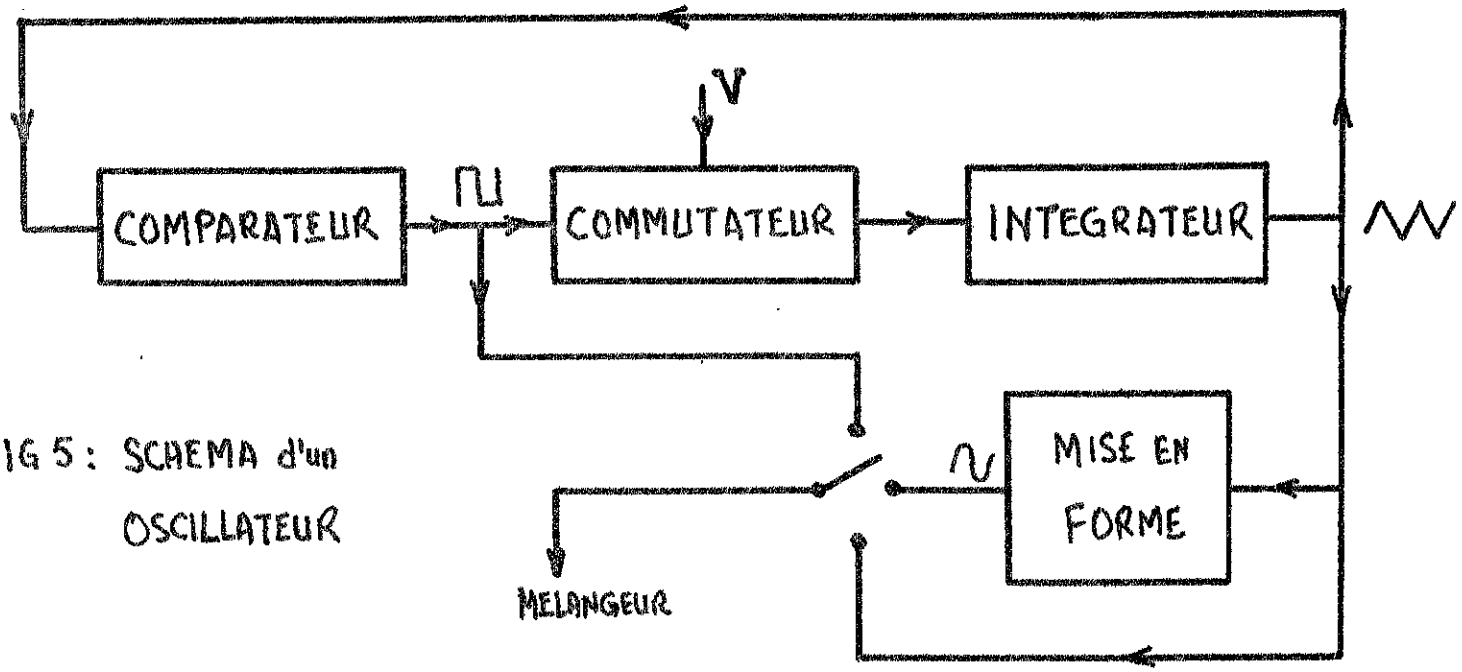


FIG 5 : SCHEMA d'un OSCILLATEUR

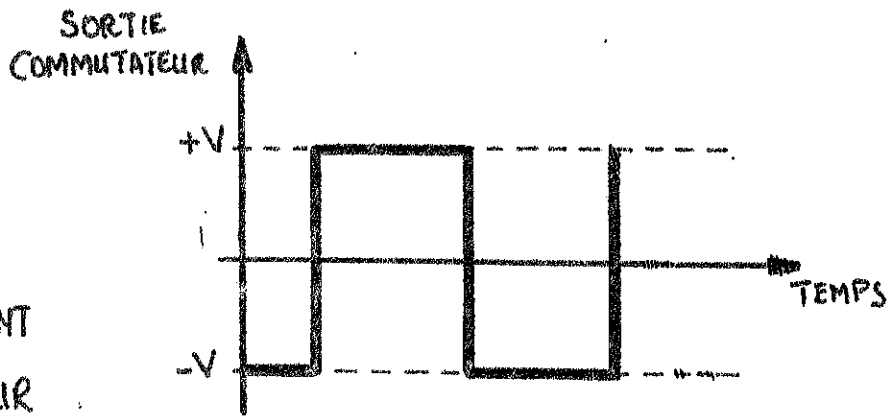
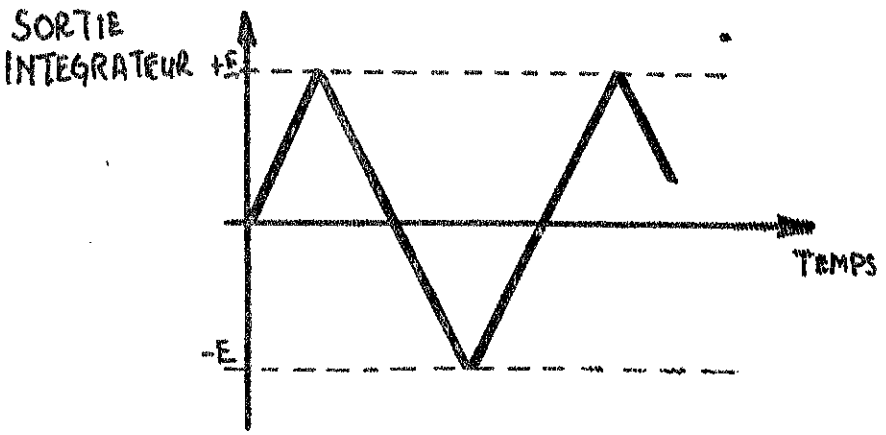


