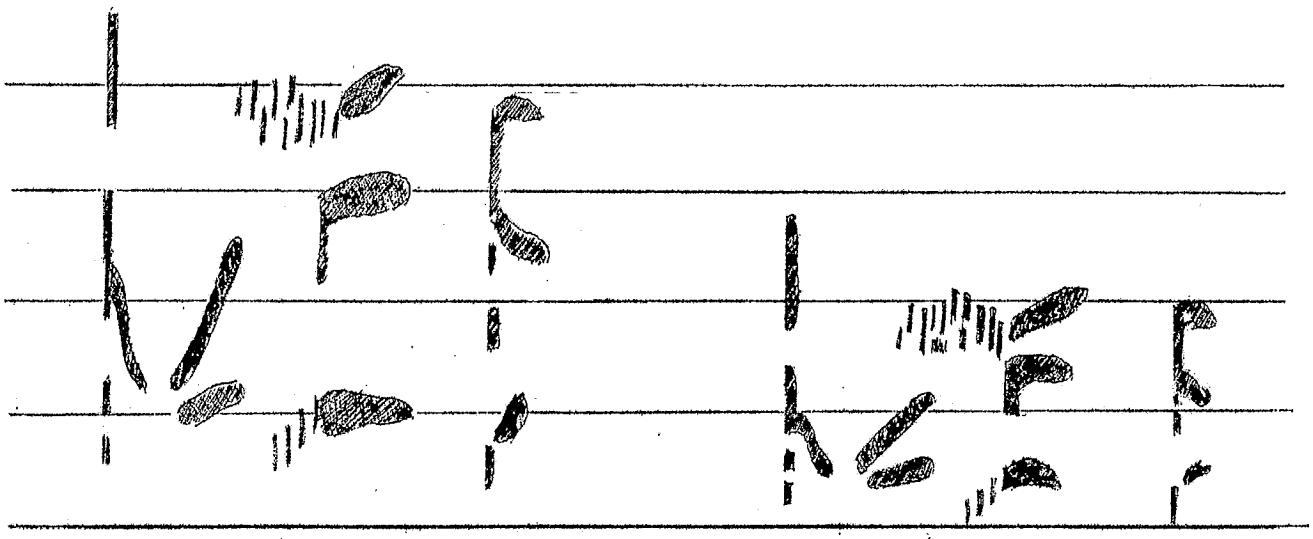
E. LEIPP

LE PROBLÈME DE L'INTELLIGIBILITÉ DE LA PAROLE

NOVEMBRE 1968 N° 37



GAM

BULLETIN DU GROUPE D'ACOUSTIQUE MUSICALE
FACULTÉ DES SCIENCES _ 8 RUE CUVIER _ PARIS 5°

G. A. M.
Groupe d'Acoustique Musicale
Laboratoire d'Acoustique
Faculté des Sciences
8, Rue Cuvier PARIS 5°

Adresse postale
9, Quai St Bernard 5°

Paris, le 14 Novembre 1968

BULLETIN N° 37

REUNION DU 8 Novembre 1968

Etaient présents

M. le Vice Doyen L. GAUTHIER qui nous avait honorés de sa présence
M. le Professeur SIESTRUNCK, Président
M. LEIPP, Secrétaire général. Melle CASTELLENGO, secrétaire

Puis, par ordre d'arrivée

M. J.S. LIENARD (Ingénieur Arts et Métiers); M. LELEUX (Ingénieur Central); Melle J. BARILLON (CAEN; SCHOLA CANTORUM); M. TRAN VAN KHE (Maître de recherche C.N.R.S.); M. LEGUY (Ecole Télécomm.); Melle Fr. LEIPP (Etudiante orthophonie); M. TEIL (CNRS); M. QUINIO (CCA, CNRS); M. SAPALY (Maître de Conférences Fac. Sciences); Melle M. GUERSTEIN (Informaticienne); Melle M.C. POUBLAN, Melle HENRION; Melle M.F. DENIS; M. FEUILLET(de la Schola Cantorum); M. LOPPION et M. ROUBAND (Ingénieurs Militaires CPE); M. FORET (Compositeur); M. MARION; M. FOURCIN (University College London); Melle M.C. LESIEUR; M. CONDAMINES (Labo.Acoustique ORTF); M. DETTON (cybernétique); M. MAILLARD (Musicologue); M. MERON (Etudiant son ENPTC); Dr VEIT et Mme BIZAGUET (Labo. Correction auditive); M. GRANGE et M. DUPREY (RAUC); Akira TAMBA (Compositeur, CNRS); Mme et M. CLEAVER (tambourinaire); Mlle Sylvie HUE (Prof. musique); M. PIMONOV (Maître de Recherche CNRS); M. MOULIN (CPE); M. VILLARD; Mme CAPETTE, M. TAVERNIER (DRME); M. BREMOND (INRA); M. DEMARAIS (ONERA); M. FONAGY et M. BOLTANSKI (Institut de phonétique); M. JOUHANEAU (CNRS); M. DUSSOUR (Langues orientales); M. JUNCK (Ets Pierret); Mme CHARNASSE (CNRS); M. CHARNOZ (cybernéticien); Mme LEIPP; M. CASTEREDE (Conservatoire National de Musique); M. MARX (Etudiant); M. OLIVA (Sté GAREN); Dr POUBLAN (médecin biologiste); M. HASSLER (prof. de Musique); Melle LHERITIER (Etudiante en musicologie); M. CHASTE; M. MELCHIOR (CGE); Melle ZIGLIER (Etudiante); M. CHARPENTIER (Ets Chedeville-Lelandais); M. GUEN (Sté Garen); M. FONTENEAU; M. MARTIN (Ing. CFTH); M. CARCHEREUX (Maître luthier); M. G. de la LONDE (CGE); M. MIZZI; Dr KADRI (Rééduc. parole); M. A. MOLES (Fac. Lettres Strasbourg); Mme et M. MOUTET (ONERA); Dr CLAVIE; M. CHENAUD (Président AFARP); M. ANDRIEU (Labo. Ac. INRA); Dr VALLENCIEN.

Excusés : M. BUSNEL; M. BLONDELET (Ets BUFFET CRAMPON); M. CH. MAILLOT; Mme FULIN; M. FAYEULLE; M. BATISSIER; M. MARCHUTZ; M. HUGOUNET; M. MANIN; M. R. LEHMANN; Mlle SOLA; Mme HELFFER; M. FRANCOIS; Mme STRAUS; M. GEORGEAIS; Dr DORGEUILLE; M. PUJOLLE; Mme GRINAUD; M. CANAC; Melle E. WEBER; M. OUNA; M. BORIS; M. DAMERON; M. CHAILLEY; M. CHESNAIS; Mme M.J. CHAUVIN; M. DUPARCQ; Melle NOUFFLARD.