FIG 6 : PRINCIPE de FONCTIONNEMENT de l'OSCILLATEUR



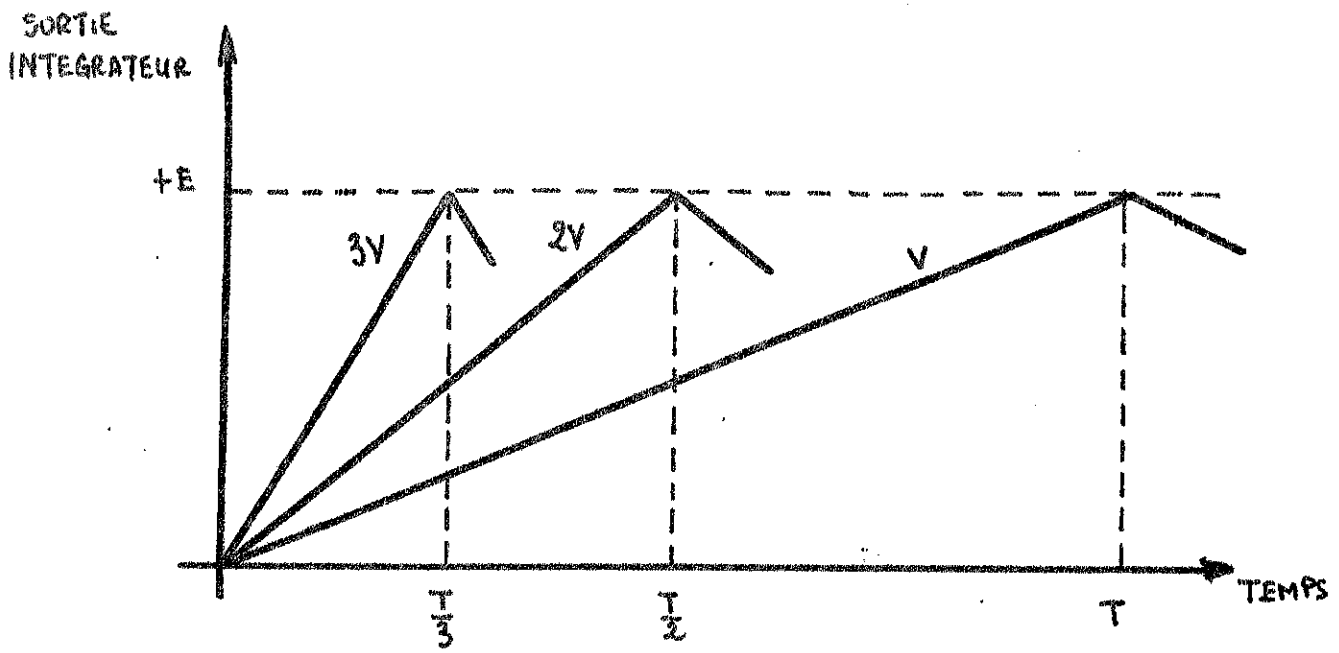


FIG 7 : LA FREQUENCE DE L'OSCILLATEUR DEPEND DE LA TENSION DE COMMANDE V

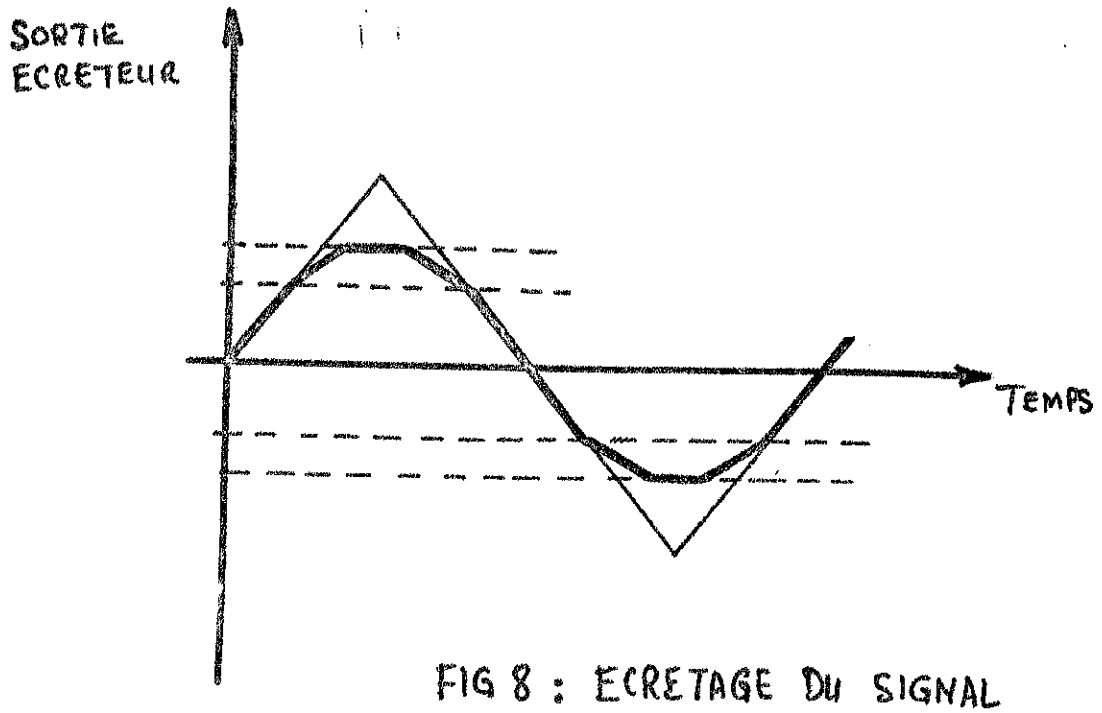


FIG 8 : ECRETAGE DU SIGNAL TRIANGULAIRE

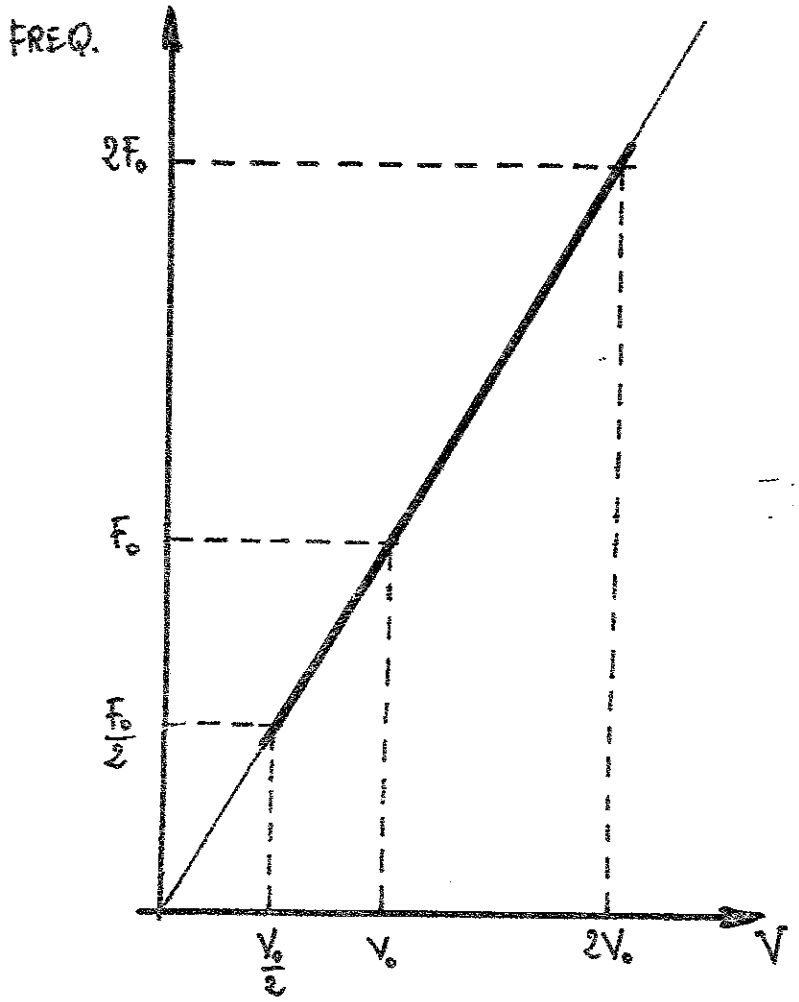


FIG 9 : CARACTERISTIQUE FREQUENCE-TENSION D'UN OSCILLATEUR

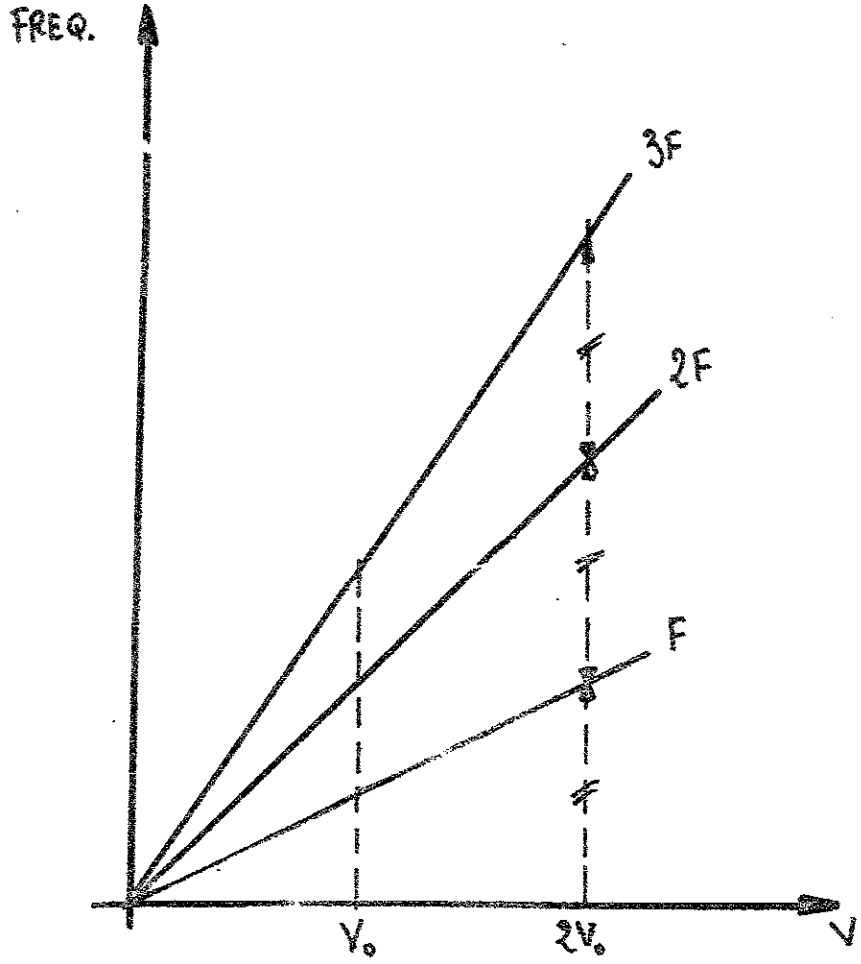


FIG 10 : CARACTERISTIQUES EN FONCTION DES FREQUENCES CENTRALES DES OSCILLATEURS

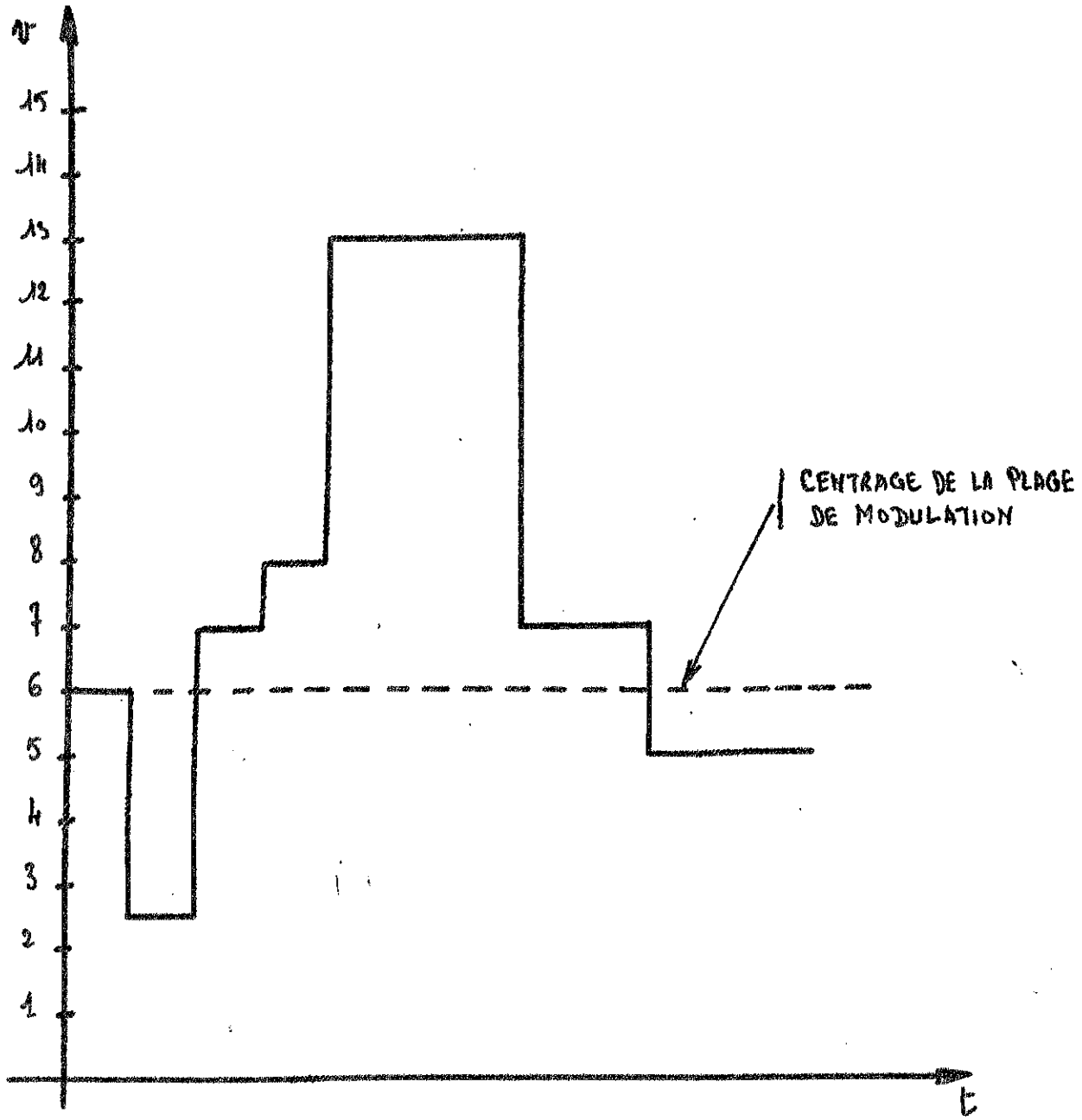


FIG 11 : SIGNAL DE COMMANDE DE MODULATION EN FREQUENCE

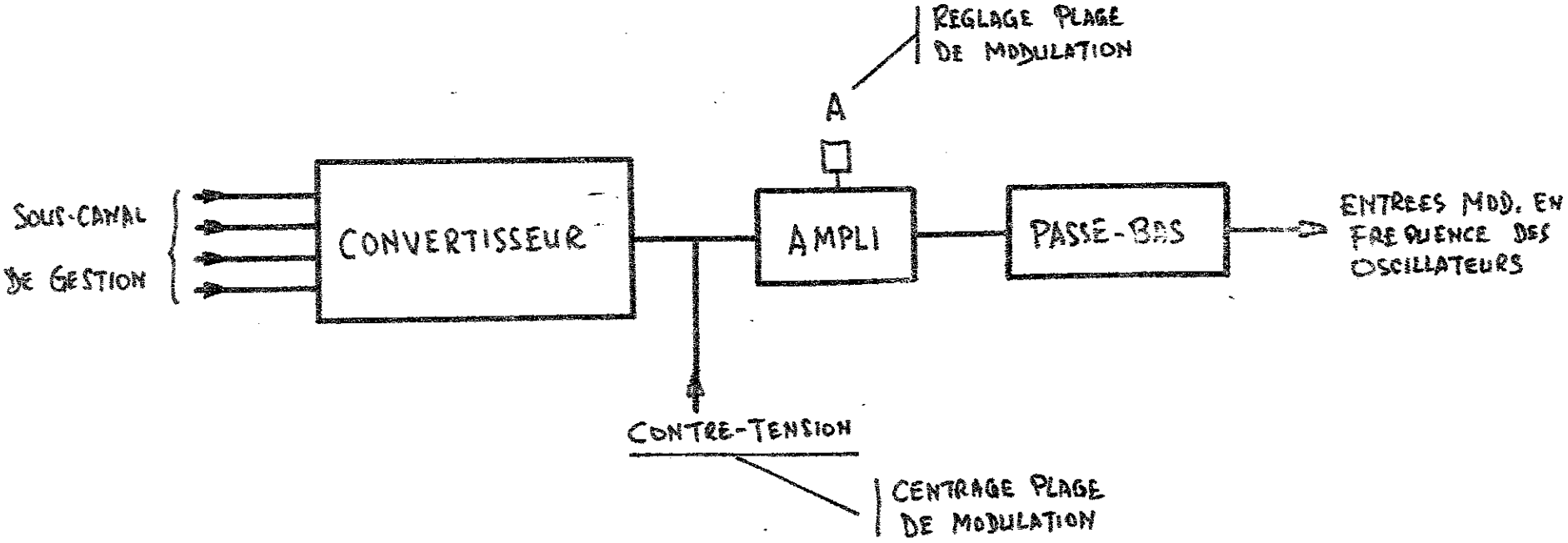


FIG 12 : MODULATION EN FREQUENCE
DES OSCILLATEURS

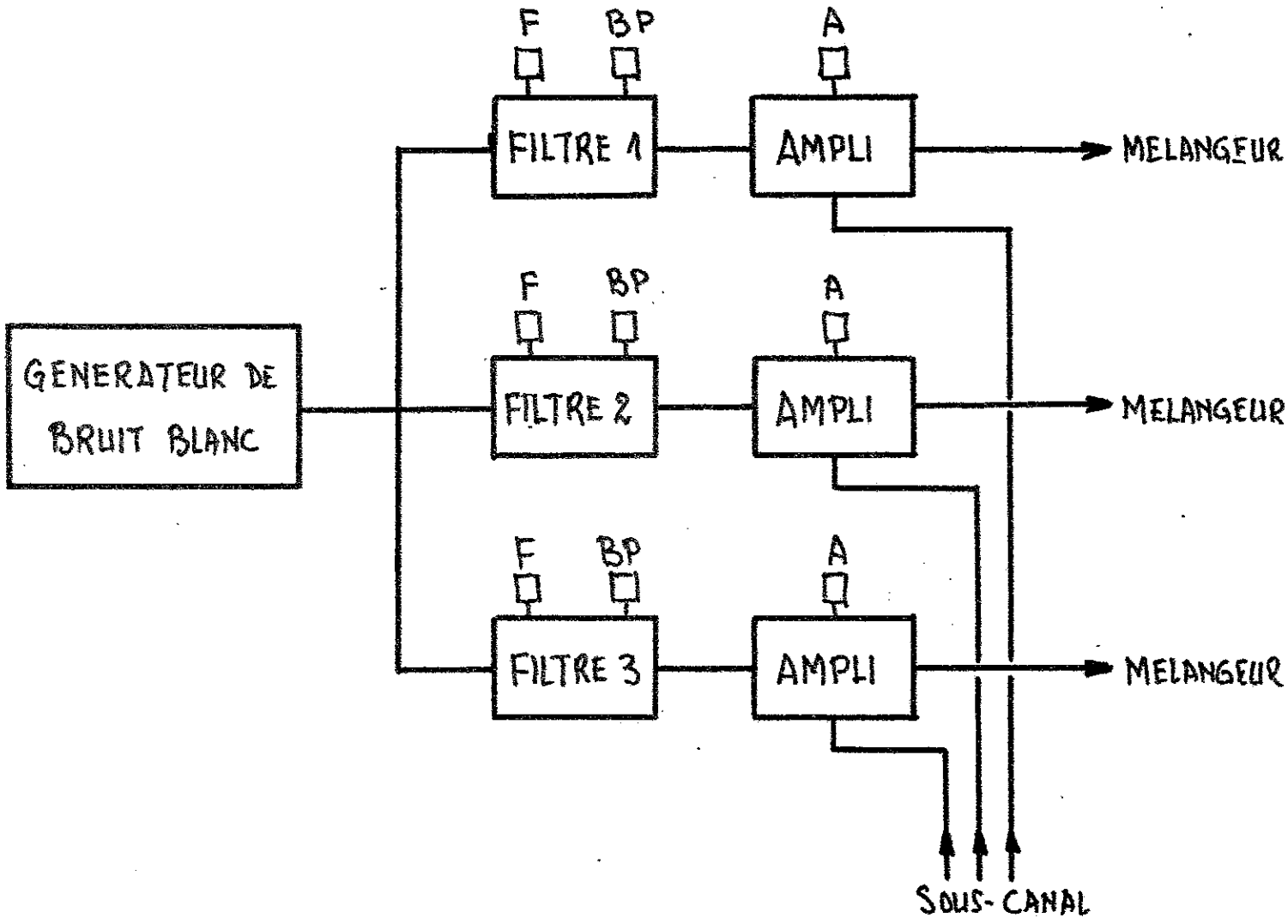


FIG 13 : GENERATEURS DE BRUIT PROGRAMMES

D. TEIL

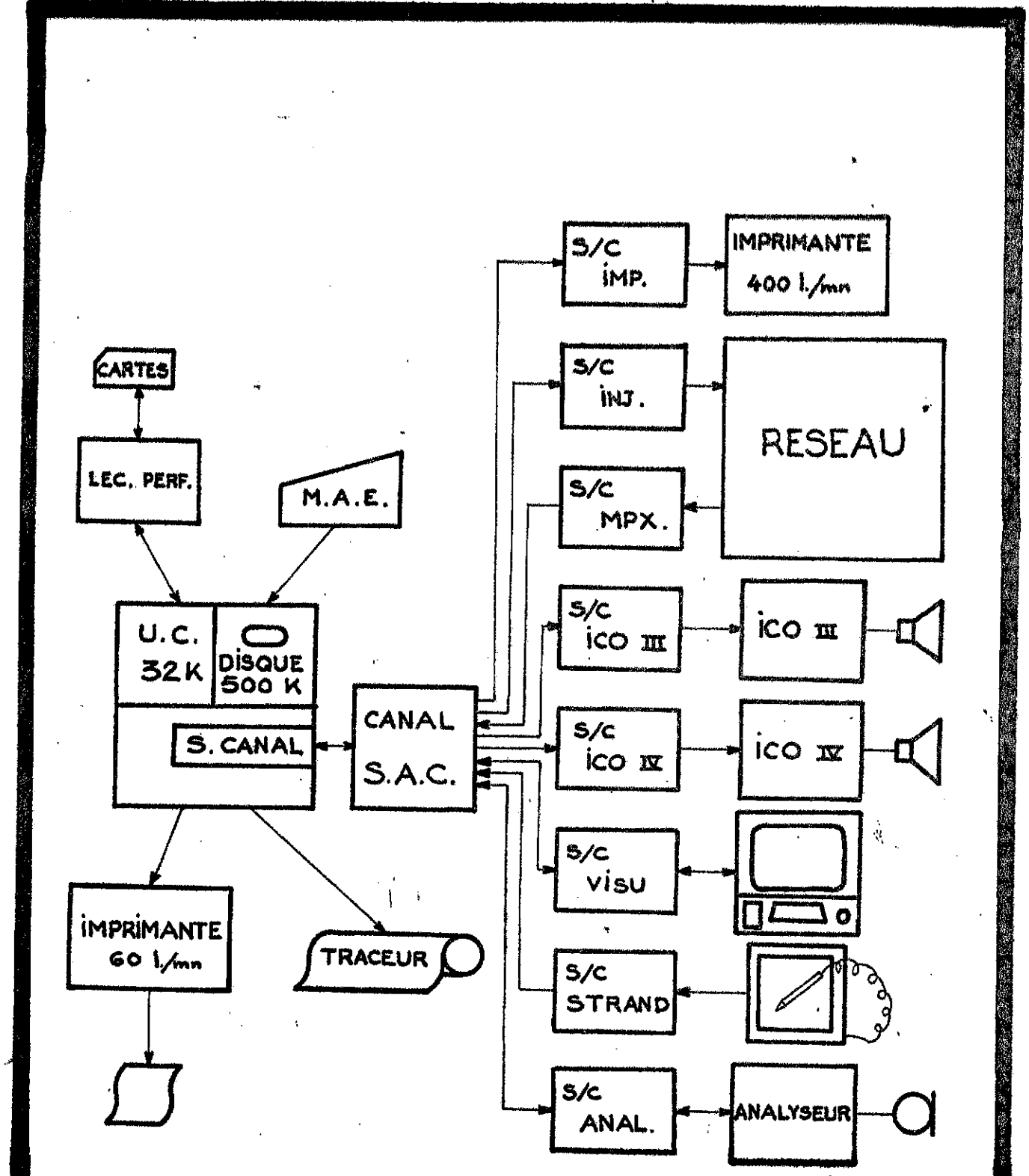
DEVELOPPEMENTS NUMERIQUES DE L'ICOPHONE

JANVIER 1971. N° 53

GAM

BULLETIN DU GROUPE D'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES . Tour 66 . Place Jussieu . PARIS 5°

TEIL - janvier 71 -



-Fig 1-

1°) INSTALLATION DU CENTRE DE CALCUL ANALOGIQUE DU C.N.R.S.

a) - Description de l'ensemble (Fig. 1)

Le système comprend un ordinateur I.B.M. 1130 muni de ses périphériques classiques : lecteur perforateur de cartes, imprimante et traceur de courbe. La mémoire centrale a une capacité de 32 000 mots de 16 bits. A cette mémoire centrale est associée une unité de disque magnétique de 512 000 mots.

Nous avons connecté à l'ordinateur un certain nombre de périphériques non classiques :

- Un réseau résistif de 1024 noeuds pour faire du calcul hybride et traiter des équations aux dérivées partielles.
- Une console de visualisation graphique qui fonctionne également en affichage de caractères. Cette console est munie d'un clavier de machine à écrire, de touches fonction, et d'un système de désignation de point du type manche à balais.
- L'Icéphone III
- L'Icophone IV

Ces périphériques sont gérés par des sous-canaux : ceux-ci mémorisent puis valident les informations échangées avec l'ordinateur, dans un sens ou dans l'autre.

Tous les sous-canaux sont reliés à l'IBM 1130 par une unité d'entrée-sortie appelée canal SAC (Storage Acces Channel). C'est elle qui assure la gestion des sous-canaux.

En projet nous avons 3 autres périphériques non standards gérés par le même canal : une imprimante rapide, un analyseur spécialisé et une tablette d'entrée graphique (STRAND).

b) - Description de la tablette.

Cette tablette est constituée d'une plaque de verre de forme carrée dont le côté est d'environ 30 cm. Sur cette vitre a été déposée une très mince couche d'un matériau conducteur. La couche est si mince que l'ensemble est tout à fait transparent.

Sur les bords sont fixées des électrodes dans lesquelles on envoie du courant. Il s'établit alors un champ électrique univorme dans la couche conductrice. Grace à un système de commutation et par un crayon connecté à la tablette on vient prélever la tension sur la couche, ce qui permet de repérer avec une grande précision l'emplacement relatif du point désigné. Les tensions sont converties en valeurs numériques et envoyées dans le calculateur.

Avec un tel système nous pouvons introduire très rapidement n'importe quel dessin de format convenable. C'est grâce à cette tablette que nous envisageons d'entrer en machine d'autres dictionnaires de phonatomes.

...../

2°) PROGRAMMES DE TRAITEMENT

Quelles sont les opérations effectuées par l'ordinateur pour passer d'un texte écrit en français à sa diction par l'Icophone ?

Le texte frappé en clair est d'abord transformé en une suite de symboles phonétiques grace à un programme de traduction phonétique. L'ordinateur se charge ensuite d'appeler en mémoire les phonatomes correspondants rangés sur le disque sous forme numérique. Cette suite de phonatomes est alors transmise au synthétiseur de parole qui restitue les sons correspondants au texte introduit.

a) - Programme de traduction phonétique :

Le programme que nous avons mis au point nous donne une bonne approximation de la prononciation courante d'un texte écrit en français. La traduction phonétique obtenue n'est pas parfaite car la prononciation est quelquefois assujettie à des impératifs grammaticaux qui dépendent de la structure et du sens de la phrase complète. Mais il est illusoire de traiter ces cas quand on sait qu'il existe une grande diversité de prononciations et d'accents différents suivant les régions.

Le travail a d'abord consisté à établir un certain nombre de règles de prononciation ainsi que les exceptions à ces règles (fig.2). Ces règles de prononciation sont des règles simples que l'on peut trouver dans les livres de lecture utilisés dans les écoles primaires pour apprendre à lire aux enfants.