...../

P L A N

		Pages
I -	GENERALITES	I
II -	LA CHAINE DE COMMUNICATION DES MESSAGES PARLES	3
1°)	L'EMETTEUR	4
a)	L'idée	4
b)	L'appareil phonatoire	4
2°)	LE CANAL	6
3°)	LE RECEPTEUR	7
A)	Le capteur	7
B)	Le centre de traitement de l'information	9
a)	Généralités	9
b)	Les mémoires	11
c)	Les opérations.....	13
III -	VARIABLES ET CONDITIONS DE L'INTELLIGIBILITE...	14
IV -	QUELQUES EXEMPLES PRATIQUES DE PROBLEMES D'INTELLIGIBILITE	17
1°)	La synthèse de la parole	17
2°)	L'intelligibilité en salle de parole.	20
3°)	L'intelligibilité de la parole dans le bruit	26
V -	CONCLUSION GENERALE	28

LE PROBLEME DE L'INTELLIGIBILITE DE LA PAROLE

I.- GENERALITES. POSITION DU PROBLEME

Les observations empiriques relatives à l'intelligibilité de la parole, au rôle du bruit dans ce problème, à la destruction des messages vocaux par le contexte architectural, remontent à l'antiquité. VITRUVE signale l'utilisation de véritables logatomes un siècle avant J.C. (bib. I)... Mais le problème a pris une acuité particulière au début de ce siècle, avec l'apparition des moyens de télécommunication, du téléphone en particulier. Des recherches systématiques furent entreprises dès lors, et on possède actuellement de nombreuses publications sur ces questions.

Très généralement on cherche à chiffrer l'intelligibilité à l'aide de différentes méthodes, plus ou moins bien adaptées au cas étudié, et sur lesquelles il convient de dire quelques mots.

- La méthode des logatomes. C'est un moyen pratique pour obtenir des taux d'intelligibilité à partir de tests statistiques. On lit une liste de monosyllabes sélectionnées selon certains critères phonétiques, et on demande à des sujets de transcrire ce qu'ils ont entendu. L'intelligibilité est définie par le pourcentage de réponses correctes.

Cette méthode appelle diverses remarques. D'abord, lors de nos recherches sur la parole, synthèse en particulier, il nous est clairement apparu que la forme physique des phonèmes varie selon les mots où ils sont incorporés; ainsi dans le logatome " ar " le phonème "r" n'a pas la même structure physique, la même forme, que dans le mot " mari ", dans le mot " sarrau " ou " garçon ".

D'autre part, considérons un certain logatome, articulé isolément dans une salle où le brouillage des échos et de la réverbération est assez faible. Il est alors parfaitement reconnu; mais le même logatome placé dans un mot normal subit, dans les mêmes conditions, le brouillage supplémentaire des échos de tout ce qui le précède comme il brouille lui-même ce qui le suit. Dans ces conditions, on comprend les objections élevées depuis longtemps par les praticiens qui ont tenté d'utiliser cette méthode, par exemple en acoustique des salles (bib. I). Les logatomes peuvent être mal articulés, mal entendus par les personnes n'ayant pas l'habitude de ce genre de " sons "; même s'ils sont bien entendus, ils peuvent être mal transcrits. Enfin, l'expérience est complétement faussée si le mot est connu d'avance : on le reconnaît toujours.

On a bien tenté d'améliorer la méthode en se rapprochant mieux du langage courant; mais on ne réussit pas à faire une bonne corrélation avec les avis formulés sur la qualité des salles de parole, tout simplement parce qu'on y utilise des artefacts n'ayant avec la réalité de la parole que des rapports assez éloignés. La popularité de la méthode des logatomes vient surtout du fait qu'elle fournit des nombres précis qui satisfont l'esprit; mais il n'est pas étonnant qu'on ait cherché d'autres voies.

Le brouillage de la parole vient en effet souvent du chevauchement des mots par l'écho des mots précédents.

...../

- La méthode des indices d'articulation. Elle consiste à découper la bande spectrale de 200 à 7000 Hz environ, contenant l'essentiel du signal physique de la parole, en 10 bandes d'" égale contribution à l'intelligibilité ", déterminées expérimentalement (bib 2). Chaque bande représente alors un dixième d'une grandeur appelée " indice d'articulation ". Cet indice est égal à I lorsque l'intelligibilité est totale.

On dicte des mots disyllabiques, des phrases, dont on détruit, par filtrage, ou masquage à l'aide de bandes de bruit, une, deux ou trois bandes etc... Le nombre de bandes nécessaires pour que les mots ou la phrase soient compris, fournit l'indice d'articulation. Par exemple si tout est compris avec 7 bandes, l'indice est de 0,7.

Comme la précédente, cette méthode a l'avantage de fournir un chiffre précis; mais nous savons maintenant que l'intelligibilité de la parole n'est pas un problème si simple qu'on puisse le trancher avec une donnée unique, si précise soit-elle, puisqu'il s'agit de la perception d'une forme. Or celle du mot " tisser ", par exemple, est située tout en haut dans l'échelle des fréquences; celle du mot " goulot ", tout en bas. Pour l'un les 4 ou 5 bandes inférieures ne jouent aucun rôle; pour l'autre, les bandes supérieures peuvent être supprimées sans inconvénient : la notion de bandes d'égale intelligibilité est donc tout à fait arbitraire et les résultats qu'on peut en attendre sont très limités. Cette méthode avait surtout été conçue pour étudier l'intelligibilité dans le bruit; elle se raccordait à la méthode d'étude des bruits qui consiste à découper le domaine des fréquences audibles en bandes d'octave, de tiers d'octave, ou en " bandes critiques ". Cette façon de faire découle de conditions pratiques relatives, en particulier, aux types d'appareils utilisés en technologie du bruit, mais il faut encore insister sur le fait que l'étude de la destruction de la parole par des bandes de bruit blanc ne peut aboutir raisonnablement à des résultats réalistes. En effet, les bruits qui détruisent la parole dans les conditions habituelles sont toujours ou presque des bruits largement évolutifs en fréquence, et le problème n'est plus du tout le même alors; nous y reviendrons plus loin. Mais d'ores et déjà on est obligé d'admettre que les méthodes classiques pour définir l'intelligibilité de la parole doivent être repensées.

Mais d'abord, que signifie donc exactement le mot : " intelligibilité " ? Les dictionnaires le définissent comme le caractère, l'état de ce qui peut être compris. Quant au terme de " compréhensibilité ", c'est " la qualité de ce qui peut être compris ". Ces deux mots sont donc pratiquement des synonymes. Cependant la terminologie des spécialistes attribue un sens différent aux deux mots. La pierre de touche de l'intelligibilité serait la capacité de reconnaitre un logatome; celle de la compréhensibilité, la capacité de reconnaître un mot, une phrase où certains éléments pourraient être détruits, mais seraient reconstitués mentalement.

..../

Il s'agit, on le verra plus loin, de subtilités; la parole est un message; elle est intelligible quand le message est compris par le récepteur. Mais celà suppose une convergence d'un très grand nombre de conditions définies par la combinatoire entre les propriétés et les interactions des nombreux maillons de la chaine de communication des messages de la parole, qu'il faut donc préalablement définir et étudier en détail. La lecture des publications sur ces questions montre malheureusement qu'en dehors des maillons techniques (canaux) on reste dans le vague; l'élément essentiel; l'homme, considéré comme récepteur-émetteur de messages parlés, reste un mystère dans une très large mesure. Les biologistes et les psycho-physiologues, malgré tous leurs efforts, ne nous ont apporté que des connaissances très lacunaires, généralement très hypothétiques. Dans ces conditions il est nécessaire de chercher une réponse par d'autres voies. Une méthode qui s'est avérée fructueuse en d'autres domaines est celle de la simulation à l'aide de modèles de fonctionnement. Ici, on ne cherche plus à extraire la vérité des observations anatomiques et physiologiques; on ne se propose plus d'expliquer la structure et le mécanisme de la réalité qui, du fait de son extrême complication et miniaturisation échappe en fait à nos sens et à nos méthodes actuelles d'investigation. Mais on se propose de définir les fonctions nécessaires et suffisantes pour simuler et expliquer la réalité; bref, on cherche à réaliser un modèle qui fonctionne comme l'original, et dans lequel il soit possible de raccorder toutes les observations objectives des physiologistes.

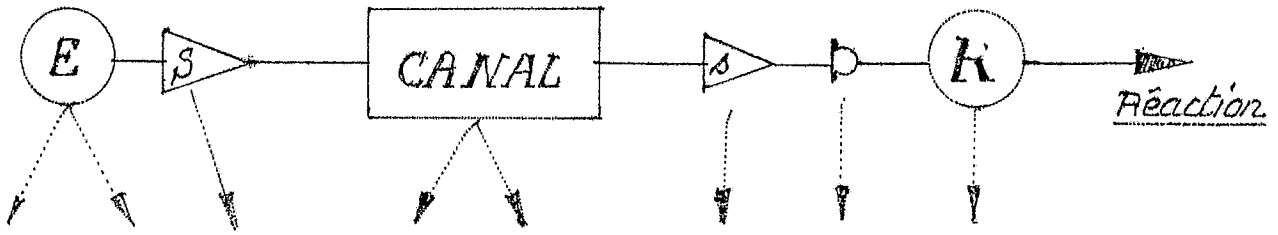
La mise au point d'un modèle de fonctionnement de la chaine de communication des messages parlés nous préoccupe depuis fort longtemps (bib 3 à 9). Mais entretemps, nous avons approfondi nos connaissances sur l'appareil phonatoire, la structure physique et la synthèse de la parole, la perception des signaux acoustiques etc. Nous avons accumulé ainsi d'importants compléments qui nous permettent à présent de proposer un modèle de fonctionnement cohérent, dont l'intérêt est évident dans le problème de l'intelligibilité, comme on va voir.

II. LA CHAINE DE COMMUNICATION DES MESSAGES PARLES

Très généralement, la communication d'un message suppose un émetteur (E) qui conçoit le message (idée) et le matérialise sous forme de signes (S). Ces signes physiques sont tout à fait arbitraires; l'important est qu'ils puissent être matériellement fabriqués par l'émetteur, qu'ils aient fait l'objet d'une convention préalable avec le récepteur (R) relativement à leur signification et qu'ils soient transportables à distance.

.../

Fig 1



	Idée	Système phonat.	Signal rayonné	naturel / artific.	Signal reçu	Oreille	Centre de traitt	musculaire ou mentale
FONCTIONS	Faire réagir R — envoyer un ordre à l'app. phonatoire	Fabriquer des Formes acoustiques	Véhiculer l'inform. chargée de redondance.	Filtrer / masquer	apporter l'inform. au capteur	Capter la forme acoustique et la transformer en configuration électrique	Traiter L'inform. en fonction du Contenu des mémoires	agir

EXEMPLE

Faire sauter

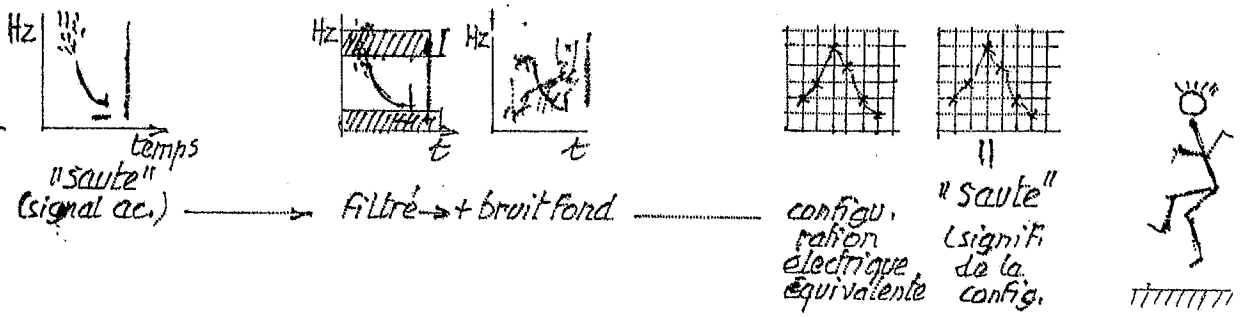
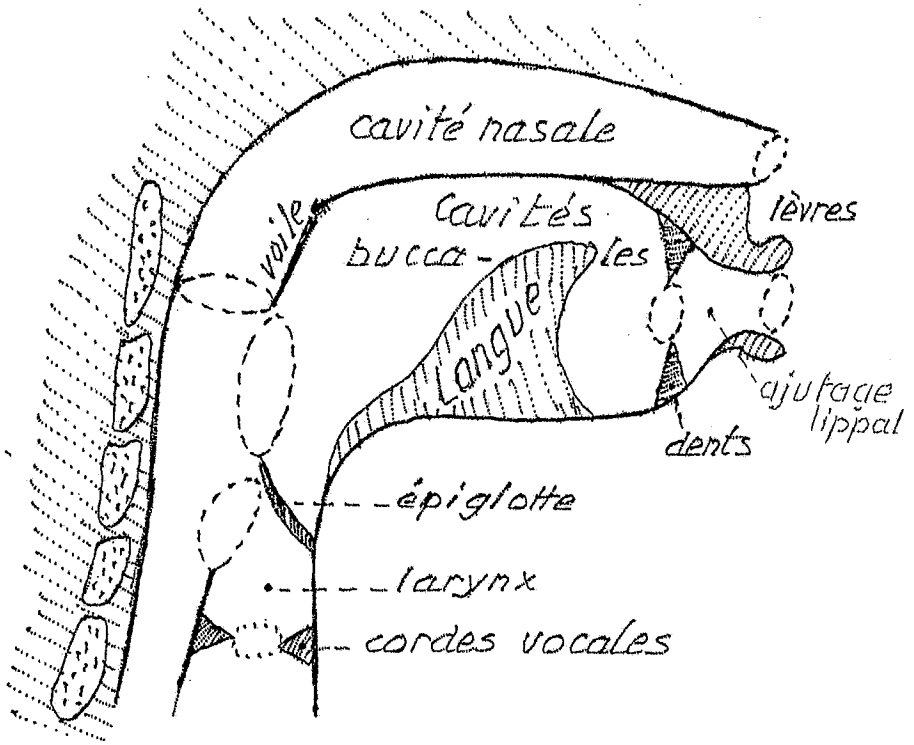