par ex. O suivi de I se prononce " OUA "

GE se prononce " JE " alors que GUE se prononce " GE " (G dur).

Mais comme à toute règle bien établie il y a des exceptions qui constituent les particularités de la langue française. Celles-ci sont directement liées à la connaissance du mot : la mémoire individuelle entre en jeu c'est le cas des mots " oignon ", " monsieur " par exemple.

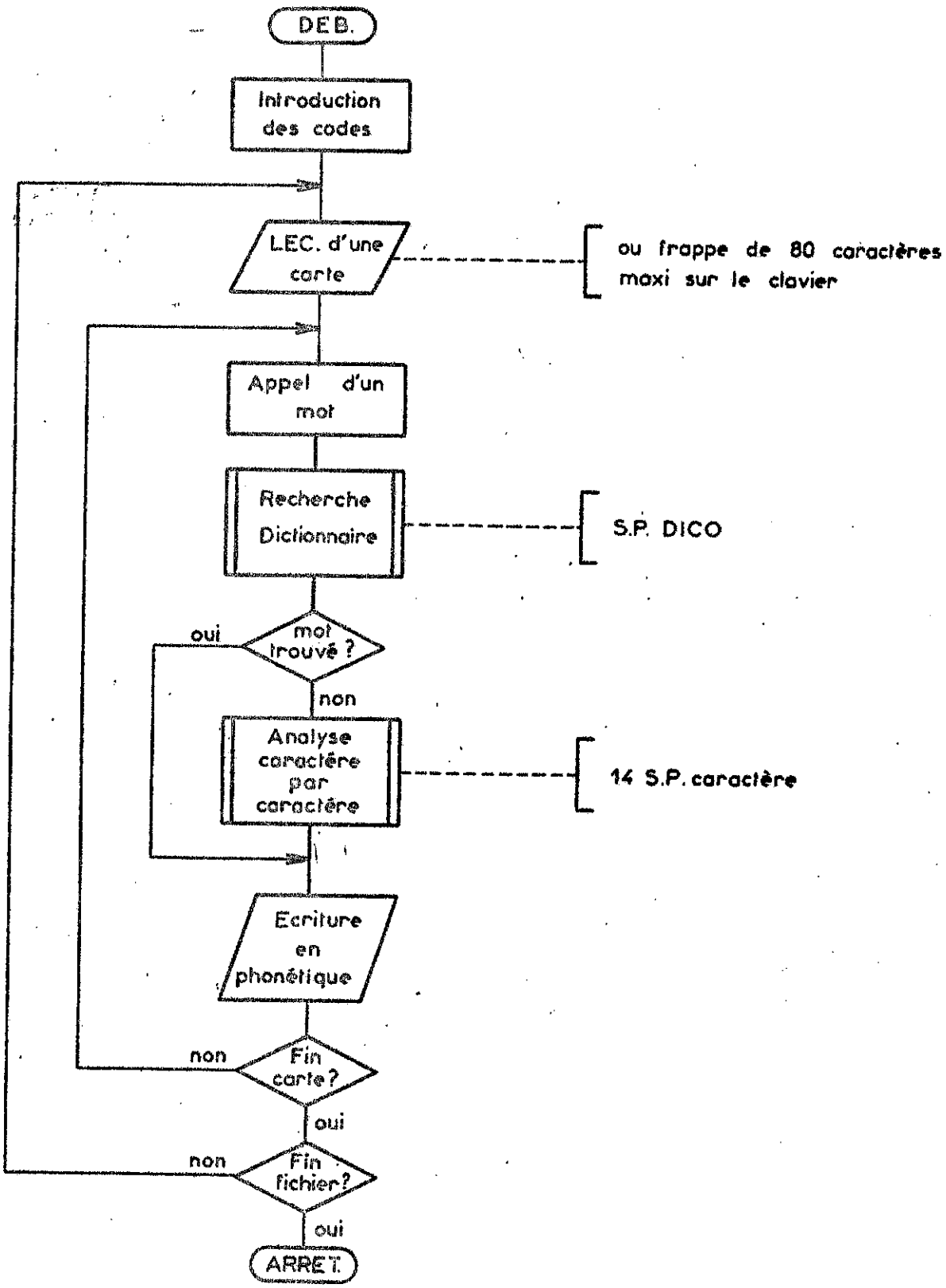
Nous avons donc constitué un dictionnaire contenant une suite de mots avec leur prononciation particulière en phonétique. Les codes utilisés pour l'introduction du texte écrit sont les 26 lettres de l'alphabet augmentées de) pour é et (pour è . Ces codes, représentant les 30 phonèmes utilisés, ont été choisis en fonction des caractères disponibles sur l'imprimante de l'ordinateur.

Ensuite les mots ou les phrases sont lus soit sur cartes soit par le clavier par groupe de 80 caractères maximum. Chaque mot est d'abord isolé en fonction des blancs ou de la ponctuation qui l'entoure.

On consulte ensuite le dictionnaire des particularités pour identifier le mot en cas de prononciation particulière. Cela suppose évidemment que le mot a déjà été répertorié. Ce dictionnaire des particularités est situé sur le disque magnétique car il peut être très important. Actuellement il est tout à fait incomplet puisqu'il ne comprend qu'une quarantaine de mots. Les mots sont rangés sur le disque par ordre alphabétique, et suivis de leur traduction phonétique. Ce dictionnaire, encore très sommaire, peut être complété chaque fois que de nouveaux éléments se présentent.

Si le mot est trouvé, sa traduction est directement transcrite dans la zone mémoire réservée au résultat. C'est à ce niveau que l'on conserve la consonne finale du mot s'il y en a une et si elle est susceptible d'engendrer une liaison avec le mot suivant. .....

TEIL, Janvier 71



-Fig 2-

Si le mot n'est pas trouvé il est alors analysé lettre par lettre par les sous-programmes d'étude de caractère. Certaines lettres ne posent pas de difficultés parce qu'elles se prononcent toujours de la même manière, suivant leur phonème associé c'est le cas des letres : B, F, J, K, L, M, N, R, V, Z. Les règles de prononciations des autres lettres sont plus complexes et peuvent être multiples. Pour le E par exemple nous avons établi 24 règles, explicitées dans 14 sous-programmes.

Ce programme de traduction utilise également 4 sous-programmes généraux. Le 1er détermine si la lettre considérée est une voyelle ou une consonne. Le 2ème précise si la consonne finale d'un mot est suivie d'un blanc ou d'un "S" auquel cas elle sera considérée comme muette. Le 3ème étudie les diffèrentes combinaisons entre les consonnes "m" et "n", et le 4ème sert à la consultation du dictionnaire des particularités.

Quand l'analyse du mot est terminée, sa traduction est rangée dans une zone résultat qui pourra être envoyée sur l'imprimante ou rangée sur disque, ou encore perforée sur cartes, suivant l'exploitation ultérieure. Le programme va ensuite chercher le mot suivant. S'il commence par une voyelle alors nous validons la liaison; ensuite le cycle continue.

Résultats :

La programmation a été faite en langage Fortran IV et l'exploitation a été réalisée sur l'ordinateur IBM 1130. Pour les essais de validité du programme nous avons pris plusieurs textes différents, l'ensemble représentant environ 17 000 phonèmes.

Le travail de l'ordinateur se résume à une suite de comparaisons, la traduction est donc rapide : il faut compter 1 seconde pour analyser une ligne de texte de 70 à 80 phonèmes.

Le programme a été testé sur un petit nombre de textes; il en résulte que certains cas de ponctuation n'ont pas été envisagés, mais l'aspect modulaire des sous-programmes permet de les corriger facilement.

Dans l'ensemble on ne trouve guère plus de 2 ou 3 erreurs d'interprétation phonétique sur une centaine de phonèmes. Si l'on exclut les mots à prononciation particulière les erreurs de traduction proprement dites sont faibles et concernent surtout certaines terminaisons de verbes en "ent" :

Soit par exemple la phrase suivante : "les renards plument rapidement les poules". Les terminaisons des 2 mots seront traduitent par (AN); à priori, il semble que seule une analyse grammaticale de la phrase nous dira où se trouve le verbe auquel cas la terminaison "ent" ne se prononce pas.

b) - Edition en temps réel :

Pour faire parler le synthétiseur on pourrait très bien ranger en mémoire tous les phonatomes constituant la phrase et une fois ce travail terminé, commander la sortie.

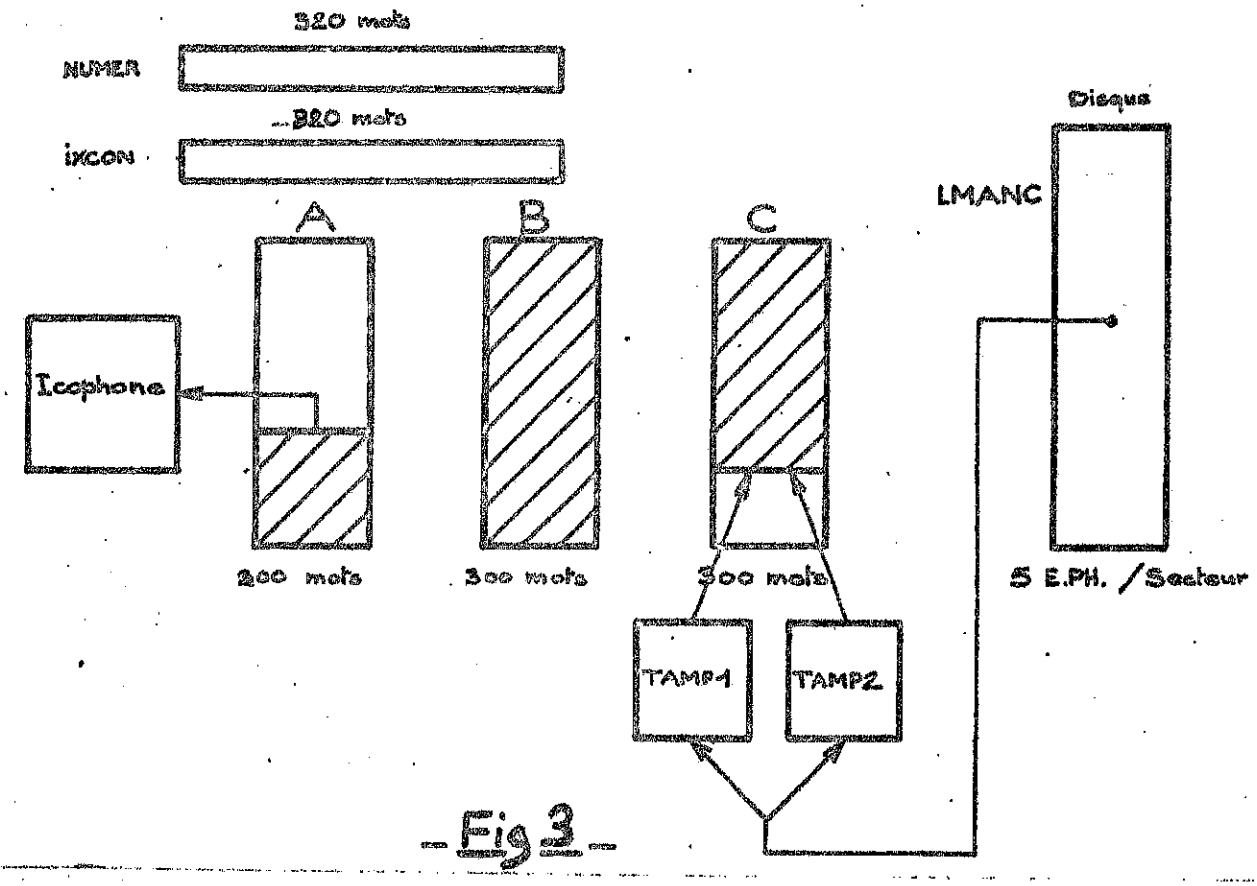
Il y a plusieurs inconvénients dans l'utilisation de ce processus :

- il nécessite beaucoup de mémoires sachant qu'il en faut 60 pour un phonatome.

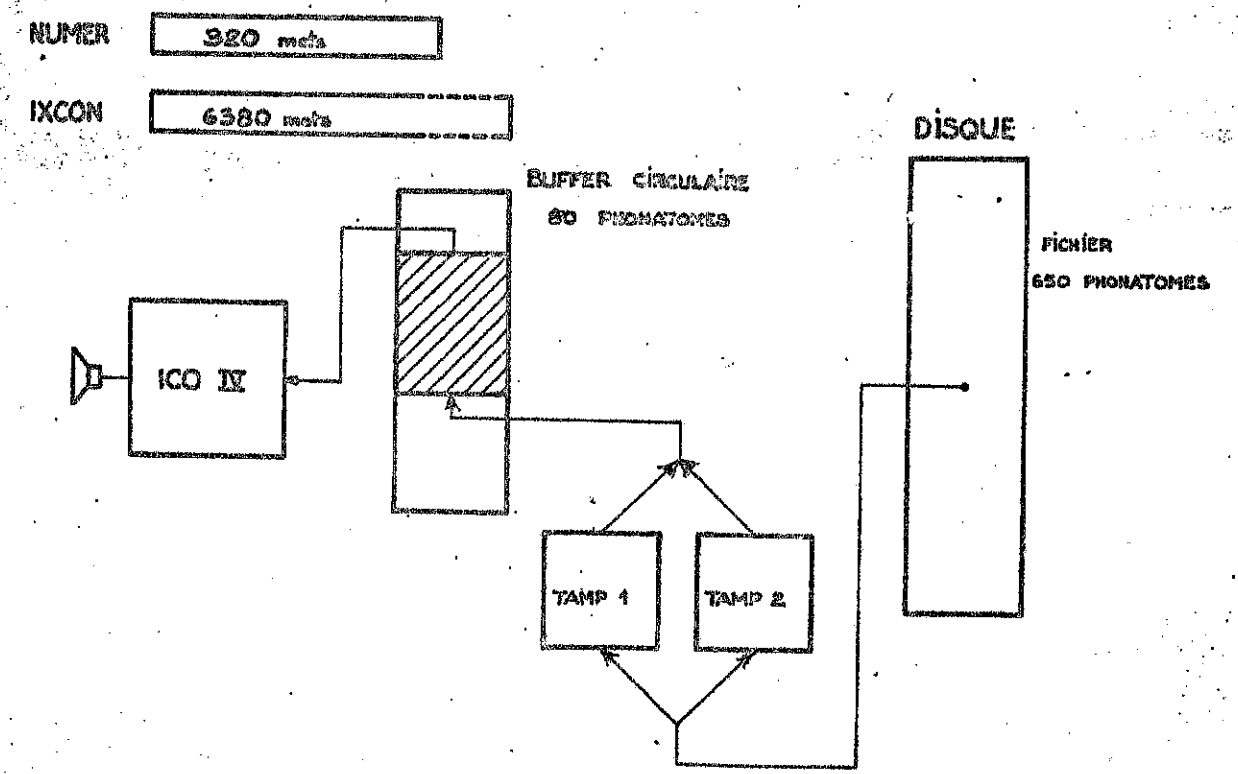
- quelle que soit la taille de l'unité centrale, nous ne pourrons ranger que quelques secondes de paroles.

...../

TEIL - Janvier 71



-Fig 3-



-Fig. 4-

- il faut attendre le chargement complet avant de pouvoir commander le synthétiseur.

Pour palier à ces inconvénients nous avons réalisé des programmes du type temps réel.

Pour l'Icophone III nous avons utilisé un système de chargement en bascule avec 5 zones tampons (fig.3) : un bloc d'information, extrait du fichier situé sur disque, est placé dans une des 2 zones tampons 1 ou 2 fonctionnant en bascule. Dans ce bloc le phonatome est prélevé et rangé dans la zone C. Le phonatome à sortir est prélevé dans la zone A, la zone B étant constamment pleine. Une permutation circulaire du rôle de ces 3 zones assure un fonctionnement continu. Ce programme nécessite une occupation mémoire de seulement 2500 mots de 16 bits.

Pour l'Icophone IV nous utilisons un buffer circulaire (fig. 4) : Le prélèvement des phonatomes dans le fichier situé sur le disque est le même que dans le cas précédent, mais les 3 buffers A, B, C sont remplacés par un seul que l'on remplit en même temps qu'on le vide. L'avantage de ce système par rapport à l'autre est sa simplicité. Dans les 2 cas il faut attendre un remplissage minimum d'une dizaine de phonatomes avant de commencer l'édition sur l'Icophone. Une fois le processus démarré le chargement des phonatomes se fait en même temps que la sortie sur l'Icophone : on peut ainsi assurer un débit de parole pratiquement illimité.

Quand on frappe une phrase d'environ 80 caractères au clavier de l'ordinateur, il faut compter actuellement environ 1 à 2 secondes pour la traduction phonétique, la recherche des numéros des phonatomes et le remplissage nécessaire au démarrage. C'est un délai très faible mais qui est encore trop grand s'il s'agit de transmettre des conversations téléphoniques.

En fait ce délai, pris en grande partie par le programme de traduction phonétique, est uniquement fonction de l'ordinateur utilisé et du type de programmation mis en jeu; il pourrait facilement être réduit à une fraction de seconde.

A. CALINET

DESCRIPTION DE DEUX PROGRAMMES POUR LA CORRECTION ET L'ANAMORPHOSE DES ÉLÉMENTS PHONÉTIQUES

JANVIER 1971 N° 53

GAM

BULLETIN DU GROUPE D'ACOUSTIQUE MUSICALE
FACULTÉ des SCIENCES . Tour 66 . Place Jussieu . PARIS 5°

CALINET - GAM Janvier 1971

DESCRIPTION DE DEUX PROGRAMMES

D'UTILISATION DE L'ICOPHONE A COMMANDE NUMERIQUE

1 - Programme " CREPH "

2 - Programme " ANAMO "

- C R E P H -

Le programme " CREPH " est utilisé pour la création et la correction des éléments phonétiques.

Les périphériques utilisés sont :

- LECTEUR
- IMPRIMANTE
- UNITE DE DISQUE
- UNITE DE VISUALISATION
- ICOPHONE 3 A COMMANDE NUMERIQUE

Nous ne décrirons pas le fonctionnement des périphériques, cependant, nous parlerons un peu de l'unité de visualisation.

C'est à partir d'elle que l'utilisateur fera ses requêtes au programme.

L'unité de visualisation comprend :

- UN CLAVIER
- TRENTE DEUX TOUCHES FONCTION
- UNE TOUCHE FONCTION SPECIALE

qui permet de déplacer un spot sur l'écran suivant huit directions.

Elle est constituée d'un tube à mémoire morte. Il n'est donc pas possible d'effacer une partie de l'écran en conservant une autre partie. Nous ne possédons que l'effacement total ce qui nécessite de garder en mémoire l'information, de venir la modifier en mémoire et d'envoyer à l'unité de visualisation le nouveau contenu du "BUFFER".

C'est à partir du clavier que l'utilisateur donne au programme l'information, puis à l'aide des touches fonction, lui indiquera de faire tel ou tel traitement sur cette information.

Nous expliquerons un peu plus loin la structure du programme du point de vue général, ce qui permettra de placer l'utilisateur par rapport au programme.

Contrairement à certains travaux ou l'utilisateur ayant donné l'ordre d'exécuter doit attendre la fin et relancer le programme pour un autre traitement, le programme " CREPH " est architecturé de telle manière que l'utilisateur en est le maitre et non l'esclave. C'est-à-dire qu'il peut interrompre n'importe quel traitement à n'importe quel moment sans trop de danger ceci par le jeu des interruptions traitées au niveau du système.

Avant de commencer l'étude du programme nous donnerons quelques indications sur les fichiers et les tableaux utilisés.

...../

FICHIERS :

LMANC, LMAND, MATOR

Les fichiers LMANC et LMAND ont été créés à partir du fichier général de base des formes acoustiques. Ce dernier est la mémorisation des matrices binaires correspondant aux sonagrammes des phonatomes. Nous ne rentrerons pas dans les détails sur la constitution du fichier des éléments, nous nous bornerons à dire que les éléments sont rangés par fréquence d'apparition, chacun d'eux étant sous forme d'une matrice de vingt colonnes et trois lignes soit 60 mots machine.

Le fichier " MATOR " contient les numéros d'ordre des phonatomes (" MATOR " étant rangé sous forme matricielle).

TABLES : "NUMER " et " IXCON "

1) " NUMER " : La table " NUMER " contient la suite des numéros d'ordre des phonatomes à éditer.

A chaque numéro correspond une forme dans le fichier " LMAND " ou " LMANC " situé sur le disque.

Le premier mot de " NUMER " indique le nombre de phonatomes à éditer.

2) " IXCON " : Cette table regroupe les mots de commande associés aux phonatomes à synthétiser.

Le reste des tableaux et variables offre peu d'intérêt.

STRUCTURE DU PROGRAMME " CREPH " (fig. 1)

Nous distinguerons trois parties :

1) Partie d'initialisation du programme qui n'est exécuté qu'une seule fois.

2) Phase où l'on donne la " main " à l'utilisateur, c'est-à-dire que l'on attend des ordres.

3) Exécution du traitement choisi et retour à la deuxième partie.

A partir du moment ou l'exécution du programme est lancée, on passe dans la première partie qui elle, ne doit pas être interrompue.

Elle consiste à initialiser les tableaux " NUMER " , "IXCON ", rechercher les adresses des fichiers, lire " MATOR " etc...

Ensuite le programme demande à l'utilisateur placé devant l'unité de visualisation d'entrer, par l'intermédiaire du clavier, une suite de phonèmes.

Le programme va se dérouter de lui même vers le module d'analyse de la phrase phonétique et de chargement de " NUMER " et " IXCON " puis se débrancher dans le module de visualisation des trois premiers phonatomes.