Fig 2



L'appareil phonatoire.

C'est une machine à tee d'une musculature élaborée. Les mouvements élémentaires des organes en présence sont synchronisés par un "programme" appris stocké en mémoire (archétype.) Elle permet de réaliser une variété infinie de signaux acoustiques différents.

Les signes acoustiques dont il s'agit ici sont les formes acoustiques à trois dimensions fabriquées par une "machine"très élaborée, l'appareil phonatoire. Ces formes passent ensuite dans un canal qui peut être le canal naturel, pour lequel la parole s'est empiriquement élaborée, ou bien un canal artificiel, technologique (microphone, téléphone, amplificateurs, haut-parleurs etc.) Pendant le transfert, les formes sont plus ou moins amputées, déformées. Finalement elles parviennent au voisinage du récepteur (R) qui les capte et traite l'information qu'elles contiennent. Le résultat normal du message est une réaction: musculaire ou mentale. Si le récepteur réagit conformément aux conventions, le message est compris: il y a intelligibilité. Il s'agit donc en dernière analyse d'un problème de fabrication, de structure, d'acheminement et de reconnaissance de formes, qui suppose nécessairement l'étude des fonctions et des propriétés de trois grands secteurs: émetteur, canal et récepteur.

I. L'EMETTEUR

Il pose trois problèmes distincts: celui de l'idée, celui du système phonatoire et de son fonctionnement, celui de la structure physique et sémantique des signaux rayonnés.

a) L'Idée. La nécessité d'envoyer des messages découle directement des conditions de vie en société, où l'homme a besoin de communiquer de l'information à son semblable. Le langage des premiers hommes était de toute évidence très rudimentaire, se limitant à un nombre restreint de borborygmes et d'onomatopées, chargés d'un sens très précis cependant, et liés à la vie ou à la survie de l'individu: cris de douleur, cris de danger, appels etc. Si l'homme a réussi à se forger le riche éventail des formes acoustiques significantes que nous observons dans les langues actuelles, c'est d'abord grâce aux particularités de son système auditif et phonatoire, mais surtout grâce aux caractéristiques, à la capacité, aux performences de son système de mémoires, qui conditionnent la possibilité de concevoir une idée élaborée et l'envie de la communiquer à quelqu'un d'autre. Cette idée s'étant fait jour, il lui faut encore un support matériel; cette fonction est remplie par l'appareil phonatoire.

b) L'appareil phonatoire: structure physique et information sémantique de la parole.

L'appareil phonatoire est une machine à fabriquer des formes à trois dimensions dont la meilleure représentation est donnée par le sonagraphe qui fournit des diagrammes fréquence-temps, où l'intensité est indiquée de façon sommaire mais suffisante (bib 10 à 14).

Cette machine comporte une série de cavités qui sont autant de résonateurs dont le volume et les ouvertures sont largement réglables au gré du locuteur (fig 2).

.../

Ces résonateurs sont couplés en série (pharynx, cavités buccales, ajutage lippal), sauf la cavité nasale qui est en parallèle. Ils déterminent des zones de résonance variables (formants) et chaque ouverture, débouchée brusquement, est susceptible de produire de petites explosions ou implosions. Le système peut être excité par le bruit d'écoulement de l'air venant des poumons, ou par le spectre que délivre un système d'anche double musculaire : les cordes vocales.

L'articulation d'un mot résulte d'un sous-programme de mouvement très compliqué, préalablement stocké en mémoire lors de l'apprentissage de la langue. Le nombre de programmes de mouvements nécessaires pour parler couramment est énorme. En fait nous avons montré que chaque mot était une " super-forme " réalisée à l'aide d'un nombre limité et relativement restreint de sous-programmes élémentaires correspondant à des mouvements partiels de l'appareil phonatoire entre deux positions limites.

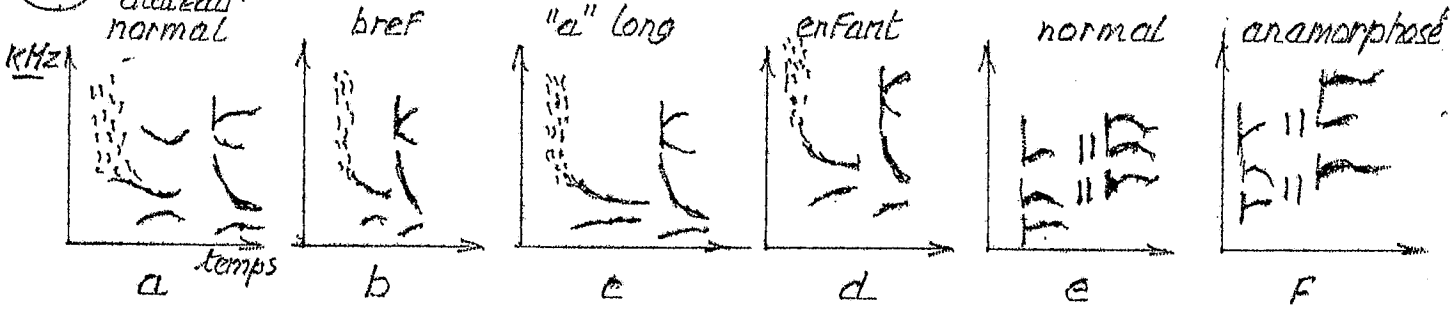
Ainsi le mot " PARIS " se décompose en trois mouvements élémentaires représentés par les " éléments phonétiques ", les " diphonèmes " PA, AR, RIS, qui se raccordent par la force des choses, chaque mouvement élémentaire s'arrêtant à la position où débute le suivant. Ces diphonèmes représentent en fait les plus petits " atomes " de langage possible, si on veut bien prendre en considération la mécanique de l'appareil phonatoire; c'est pourquoi nous les appelons aussi " phonatomes ".