...../

CALINET . Janvier 1971

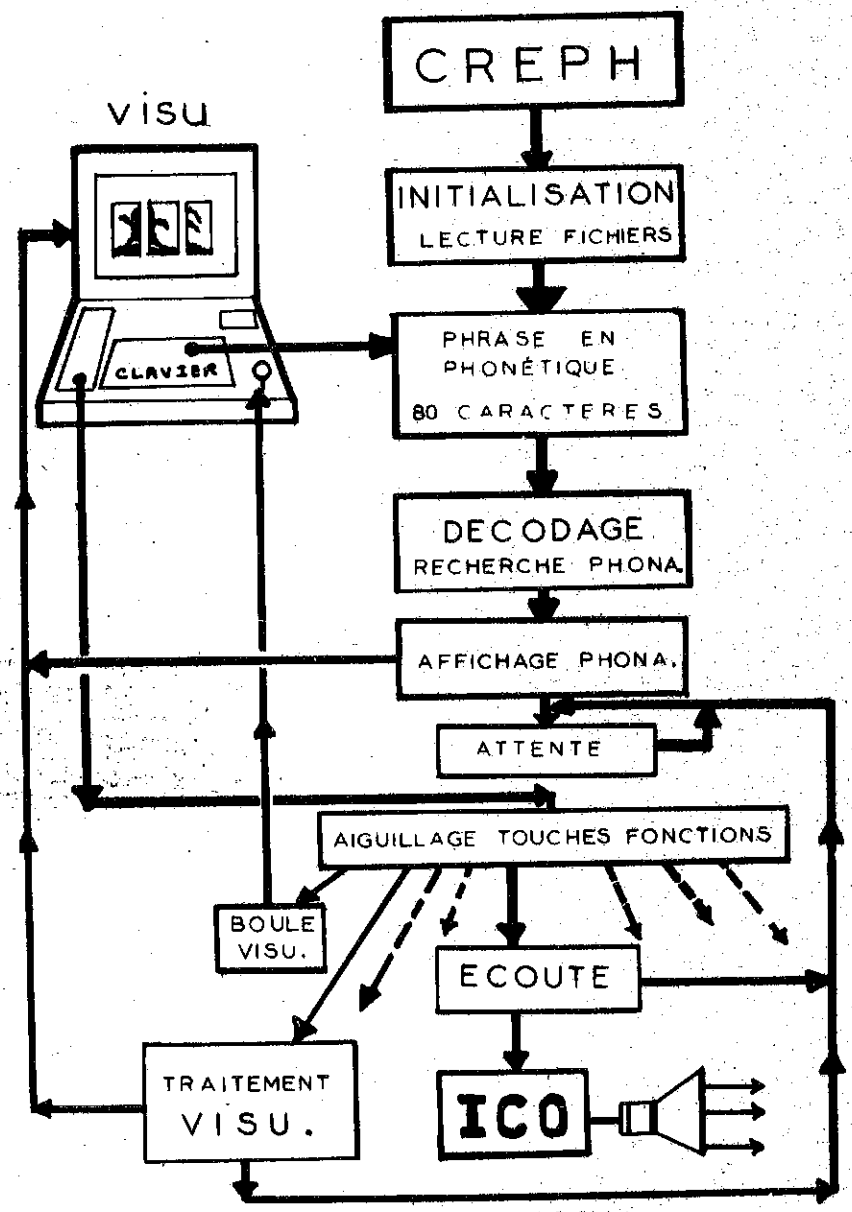


Fig 1.

ORGANIGRAMME GENERAL

du programme CREPH.

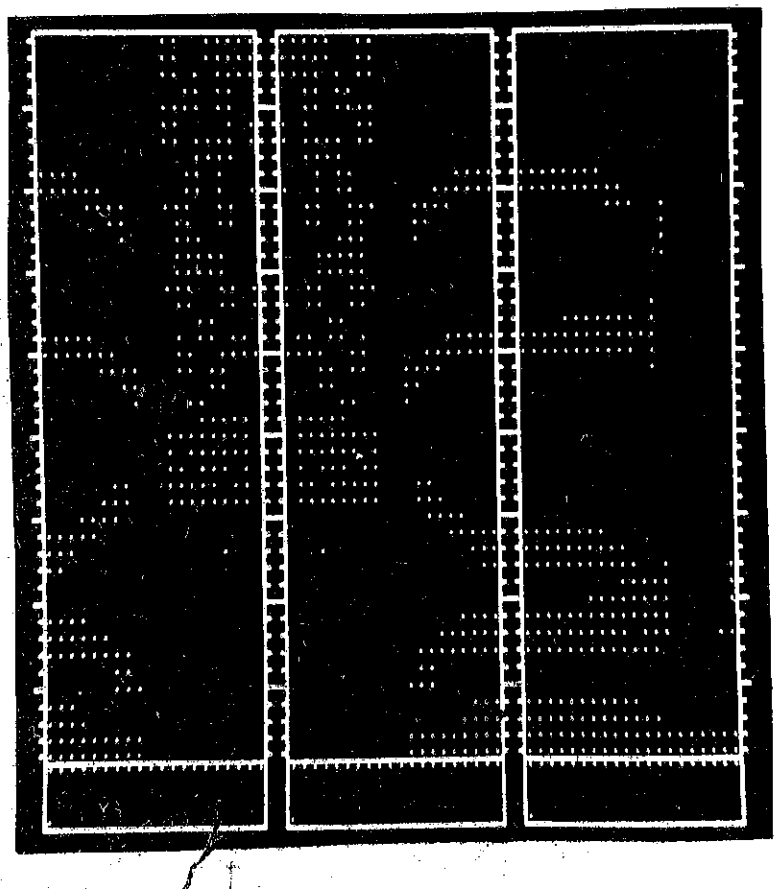


Fig 2.

AFFICHAGE DES PHONATOMES

AX - XA - AR

Visualisation par le module 2

Si l'utilisateur se sert à ce moment là des touches fonctions cela aura pour effet de produire une interruption qui ne pourra pas être traitée par le programme " CREPH " du fait que ce dernier n'est pas entré dans la deuxième partie.

Pour le moment l'utilisateur ne peut rien demander. Cette phase est cependant très rapide.

La deuxième partie est très courte. Elle consiste à appeler un sous-programme qui gerera l'interruption sur les touches de fonction. Une fois entrés dans ce sous-programme nous attendons l'interruption. Plus rien ne s'exécute (exceptée une seule instruction qui a pour effet de boucler sur elle-même). Cette phase est très intéressante car l'utilisateur a tout le temps devant lui pour réfléchir au traitement dont il demandera l'exécution.

L'utilisateur appuyant sur une touche débloquera l'attente. Le sous-programme traitant l'interruption, déterminera le numéro de la touche qu'il donnera à CREPH.

CREPH entre alors dans la troisième partie qu'il n'exécutera pas entièrement celle-ci étant constituée actuellement de 29 modules. En fonction du numéro N de la touche il ira exécuter le module N puis retournera au début de la deuxième partie.

Il est nécessaire de préciser que pendant l'exécution d'un module, l'utilisateur peut demander à exécuter un autre module. Cette action ayant pour effet d'arrêter l'exécution de ce dernier sans détruire le programme.

Nous n'expliquerons pas en détail tous les modules. Cela nous emmènerait trop loin. Nous ne parlerons que de la fonction de chaque module et pourquoi certains ont été faits alors qu'ils semblent inutiles.

Nous allons sur un exemple voir les différents traitements que nous pouvons faire et expliquer chacun des modules par lequel nous passons.

La phase d'initialisation étant faite, le programme se trouvant dans la deuxième partie, l'utilisateur commence à travailler.

Supposons que l'utilisateur désire créé un phonatome.

Sélection touche 16 .

CREPH se déroute dans le module 16 et demande à l'utilisateur le nom du phonatome ainsi que son numéro. En fonction de ces données CREPH modifie le tableau MATOR et réserve dans le fichier la place du phonatome qui est à l'état vide pour le moment. Le programme revient à l'état 2. L'utilisateur va pouvoir remplir le phonatome.

touche 1.

L'exécution du module 1 sélection le clavier de la vision ce qui permet à l'utilisateur d'entrer une phrase en phonétique puis de visualiser les trois premiers phonatomes (fig. 2).

...../

CREPH n'admet que des corrections sur le phonatome central. C'est pourquoi si l'utilisateur veut créer le phonatome de nom /J il devra taper par exemple B/JWR (bonjour).

Les phonatomes visualisés seront alors B/ /J JW ainsi /J est bien situé au centre.

touche 3 .

L'utilisateur peut maintenant remplir le phonatome. CREPH se trouve alors dans le module 3. A partir de cet instant il donne à l'utilisateur le contrôle de la touche fonction spéciale qui permet de déplacer un spot lumineux.

L'utilisateur dispose d'une touche de validation du point qu'il désigne sur l'écran. Le module 3 attend donc la validation, il récupère les coordonnées du point et vient ranger en mémoire dans la matrice représentant le phonatome le point, c'est-à-dire que l'on met à 1 le bit correspondant se trouvant à l'intersection de la fréquence déterminée par le y et l'évènement par x .

Au même instant le caractère 1 s'allume à l'endroit où le point a été validé. Ce qui permet à l'utilisateur de savoir les points qu'il vient d'entrer.

L'utilisateur dispose également de 17 touches fonction lui permettant de remplir ou d'effacer des points, des blocs, des colonnes, des lignes, du pont de phonatomes. Ceci étant très rapide et très souple. Je ne rentrerai pas dans les détails sur la manière dont on a fait ces différents traitements qui se passent purement au niveau machine. Retenons simplement que le jeu de ces touches joue le même rôle qu'une gomme et un crayon.

L'utilisateur a rempli le phonatome qui n'est pas encore parfait et veut l'écouter.

touche 4 .

Le module 4 prend le phonatome en mémoire centrale, le ramène de sa forme matricielle 20, 48 en 20,3 V , la copie sur le disque dans le fichier LMAND à une place déterminée et se débranche dans le S/P de sortie sur icophone en lui transmettant les paramètres nécessaires à la sortie.

Puis retour à l'état (2)

l'utilisateur peut redemander l'écoute autant de fois qu'il le désire en appuyant sur la touche 4.

L'utilisateur peut être amené à retoucher le phonatome. Il est évident que s'il modifie le phonatome en mémoire il ne pourra plus le comparer du point de vue écoute avec le phonatome nouvellement corrigé.

...../

Pour cela, l'utilisateur, par la touche 25 demande à CREPH de copier dans le fichier à l'endroit qu'il avait précédemment réservé, le phonatome.

Le module 25 a donc pour fonction de copier le phonatome et de redonner le contrôle à l'utilisateur.

Ce dernier va donc pouvoir appeler à nouveau le phonatome par la touche 1, le corriger puis l'écouter par la touche 4

Touche 28 .

L'utilisateur écoute le phonatome ancien rangé sur disque. Le module 28 charge dans NUMER le numéro du phonatome puis se débranche dans le module 4 . On sait déjà ici la notion qu'un module peut déclancher l'exécution d'un autre module bien que chacun d'eux soit indépendant, excepté le module 31 qui ne peut s'exécuter que si le module 27 a été exécuté.

Si l'utilisateur juge que la nouvelle version est bonne il sélectionne la touche 25 dont le module ira écraser l'ancien phonatome.

Maintenant l'utilisateur désire écouter le phonatome non plus isolé, mais dans l'ensemble d'une phrase.

Sélection de la touche 24 .

L'utilisateur a le contrôle du clavier. Il entre une phrase en phonétique puis valide la touche fin de zone. Le module 24 prend la phrase qui se trouve dans la mémoire tampon puis analyse cette phrase, crée NUMER et IXCON , se débranche ensuite dans le S/P de sortie sur icophone. Dès que le remplissage du BUFFER est terminé, ce qui ne veut pas dire que l'icophone est en état statique, le contrôle est donné à l'utilisateur toujours en module 24. Ceci est intéressant par le fait que, pendant la sortie sur icophone, l'utilisateur peut entrer à nouveau une phrase. Signalons qu'une clé au pupitre de l'IBM 1130 permet de boucler sur la sortie de l'icophone.

L'utilisateur trouve que le phonatome n'est pas parfait. Il demande alors la visualisation des phonatomes issus de la phrase.

Touche 27.

Visualisation de la phrase (fig.3 et fig.4). Cet affichage est composé de la manière suivante.

Si nous avons 15 ou moins de 15 phonatomes, ils sont affichés sur une seule bande.

Sinon sur deux bandes, la lecture se faisant de la gauche vers la droite et de haut en bas. Un cadre est tracé pour délimiter chaque phonatome.

L'utilisateur a la possibilité de faire un effet de ZOOM sur le phonatome choisi ceci par la touche 31.

Le module 31 va pouvoir s'exécuter du fait que l'on sort du module 27 étant repassé par la phrase 2 d'attente.

L'utilisateur a le contrôle du spot lumineux qu'il déplace sur le phonatome, puis valide le point.Le phonatome va alors être visualisé accompagné du précédent et du suivant dans un format plus grand pour permettre les corrections.

L'utilisateur peut donc recommencer ce processus autant de fois qu'il le désire.

Signalons que la touche 2 permet de visualiser les 3 phonatomes dont les numéros se trouve dans NUMER (si NUMER ne contient qu'un seul phonatome il sera affiché accompagné de deux autres qui seront vides).

Cette utilisation de touches semble inutile du fait que l'on a déjà les éléments visualisés. Cependant, si nous précisons qu'au cours de la correction d'un phonatome, à chaque validation du point, s'allume le caractère 0 si le mode est "effacement" ou le caractère 1 si le mode est "lecture" l'utilisateur ne reconnaîtra plus grand chose du phonatome au bout de quelques instants. Ainsi par ce module, il visualise le phonatome à l'état actuel mais avec des allumage de petits points.

Touche 29.

Si nous rappellons qu'une clé du pupitre 1130 permet de boucler sur la sortie de l'icophone on comprend l'intérêt d'une touche fonction permettant de relancer l'écoute si l'utilisateur oublie la clé du pupitre. Ainsi s'il doit écouter plusieurs fois une phrase et s'il oublie de baisser la clé il ne sera pas obligé de recommencer la frappe d'une phrase.

Le programme CREPH présente donc une grande souplesse dans la création et la correction des éléments. Ceci est dû à sa structure modulaire, et aussi au fait que, lorsque l'utilisateur demande un traitement, il est servi dans les secondes qui suivent - puis CREPH attend de nouveau - ceci laisse du temps à l'utilisateur pour réfléchir.

...../

CALINET. Janvier 1971

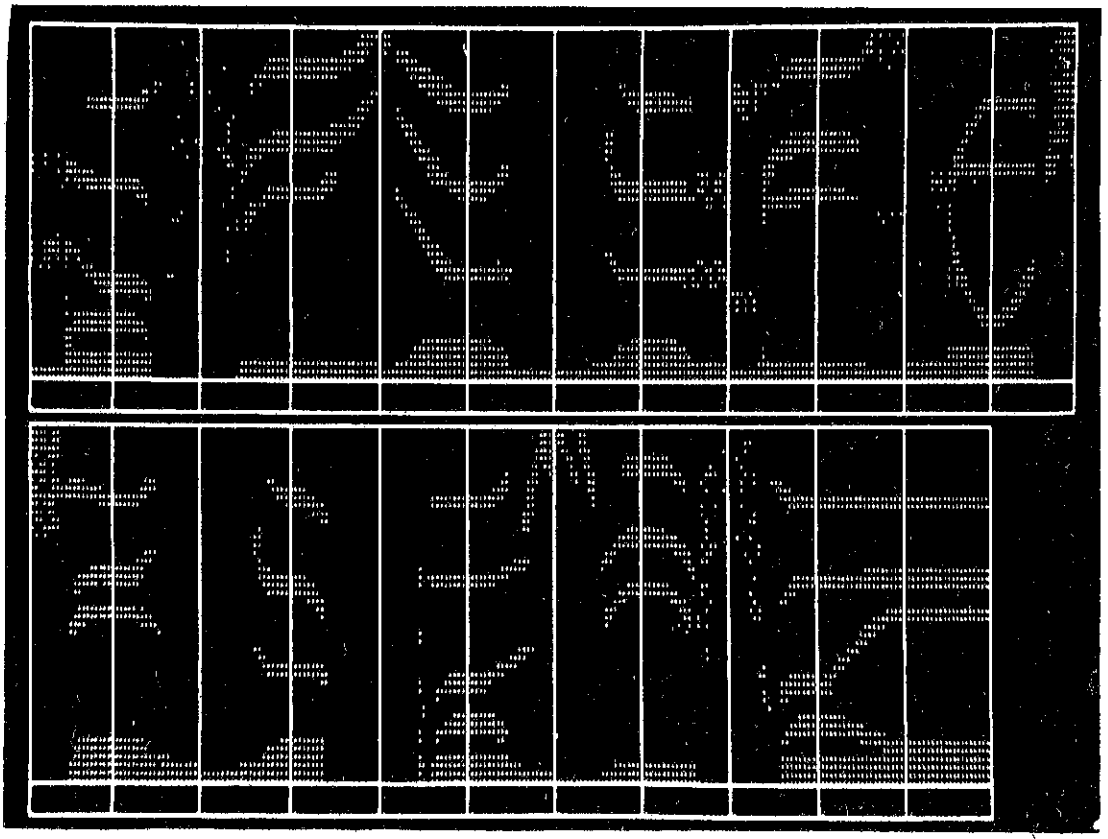


Fig 3.

VISUALISATION DE LA PHRASE :

La fille de Minos et de Pasiphae .

Module 27

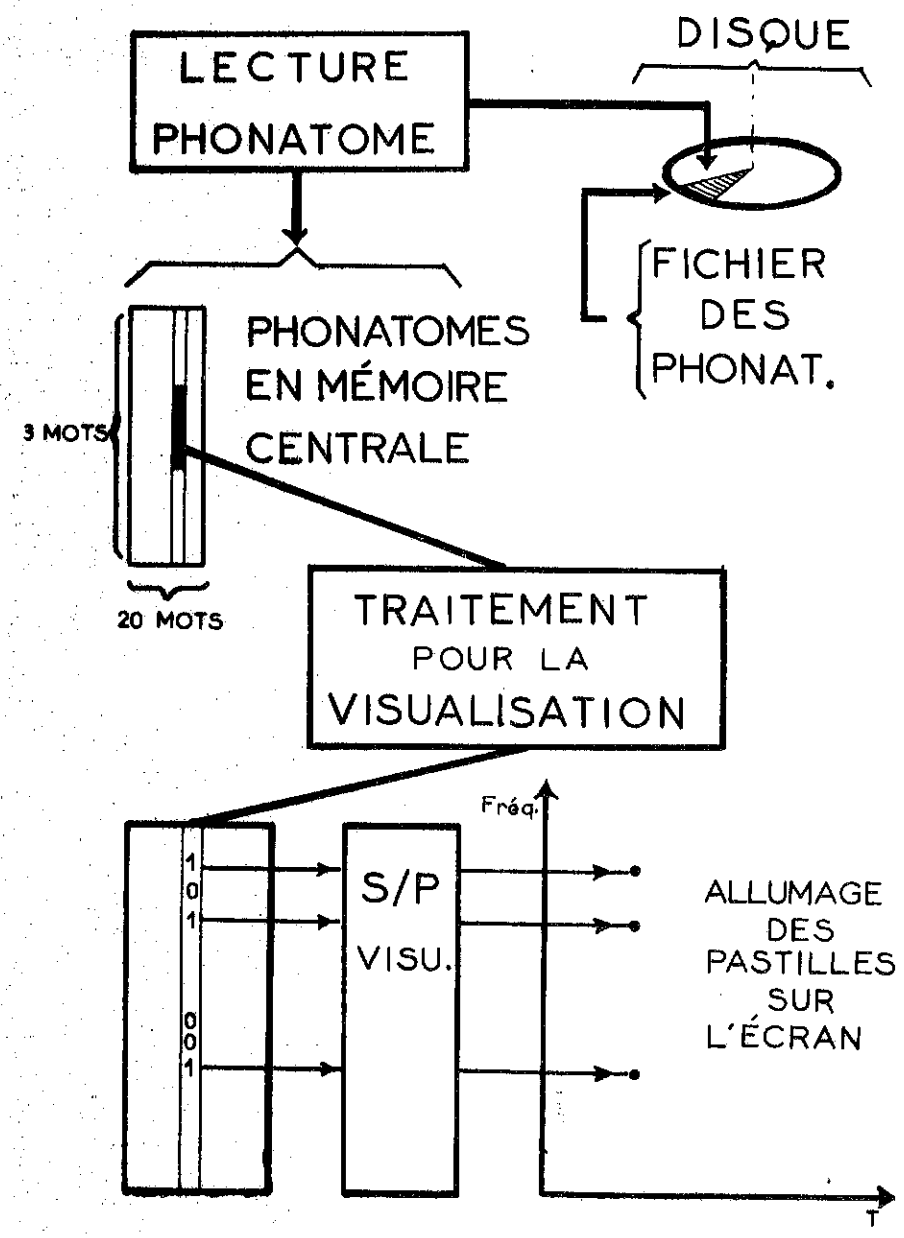


Fig 4.

ORDINOGRAMME du module de visualisation du phonatome

Voici quelques intérêts de la structure modulaire du CREPH.

- Si un module n'a plus de raison d'exister on supprime le lien qui existe entre la 2ème partie de CREPH et ce module.

- des touches fonction de la visu peuvent tomber en panne. Il suffit de changer le numéro du module et de le remplacer par le numéro d'une touche disponible à condition que celle-ci soit valide.

- A N A M O -

Le programme ANAMO permet d'étudier l'intonation avec possibilité d'anamorphoser les éléments phonétiques.

- Bien que les fonctions d'ANAMO soient différentes de CREPH nous retrouvons la même installation et la même structure de programme. Cependant un seul périphérique diffère et c'est le plus important : l'ICOPHONE IV à commande numérique. Signalons que le programme utilise un traceur de courbe CALCOMP (fig.5).

Nous prendrons, comme dans l'exposé sur CREPH, un exemple d'utilisation du programme montrant le déroulement de celui-ci avec ses différentes phases.

- Avant cela donnons quelques précisions sur les éléments de base du programme.

- L'unité de visualisation a le même rôle que dans CREPH, c'est-à-dire que tous les ordres partent de l'unité de visualisation et qu'une partie des résultats est affichée par l'écran.

- Tables utilisées par ANAMO.

  - NUMER : même signification que dans CREPH

  - IXCON : table des mots de contrôle diffère de celle de CREPH. Sa structure sera exposée plus loin.

  - IXFM et IYFM : Ces deux tables sont destinées à garder les coordonnées de chaque point de la courbe de modulation de fréquence.

- Fichiers

  - MATOR)
  - KOD (voir CREPH
)

  - LMM4 Fichier des éléments phonétiques pour l'ICOPHONE IV (équivalent de LNAMAND pour CREPH)

  - LMIND Fichier de travail. Contient les phonatomes pour la phrase en cours. LMIND possède un double en mémoire centrale ce qui permet une souplesse d'utilisation importante dans le traitement de l'anamorphose.

..../

ANAMO .

Fig 5.

ORGANIGRAMME GENERAL

du programme ANAMO .

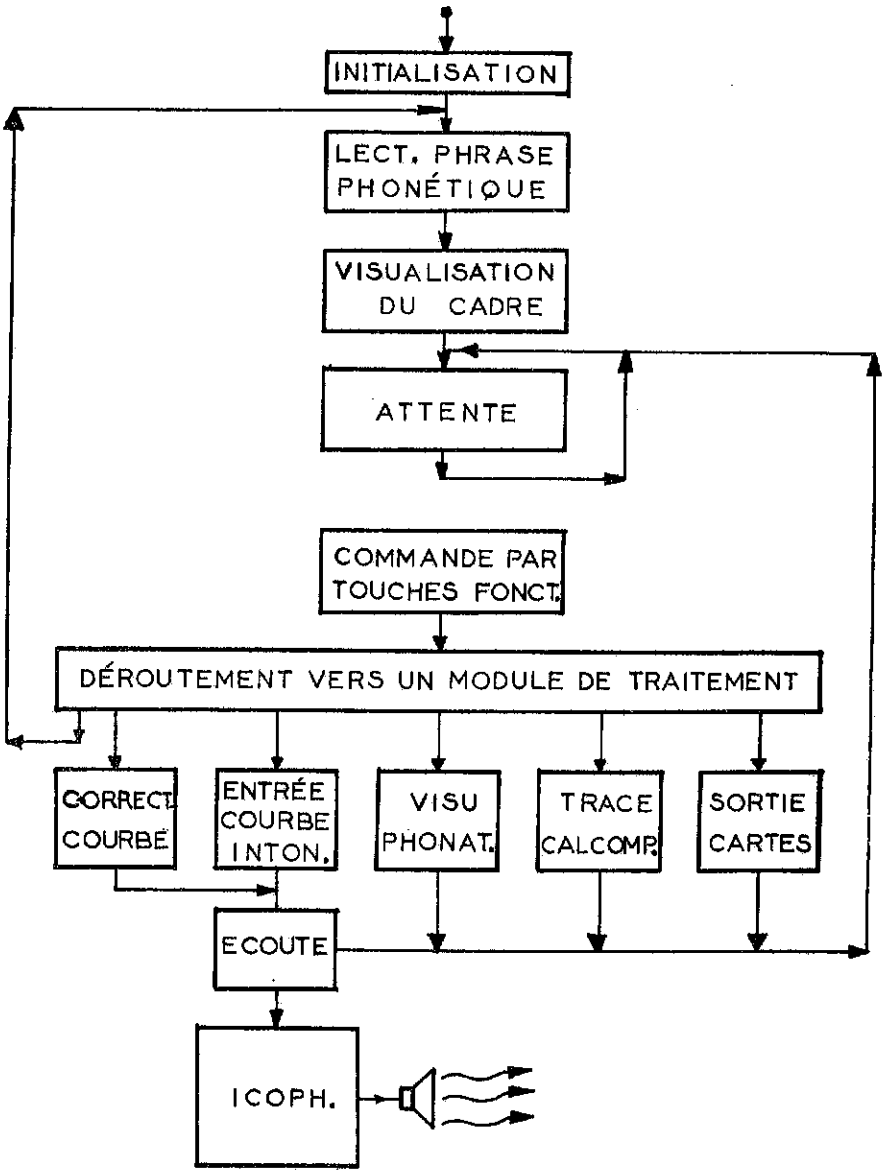
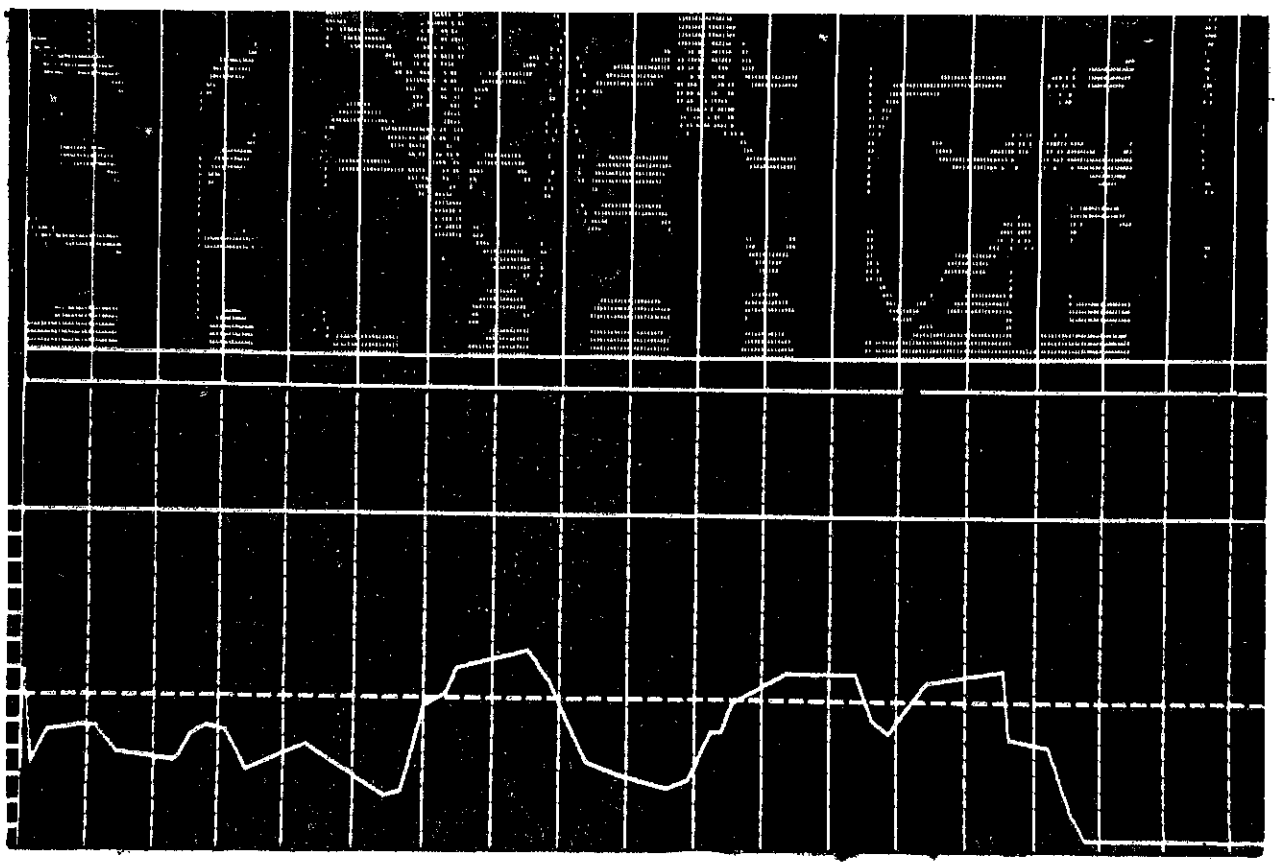


Fig 6.