Dans ces conditions, au lieu de stocker en mémoire autant de formes globales très compliquées qu'il existe de mots dans le dictionnaire employé, il suffira de stocker les formes élémentaires des quelque 900 phonatomes existants (association deux à deux des " phonèmes " classiques) et autant de programmes d'association qu'il faut de mots. Les programmes d'association étant très simples, on réalise ainsi une énorme économie de mémoires, comparativement à ce qui serait le cas si l'on était obligé de stocker l'intégralité des formes pour chaque mot du dictionnaire utilisé. On se rappellera de ce point de vue quelques chiffres intéressants. Un enfant de deux ans articule normalement, de façon intelligible, quelque 300 mots; ce nombre passe à 1000 autour de trois ans et à 2000 à cinq ans. Un adulte cultivé utilise quelque 25 000 mots ! Les oeuvres de Shakespeare contiennent environ 15 000 mots différents.....

Le sous-programme de mouvement correspondant à un phonatome représente bien entendu un archétype, variable d'un individu à l'autre et dont la réalisation matérielle conduit à des formes acoustiques chaque fois très différentes, en raison de l'imprécision de la mécanique de l'appareil phonatoire, ainsi que de ses dimensions, non normalisées d'un individu à l'autre. Il est donc évident que les signaux de parole fabriqués par la société qui nous entoure varient dans une très large mesure, mais la forme de chaque mot reste reconnaissable sans ambiguïté parmi toutes les autres puisque nous comprenons à peu près tous les locuteurs parlant la même langue.

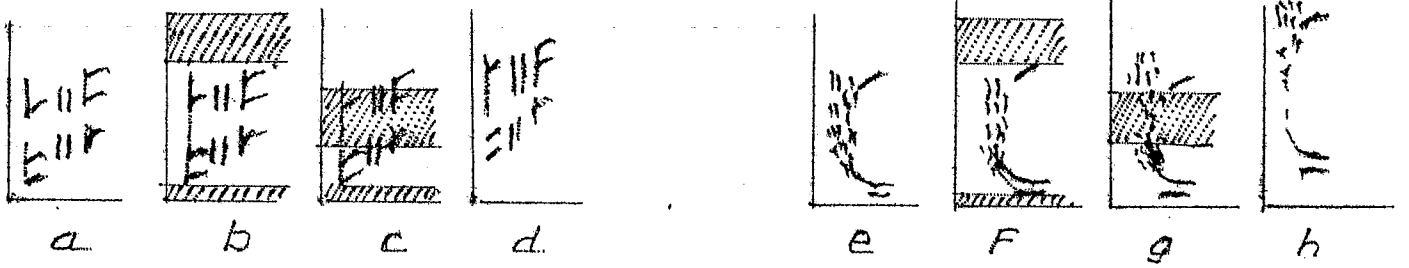
...../

Fig.3 Formes sémantiques des mots "château" et "Paris"



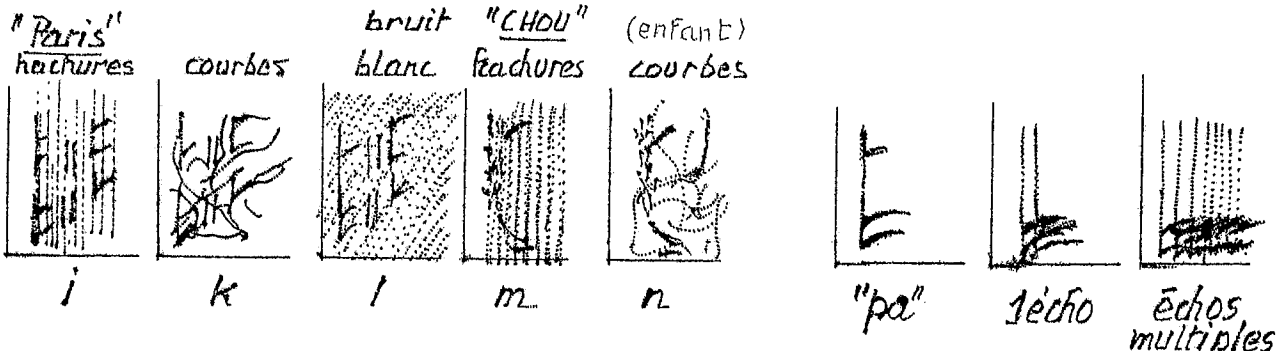
Les formes sémantiques sont largement anamorphosées d'un locuteur à l'autre. mais on ne confondra en aucun cas "château" avec "Paris". Le langage s'est établi empiriquement en tenant compte de la nécessité d'éviter la confusion et l'ambiguïté. d'où le choix de formes "fortes".

Fig 4 a-h Le mot "Paris" (enfant) Le mot "chou" (enfant)



Le canal peut détruire diversement les formes par filtrage: c'est chaque fois un cas particulier résultant du rapport entre forme et situation des bandes de rejection.

Fig 4 i-n



Les formes sont susceptibles d'être détruites par masquage total ou partiel, selon l'allure du bruit de fond par rapport à celle de la forme. La reconnaissance de la forme est une question d'émergence d'une forme sur un fond et non de rapport signal-bruit ou de bandes "d'égale contribution à l'intelligibilité.

Cette forme est une véritable image acoustique, et nous avons montré ailleurs que la meilleure manière de la mettre en évidence, consistait à relever les sonagrammes de voix chuchotée. Quelques exemples montreront clairement que les " formes sémantiques " obtenues ainsi restent reconnaissables malgré l'absence de normalisation.

Tel locuteur prononce le mot " château " normalement (fig.3a): l'autre aura un débit rapide (3b), un troisième traînera sur la voyelle "a" (3c); enfin un enfant, dont les cavités sont plus petites, dessinera la forme (3d). Toutes ces formes sont en fait un seul et même graphisme plus ou moins anamorphosé; mais quelle que soit l'anamorphose, on ne les confondra jamais avec celles du mot " PARIS " (3e et 3f); or c'est celà qui importe.

La notion de forme a fait l'objet de nombreuses recherches de la part des psychologues, et en particulier des théoriciens de la " Gestalt " dont von EHRENFELS fut le promoteur à la fin du siècle dernier. Les principales conclusions des gestaltistes se retrouvent ici : une forme est une totalité perçue en bloc; c'est quelquechose de plus que la somme de ses parties. Une forme est transposable et anamorphosable : elle ne peut donc pas être définie par des fréquences ou des durées absolues lorsqu'il s'agit de formes acoustiques, mais par des rapports de fréquences associés à des rapports de durées. Autrement dit, l'expérience montre que le récepteur humain ne mesure pas des grandeurs absolues mais apprécie des rapports de grandeurs qui définissent précisément des formes acoustiques.

Nous avons longuement vérifié toutes ces propositions par d'innombrables analyses au sonagraphe, mais aussi par la synthèse de la parole réalisée à l'ICOPHONE, appareil que nous avons décrit en détail ailleurs et qui, en fait, est un convertisseur de formes graphiques optiques en formes acoustiques. Retenons donc que l'appareil phonatoire fabrique des formes acoustiques non normalisées mais que nous reconnaissons, à condition, bien entendu qu'elles ne soient pas détruites par le canal, dont il convient à présent de dire quelques mots.

3°) LE CANAL

Le canal doit véhiculer les formes acoustiques de la parole à distance; mais nous savons que tout canal possède deux particularités :

- celle de détruire partiellement le signal par filtrage en jouant le rôle d'un filtre de réjection, susceptible de couper plusieurs bandes qui amputent ainsi le signal plus ou moins.

- celle de masquer certaines parties du signal par le bruit de fond obligé de tout canal, qu'il soit naturel ou artificiel.

...../

De toutes façons les formes de la parole sortent du canal plus ou moins détruites selon les cas et mélangées avec d'autres " formes acoustiques " étrangères au message. Nous avons étudié ailleurs (bib. 15) les conditions générales d'émergence d'une forme acoustique sur un fond. Quelques exemples montreront ce qu'il en est lorsqu'il s'agit de parole (fig.4).

Prenons encore la forme du mot " PARIS " (4a). Deux bandes de réjection de même largeur peuvent, selon le cas, laisser subsister l'intégralité de la forme (4b) ou la détruire partiellement (4c), sans qu'il puisse cependant y avoir ambiguïté avec la forme du mot " chou " dans les mêmes conditions (e, f, g,). Par contre le mot " chou "prononcé par un enfant peut être parfaitement compris (d, h) avec les mêmes bandes de réjection.

L'effet du bruit de fond peut de même varier à l'infini, :le masquage est susceptible de détruire plus ou moins complètement la forme. A titre d'exemple, reprenons le mot " PARIS ". Sa forme peut être rendue méconnaissable par une série de hachures verticales (bruit de motocyclette par exemple) comme on voit en 4j. Elle peut devenir totalement invisible (4 k) dans un graphisme de quelques lignes courbes surajoutées (sons évolutifs en hauteur). Par contre la forme émerge totalement sur un bruit blanc, même assez intense, tant qu'il n'y a pas saturation de l'oreille (4 l). Par contre le mot " chou " peut être totalement détruit là ou " PARIS " reste tout à fait perceptible et invers ement (4 m, n). Signalons encore que la forme peut se détruire elle-même; c'est le cas de l'écho simple ou multiple (fig.4 ou fig.9).

Tout celà montre d'une part l'impossibilité de définir des " bandes d'égale contribution à l'intelligibilité ", d'autre part l'inadéquation de la notion de signal/bruit chère à la technologie acoustique dans le problème de l'intelligibilité de la parole.

Il est donc clair que le rôle du canal est déterminant et, en tout cas, le système auditif devra être capable de " reconstituer " ce qui a été coupé par le canal et de reconnaître une forme dans le mélange signal-bruit. Disons tout de suite qu'il est admirablement organisé pour celà.

4°) LE RECEPTEUR

C'est un véritable centre de traitement de l'information que nous envoie le monde extérieur et qui doit être préalablement captée par l'oreille.

A) Le capteur (fig.6). Si l'anatomie de l'oreille est bien connue, sa physiologie reste largement hypothétique. On sait cependant sûrement qu'il s'agit d'un capteur électro-acoustique, transformant les variations de pression acoustique du signal (la forme acoustique si on préfère) en forme électrique, en configu-

..../

ration d'impulsions. Sous cette forme, l'information peut être véhiculée vers le centre de traitement par les voies nerveuses. Il est certain que la configuration impulsionnelle n'a plus aucun rapport avec la structure acoustique du signal original; celle-ci est codée, autrement dit. Celà n'a aucune importance du point de vue de la reconnaissance de la forme acoustique, c'est-à-dire du message qu'elle véhicule. En effet, il suffit qu'il y ait correspondance univoque : le même mot donne toujours la même forme acoustique; la même forme acoustique donne toujours la même configuration impulsionnelle : mot, forme acoustique et configuration impulsionnelle représentent la même chose, et l'une appelle l'autre. Nous nous proposons de revenir en détail sur ces questions dans un travail ultérieur; il suffit ici d'avoir attiré l'attention sur cette question.

Il est nécessaire, à présent, d'insister sur une autre particularité du capteur, celle de réaliser un système asservi au niveau du signal grâce à la chaine ossiculaire. On sait de quoi il s'agit. Trois osselets couplés communiquent les vibrations du tympan à la fenêtre ovale; leur musculature respective permet de modifier la raideur du système global, donc de réduire les amplitudes mécaniques du système en cas de besoin. Ce dispositif permet d'abord de protéger l'oreille interne contre les trop grands niveaux acoustiques qui risqueraient de détruire les organes fragiles de la cochlée. L'adaptation se fait par voie réflexe, mais il faut un certain délai pour que le mécanisme joue; on peut faire un parallèle avec la pupille de l'oeil lorsque la lumière devient trop violente. Mais par la même occasion, ce dispositif permet d'éviter la saturation électrique des centres supérieurs, saturation qui se traduit par un bruit qui masque alors toute forme acoustique éventuelle. Pour éviter la saturation, les appareillages électroniques comportent un potentiomètre d'entrée; les osselets jouent précisément le rôle de ce potentiomètre, avec la différence qu'ici il est asservi au niveau. Quand celui-ci devient trop fort, le système ossiculaire intervient automatiquement. Diverses observations pratiques que nous avons faites, nous font penser que l'atténuation est très importante, de l'ordre de 25 ou 30 dB, et il découle de ce fait des conséquences importantes. Jusque vers 70 dB, l'oreille ne risque normalement pas d'être saturée. Par contre, au-dessus de 100 dB la saturation est inévitable. Mais entre 70 et 100 dB, c'est-à-dire dans la grande majorité des cas, notre système auditif peut être saturé ou non selon la prévisibilité du signal. Si nous sommes prévenus à temps pour que le réflexe ossiculaire puisse entrer en jeu, un signal de 110 dB sera effectivement perçu comme s'il n'avait par exemple que 110 - 30 dB, c'est-à-dire de 80 dB et la saturation n'a pas lieu. Nous pouvons alors percevoir une forme de parole contenue dans un bruit intense, ce qui serait impossible sans le mécanisme ossiculaire.