VISUALISATION de la phrase :

"Le petit chat fait sa toilette".



- Données pour le programme.

L'utilisateur a la possibilité d'entrer une phrase en phonétique, soit en utilisant le lecteur de cartes, soit en utilisant le clavier de l'unité de visualisation.

S'il n'y a pas de cartes à lire au lecteur, le programme ANAMO donne automatiquement le contrôle du clavier à l'utilisateur. Dans le cas contraire le clavier est vérouillé et les cartes lues.

- Supposons que l'utilisateur entre une phrase par l'un des deux moyens. ANAMO appelle le sous-programme PHONU (commun à CREPH) qui analyse la phrase et crée la table NUMER et une partie de IXCON. A ce niveau la table IXCON contient dans chaque mot le niveau sonore et le temps d'édition par évènement d'un phonatome.

Commence ensuite la phase de lecture des phonatomes dans le fichier LMM4. Chaque phonatome lu est rangé à la fois sur disque dans le fichier LMIND et gardé en mémoire dans le double du fichier LMIND. Le double de ce fichier permet d'anamorphoser les phonatomes pour une phrase donnée autant de fois que l'utilisateur le désire. Nous verrons plus bas le principe de l'anamorphose fréquentielle.

- L'utilisateur jusqu'à présent n'a pas encore donné d'ordres. Il doit attendre la visualisation d'un cadre sur l'écran de la visu, qui lui servira de "support" pour tracer sa courbe d'intonation.

- Le cadre est affiché sur la partie basse de l'écran. L'autre moitié étant réservée pour la visualisation des phonatomes (fig.6).

En ordonnées, nous trouvons les intervalles de fréquences, gradués de 0 à 15 ce qui nous fait 16 échelons. En abcisses le temps. A chaque frontière de phonatome est tracé verticalement une limite en pointillé. Un axe horizontal est tracé pour indiquer le fo (voir exposé de Monsieur SAPALY).

Dès la fin du tracé de ce cadre l'utilisateur a le contrôle des touches fonction.

Le programme ANAMO entre dans une phase identique à la deuxième phase de CREPH.

- Visualisation des phonatomes

La touche fonction 2 déclenche l'exécution du module 2 qui appelle un sous-programme utilisé par CREPH pour visualiser les phonatomes. Ici les phonatomes sont toujours affichés sur une seule bande, selon un format variable dans le Temps en fonction du nombre de phonatomes.

- L'utilisateur en sélectionnant la touche 3 va pouvoir entrer la courbe de modulation de fréquence.

...../

Le module 3 sélectionne la touche fonction spéciale permettant à l'utilisateur de déplacer un spot sur l'écran. Ce dernier, à l'aide du spot, valide les joints les uns après les autres. A chaque validation les coordonnées sont analysées pour déterminer si le point est hors du cadre ou non. Si c'est le cas un message est imprimé au pupitre du 1130 et le contrôle de la touche redonné. Sinon les coordonnées sont enregistrées dans les tables IXFM et IYFM et un vecteur est tracé du point précédent ou de l'origine, si c'est le premier point, au point validé.

L'utilisateur construit sa courbe point par point en ayant à chaque validation une visualisation.

- Le Module 4 est appelé pour indiquer au programme que la courbe est finie. Le dernier vecteur est tracé et une fin de zone est placée dans les tables IXFM et IYFM. Puis commence le calcul des mots de contrôle.

- Principe du Calcul.

La table IXFM est parcourue de la gauche vers la droite et par pas de deux. C'est-à-dire que son index évolue toujours de la valeur N à la valeur N + 1 puis de N + 1 à N + 2 ainsi de suite. Ainsi à chaque exploration nous récupérons les abscisses des extrémités de chaque segment de droite. Le balayage est le même pour la table IYFM.

A chaque exploration sont réalisés :

1) le calcul de l'équation de la droite

2) l'équation aux abscisses entre cette droite et tous les évènements compris entre X1 et X2 . Donc à chaque intersection nous déterminons une valeur de Y qui nous permet, à l'aide d'un calcul intermédiaire, de déterminer la commande de fréquence

Nous venons ensuite placer cette valeur dans le mot de contrôle associé à l'évènement. A la fin du balayage de la table IXFM la table IXCON est chargée et prête.

Le sous-programme de sortie sur ICOPHONE 4 est ensuite appelé. Un test de clé au pupitre indique si le programme doit boucler sur la sortie des phonatomes ou non. L'utilisateur a aussi à sa disposition une touche fonction pour relancer l'écoute.

- Correction de la courbe.

Trois touches fonction sont à la disposition de l'utilisateur pour modifier la courbe.

Touche 11 : Correction d'une portion de courbe ou de plusieurs portions.

" 15 : Correction à partir du début de la courbe.

" 12 : Fin de correction.

La fonction du module 11 est de venir insérer les coordonnées des points entrées dans les tables IXFM et IYFM tandis que le module 15 remplace ces mêmes

...../

coordonnées à partir du début des tables jusqu'au moment où l'on arrête la correction.

- Deux modules ont pour fonction de conserver l'état des travaux de l'utilisateur.

  - Module 9 : Sortie sur traceur de la courbe et du cadre avec écriture de la phrase. Ceci est un document.

  - Module 13 : Sortie des tables IXFM et IYFM sur cartes avec éventuellement la phrase.

  - Module 14 : L'utilisateur peut introduire à nouveau en machine les travaux précédents par l'intermédiaire des cartes perforées par le module 13.

    Le module 14 recharge à partir de ces données les tables IXFM et IYFM, affiche le cadre, les phonatomes, calcule les mots de contrôle et effectue la sortie sur icophone. Tous les modules faisant ces tâches sont exécutés par le module 14. A ce niveau ce n'est plus l'utilisateur qui en demande une exécution séparée.

Anamorphose fréquentielle :

Le module 6 peut être exécuté de deux façons :

1) L'utilisateur appelle le module pour anamorphoser les phonatomes et affecte à chaque phonatome un coefficient d'anamorphose.

2) C'est le programme qui de lui-même va anamorphoser les phonatomes en calculant les coefficients à partir de la courbe de modulation de fréquence et du réglage de l'icophone.

Le principe de l'anamorphose est le suivant. Il consiste à dilater ou à comprimer les événements un par un en fréquence. Ceci en fonction d'un coefficient qui peut varier entre 0,6 et 1,7.

Cas 1 :

- L'utilisateur par la touche fonction 6 demande à anamorphoser lesphonatomes qui se trouvent dans le double du fichier LMIND. L'état d'une clé au pupitre indique au module s'il doit demander ou non à l'utilisateur un coefficient. Si ce n'est pas le cas le coefficient reste à sa dernière valeur.

  A partir de NUMER le module va chercher un phonatome dans le double de LMIND, l'isole dans un tableau intermédiaire, et l'anamorphose évènement par évènement. Le coefficient d'anamorphose est le même pour les vingt évènements. Le phonatome anamorphosé est ensuite recopié dans LMIND sur disque. Quand la table NUMER a été explorée jusqu'à la fin, le contrôle est à nouveau donné à l'utilisateur.

...../

Cas 2 :

- A la fin du tracé de la courbe de modulation de fréquence, suivant l'état d'une lcé au pupitre, le module quatre appelle le module six, qui avant de commencer le traitement calcule les coefficients d'anamorphose pour chaque évènement de chaque phonatome. Dans ce cas l'anamorphose est automatique. A la fin de cette tâche même débranchement que dans le cas 1.

Le programme ANAMO est donc un outil de travail intéressant pour l'étude de l'intonation avec anamorphose par sa simplicité d'utilisation et ses capacités de traitement.

M. MLOUKA

CODAGE ET ANALYSE SPECTRALE DE LA PAROLE EN VUE DE LA RECONNAISSANCE AUTOMATIQUE

JANVIER 1971 N°=53

GAM

BULLETIN DU GROUPE d'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES . TOUR 66 . Place Jussieu - PARIS 5°

Les exposés qui précèdent ont montré comment il était possible de générer une parole synthétique intelligible par la méthode de digrammes phonétiques, préalablement quantifiés et numérisés, correspondant aux phonatomes ou diphonèmes. On part d'une frappé sur clavier en phonétique ou en littéral.

Le problème que nous allons aborder à présent consista à étudier la possibilité de commander automatiquement cette frappe à partir de la parole normale d'un locuteur quelconque. Celui-ci parle devant un microphone et on l'enregistre sur bande magnétique. Cette bande est alors traitée, car il s'agit de faire entrer les formes sémantiques particulières à ce locuteur (donc anamorphosés) dans le cadre normalisé que nous avons adopté pour la synthèse.

La manipulation consiste à simuler d'abord en machine ce que fait le sonagraphe, c'est-à-dire réaliser une analyss fréquence-temps. Cette manipulation se décompose de la manière suivante :

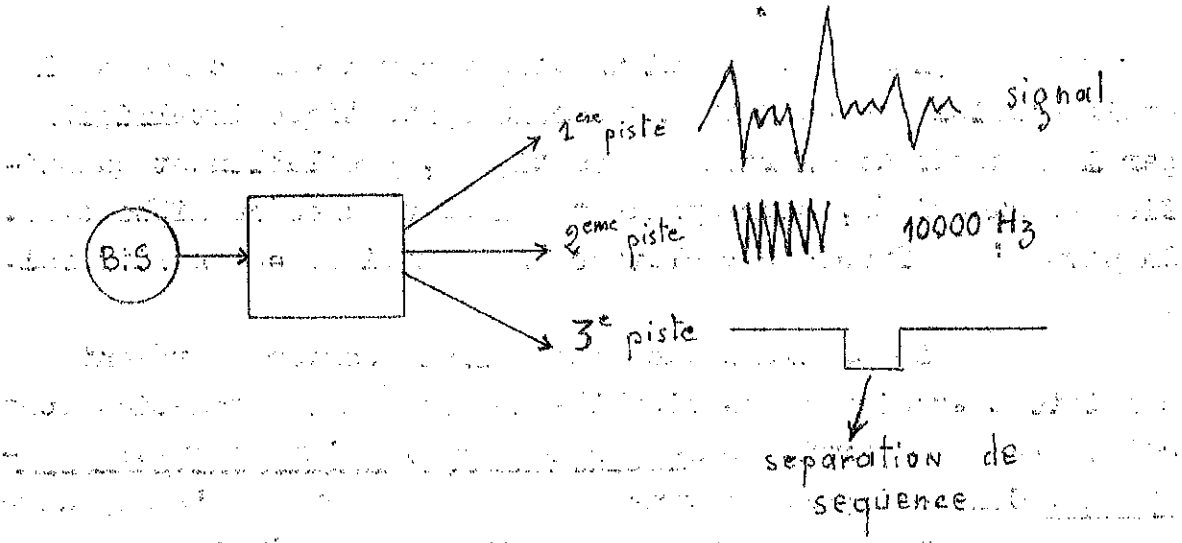
1) Entrée de la parole en machine :

a) Enregistrements sonores :

Nous avons enregistré sur bandes sonores des séquences de parole (mots, phrases, phonatomes) pour 8 locuteurs différents, hommes et femmes. Toutes ces séquences ont été regroupées sur une seule bande.

Celle-ci a été transférée sur une bande "AMPEX" à modulation de fréquence sur laquelle nous trouvons une piste pour le signal proprement dit, une piste avec un signal 10.000 Hz (pour le futur échantillonage) et un signal en créneaux indiquant les séparations de séquences.

.../...



b) Conversion analogique-digitale :

A partir de cette bande, nous nous sommes adressés au G.R.I. (Groupe de Recherche Ionosphérique - C.N.R.S.) afin d'obtenir la conversion de ces séquences en numérique.

Cette conversion a soulevé plusieurs difficultés d'ordre pratique (magnétophone en panne, séquences détruites, etc...). Finalement l'opération a été réalisée grâce à un convertisseur analogique digital monté sur un calculateur P.D.P. 5. Cette conversion nous a fourni trois bandes numériques 7 pistes, le signal étant digitalisé avec des mots de 12 bits. Nous disposons donc en machine d'échantillons dont l'amplitude est comprise entre 0 et 4096, ceci à la cadence de 10.000 par seconde.

.../...

c) L'entrée proprement dite en machine :

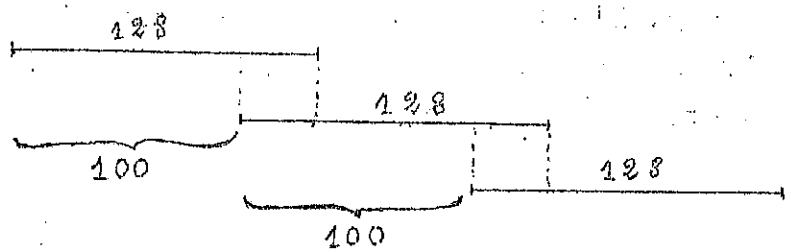
Le traitement de ces informations devant se faire au C.I.R.C.E. (Centre de Calcul Numérique du C.N.R.S. à Orsay), le premier problème qui s'est posé a été la compatibilité des informations avec l'ordinateur du C.I.R.C.E. (IBM 360-75). En effet, ce calculateur travaille habituellement avec des bandes 9 pistes sur des mots de 32 bits.

Ces factuurs nous ont obligés à écrire des programmes de conversion. Nous n'entrerons pas dans les détails de ces manipulations qui n'offrent aucun intérêt théorique pour notre problème.

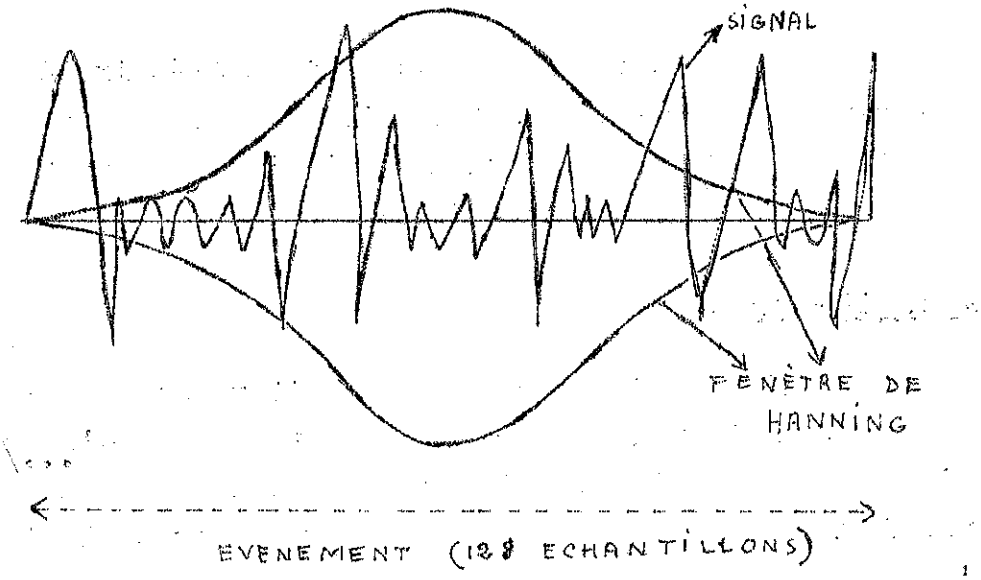
2) Le traitement :

L'échantillonnage a été fait toutes les 100 microsecondes. Il faut signaler la quantité énorme d'informations à traiter : pour les 20 minutes de parole que nous avons enregistrées, nous nous sommes retrouvés avec 12 millions d'échantillons.

Nous avons voulu reprendre la notion d'"évènement" utilisée en synthèse. Autrement dit, le signal est analysé par tranches de 10 millisecondes (soit 100 échantillons) au moyen d'une transformation rapide de Fourier (FFT). Mais les programmes performants pour cette analyse imposent un nombre d'échantillons égal à une puissance de 2. Aussi avons-nous pris des prélèvements de 128 échantillons en avançant de 100 en 100.



Le résultat de ce traitement nous a révélé des spectres paraxites provenant d'un découpage trop abrupt du signal au début et à la fin de chaque prélèvement. Ces chocs ont été estompés en multipliant le signal par une fonction de prélèvement classique ; la fenêtre de Hanning, représentée sur la figure suivante :

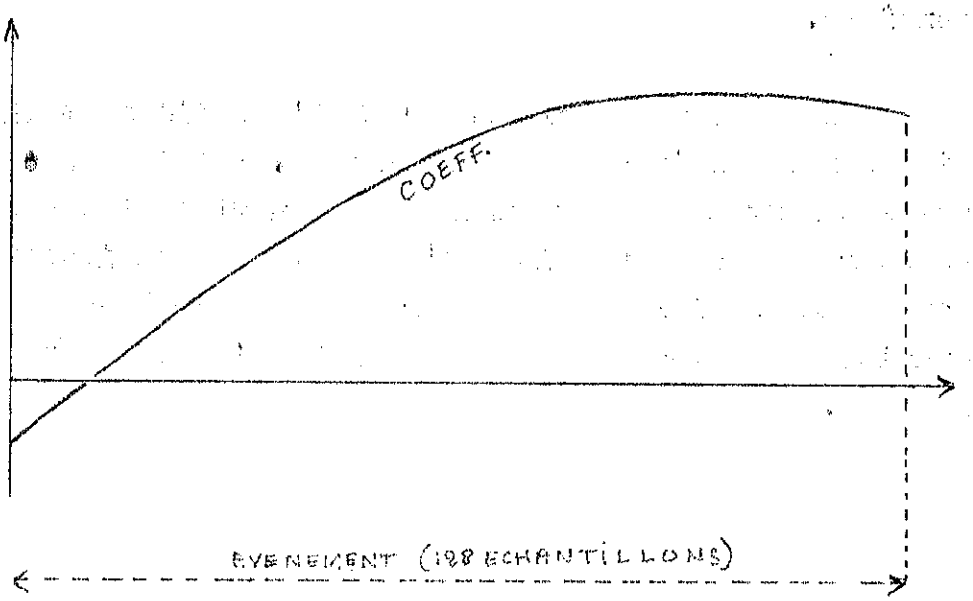


Les valeurs spectrales résultant de l'analyse ont été ramenées dans une échelle logarithmique de 0 à 9, chaque unité de l'échelle correspondant à un niveau, une différence de niveau égale à 1 correspondant à 5 décibels. A partir de cela, nous avons mis au point un programme d'édition sur papier-imprimante. Les valeurs spectrales correspondant à un "évènement" temporel ont été éditées sur deux colonnes, l'une chiffrée de 0 à 9, l'autre composée uniquement d'astérisques pour les niveaux forts. Sur les exemples donnés ci-dessous, le seuil a été fixé au 3^{e} niveau.

.../...

Le sonagramme obtenu n'était pas parfait : les niveaux correspondant aux fréquences moyennes et aiguës de la parole n'étaient pas assez marqués. De plus, on entrevoit les formants, mais il ne sont pas assez compacts.

Pour remédier à ces inconvénients, nous avons fait plusieurs essais. Nous avons retenu un réajustement des niveaux en fonction des fréquences, en multipliant par des coëfficients ayant l'allure de la figure ci-dessous et qui reflètent les résultats classiques de l'énergie de la parole en fonction de la fréquence.



Les sonagrammes ainsi obtenus étaient plus encourageants. (voir photo No. 1) Néanmoins, les valeurs spectrales restaient trop disperséos, et nous ne retrouvions que difficilement les formes sémantiques de la parole.

C'est pourquoi nous avons appliqué un lissage sur les valeurs spectrales, ce lissage consistant à affecter à à un point spectral la valeur moyenne obtenue aux points connexes.

.../...

Exemple : pour le point X nous prendrons comme valeur la moyenne de X et des points qui l'entourent.

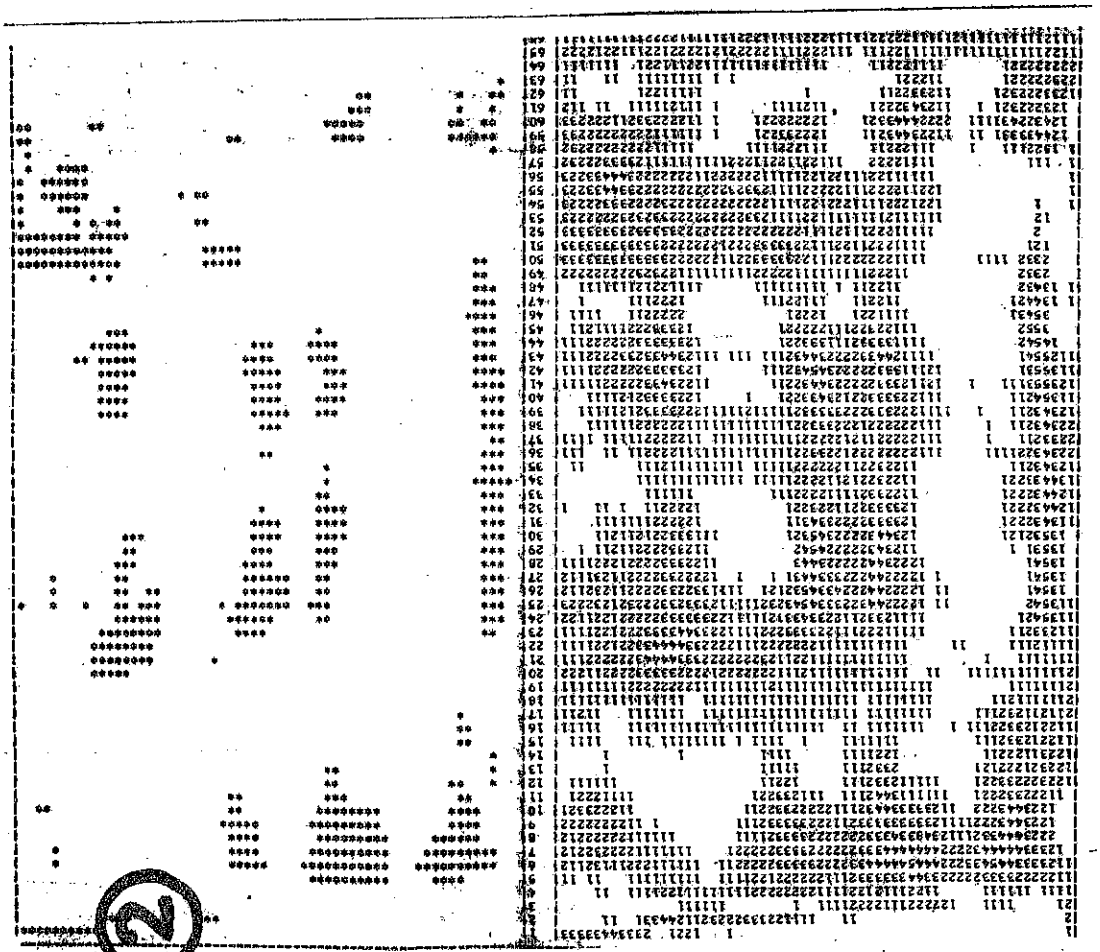
$$\begin{matrix} A_1 & A_2 & A_3 \\ A_4 & X & A_5 \\ A_6 & A_7 & A_8 \end{matrix} \qquad X \Longrightarrow \frac{X + \sum_{i=1}^{i=8} A_i}{9}$$

Ces traitements nous donnent des sonagrammes à peu près satisfaisants (voir photo No. 2).

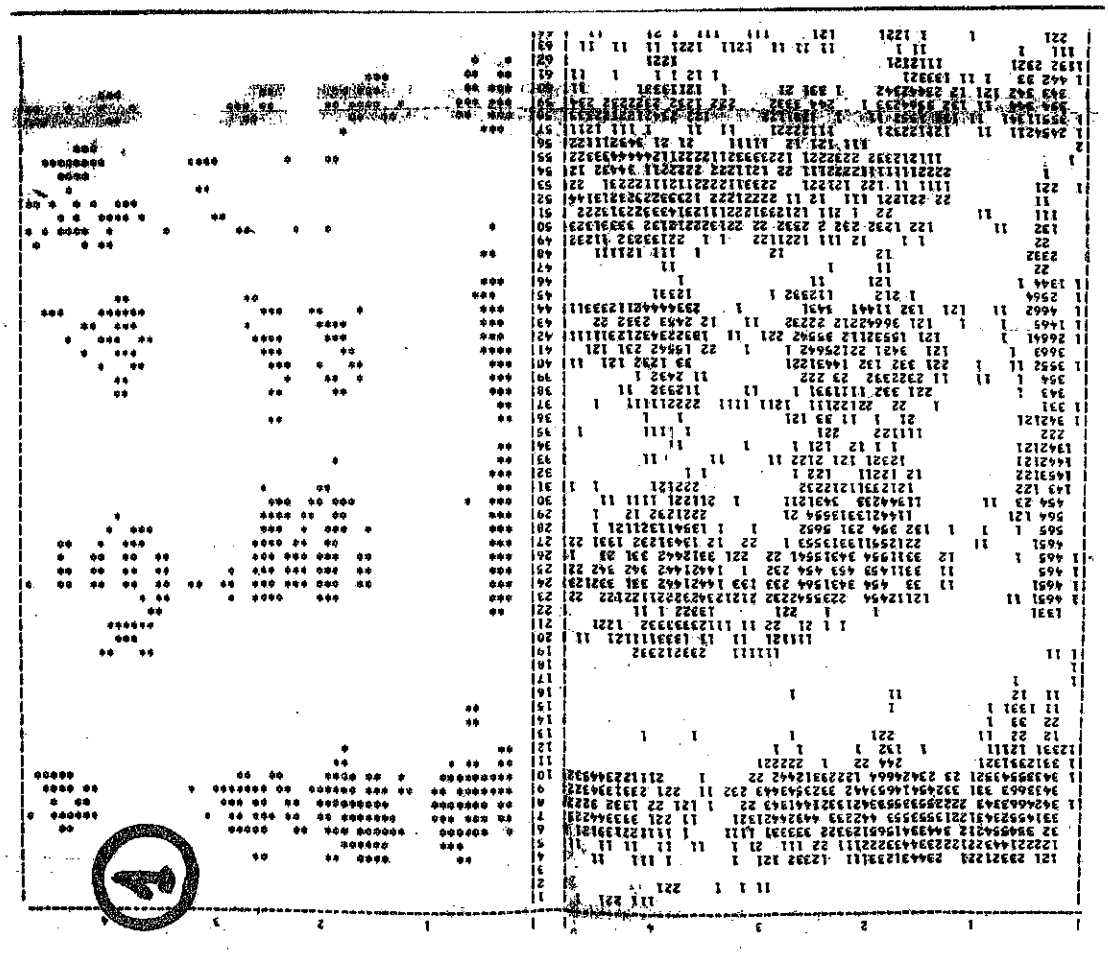
Il est clair que l'étude qui précède est bien la suite logique des travaux que nous avons effectués en synthèse. Ceux-ci nous ont appris à connaître les formes sémantiques de la parole ; nous essayons à présent de les reconnaître.

Nous avons encore beaucoup d'expérimentations à faire sur le traitement de ces images. Les images obtenues semblent cependant suffisamment nettes et démonstratives pour que nous puissions d'ores et déjà envisager d'aborder la première phase du problème général de la reconnaissance automatique de la parole d'un locuteur quelconque.

=:=:=:=:=:=:=



AS - TU VU CE...



AS - TU VU CE...