Autre conséquence : si la parole est trop faible, nous pouvons " tendre l'oreille ", c'est-à-dire accommoder la musculature ossiculaire de manière à obtenir des amplitudes maxima au niveau de la fenêtre ovale !.

On comprendra aisément dans ces conditions que l'intelligibilité de la parole puisse changer selon les capacités d'

..../

adaptation de la chaire ossiculaire, selon ses possibilités de compression dynamique; toutes choses égales, l'intelligibilité diminue pour cette raison avec le vieillissement ou la fatigue, puisqu'il s'agit de déficiences musculaires au niveau des osselets.

Finalement on retiendra que le système ossiculaire est un système automatique de compression de dynamique, permettant aux centres supérieurs de recevoir de l'information dans une gamme dynamique beaucoup plus large que celà ne serait possible sans les osselets. Voyons à présent comment on peut imaginer le centre de traitement à la lumière de ce que nous a appris l'informatique.

B) Le centre de traitement de l'information

a) Généralités.

Le cerveau et son fonctionnement continuent à intriguer de très nombreux chercheurs sans qu'on soit arrivé à comprendre ce qui s'y passe exactement. Les éléments en présence sont tellement miniaturisés, les composants électro-chimiques tellement nombreux, les " cables " tellement enchevêtrés qu'il est impossible de s'y retrouver par l'observation anatomique qui ne fournit d'ailleurs à peu près jamais d'explication physiologique puisqu'il s'agit partout de systèmes électriques.

On a proposé à diverses reprises, des modèles de fonctionnement; mais aucun ne se raccordant avec nos observations journalières sur la perception des signaux acoustiques , nous avons été amenés à en concevoir un à notre usage, et dont nous voudrions dire quelques mots.

L'idée est centrée autour de ce que nous a appris l'informatique, dont le but est précisément l'étude très générale des moyens possibles pour traiter l'information sous tous ses aspects. On trouve maintenant sur ces questions une importante littérature scientifique et de vulgarisation à laquelle nous renvoyons pour le détail. Précisons cependant quelques points.

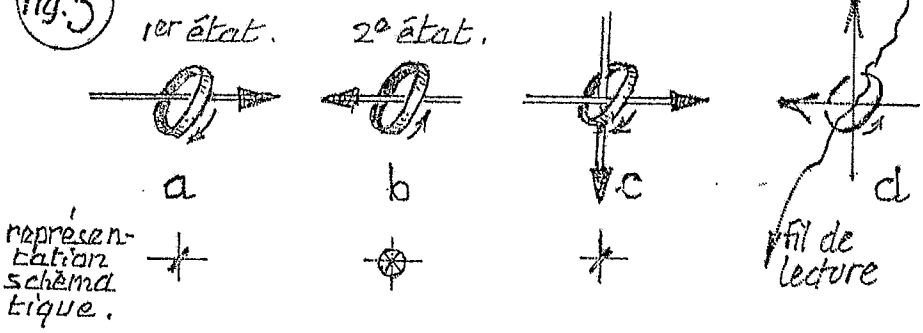
Toute information est quantifiable, c'est-à-dire décomposable en micro-éléments, en quantas acoustiques ici, qui sont par exemple les points de l'image sonagraphique du mot " PARIS ". Chacun de ces points est définissable par deux nombres, comme dans une grille de mots croisés.

Une configuration d'impulsions électriques peut de même être décrite par des nombres; chacun de ceux-ci étant réductible à deux chiffres seulement, un et zéro par exemple si on adopte la numération binaire. Il suffit pour cela que l'on ait matérialisé au préalable la configuration électrique sur une matrice à tores.

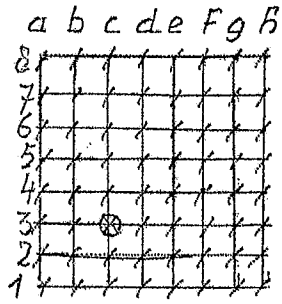
...../

Fig.5

TORES MAGNÉTIQUES

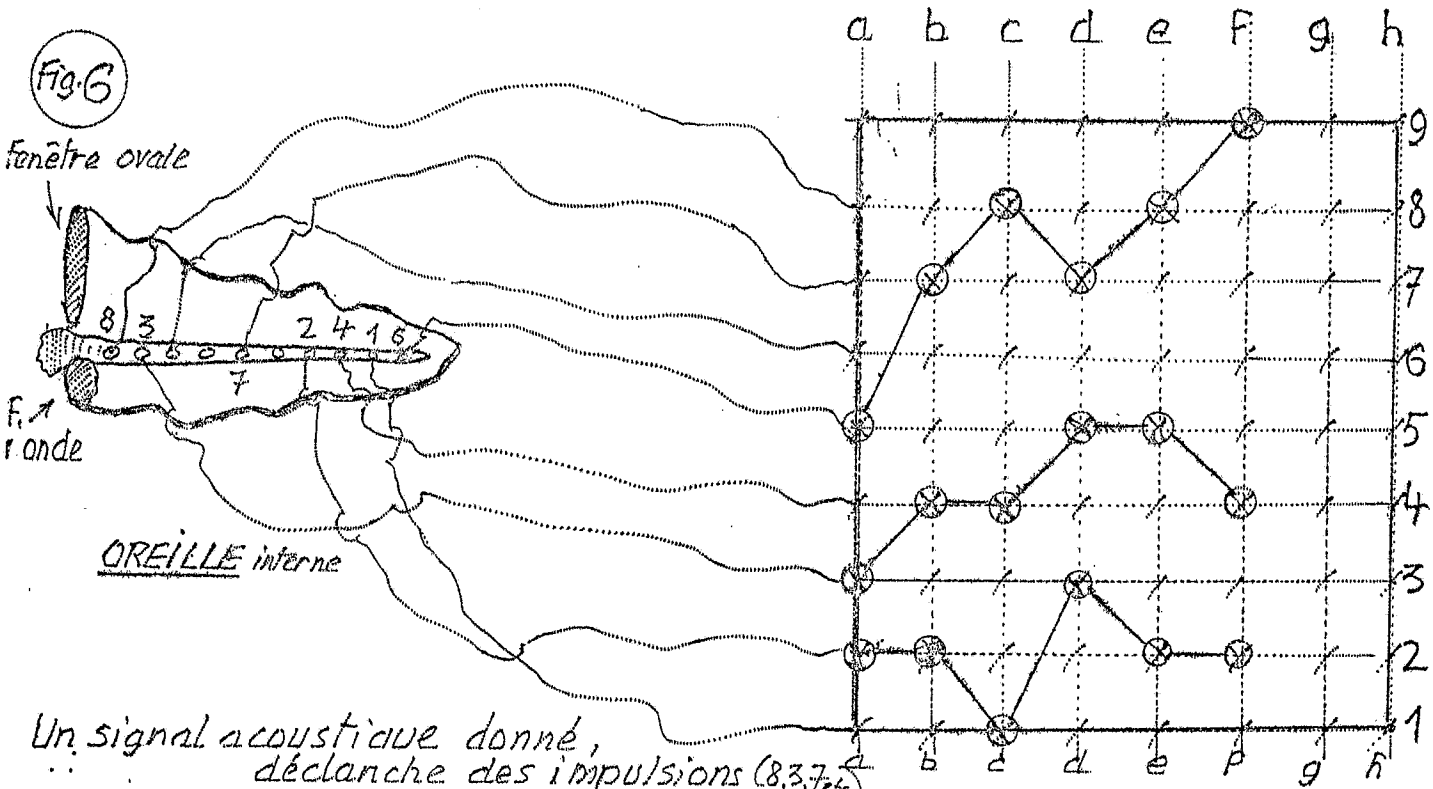


Matrice à tores



Quand on envoie une impulsion de gauche à droite, le tore prend un certain état magnétique (a). En inversant le sens de l'impulsion. Le tore bascule dans l'état inverse (b). On peut obtenir le même résultat en faisant passer dans le tore deux fils croisés et en envoyant dans chacun une intensité plus faible (moitié du cas précédent.)(c). En inversant le sens du courant dans ce système le basculement du tore envoie un courant induit dans un troisième fil passant dans le tore (fil de lecture) (d)

Fig.6



Un signal acoustique donné, déclanche des impulsions (8,3,7 etc)

Les cellules de l'organe de Corti envoient alors ces impulsions sur la matrice à tores de la mémoire instantanée où l'on met successivement sous tension les fils verticaux a, b, c, d etc. Les tores concernés à chaque instant basculent, dessinant sur la matrice une configuration de points dont la forme n'a rien à voir avec la forme acoustique. mais qui la représente et permet de la reconnaître.

Un fil de lecture reliant tous les tores permet de "recopier" cette matrice et de l'envoyer plus loin, en mémoire II.

NOTA : Pour le lecteur non initié rappelons sommairement le principe. On sait qu'un courant électrique (ou une impulsion) passant dans un fil, détermine autour de celui-ci un champ électrique. Si on place dans ce champ un tore en matériau magnétisable, ce tore est donc magnétisé avec une certaine polarité. (Fig.5a). Le matériau est choisi tel que lorsqu'on coupe le courant, le tore reste magnétisé; on se rappellera qu'il faut une certaine intensité de courant pour que la polarisation ait lieu, un certain seuil.

Lorsqu'on envoie maintenant dans le tore polarisé un courant en sens inverse, pour peu qu'il soit suffisamment intense, la polarisation va changer de sens : le tore bascule de l'état précédent à un deuxième état (fig.5b).

Faisons à présent passer deux fils perpendiculaires par le tore. Mettons maintenant par exemple le fil horizontal sous une tension plus faible que le seuil du tore. Celui-ci restera dans l'état où il est. Mais si nous envoyons sur l'autre fil, vertical, un courant tel que le supplément de champ magnétique s'ajoute à celui du fil horizontal au point que la somme dépasse le seuil, et dès lors le tore bascule en inversant sa polarité.

Faisons passer à présent un troisième fil dans le tore que nous appelerons " fil de lecture ". Lorsque le tore basculera il se produira dans ce fil un courant induit que l'on peut véhiculer où on le désire. Tout celà est très classique, mais nous fournit la solution de ce que nous désirons : fabriquer une " image électrique ".

En effet, construisons une matrice à tores, un " réseau maillé " comportant d'une part une série de fils horizontaux, d'autre part des fils verticaux et enfin un fil de lecture qui relie tous les tores entre eux, les uns à la suite des autres.

Il suffit alors d'envoyer sur tel fil horizontal et simultanément sur tel fil vertical respectivement un courant dont chacun isolément est trop faible, mais dont la sommation est suffisante pour faire basculer un tore situé au point de croisement des fils, et là seulement. En envoyant deux impulsions sur les fils convenables, nous pouvons donc " dessiner un point " sur la matrice à tores. Comme une forme est la somme de points, il est facile de dessiner ainsi une forme quelconque sur la matrice.

Lorsque toute l'image sera inscrite sur la matrice, on pourra aussi " effacer " l'image en envoyant en bloc des impulsions en sens contraire, par exemple dans la totalité des fils horizontaux. Dans ce cas les tores polarisés d'une certaine façon ne bougeront pas; mais ceux qui sont inversement polarisés vont basculer dans leur état

..../

d'origine et on aura sur le fil de lecteur général une série d'impulsions électriques disposées spatialement d'une certaine manière. Cette série d'impulsions peut être véhiculée ailleurs : elle représente de toutes façons la même configuration que celle qui se trouvait précédemment sur la matrice.

Dès lors nous allons comprendre comment fonctionne l'oreille. Imaginons la cochlée associée avec une matrice à tores (mémoire instantanée, fig.6). On sait que l'organe de Corti, dans la cochlée, porte un certain nombre de cellules qui envoient une impulsion chaque fois qu'elles sont excitées par une vibration mécanique. L'oreille interne est un système élastique très compliqué de forme irrégulière, qui se met en vibration pour tout signal acoustique extérieur. Chaque point du système suit une certaine loi de mouvement élémentaire. L'ensemble des points représente alors une configuration de mouvements élémentaires, toujours la même pour un même signal acoustique donné. Un tel signal se traduira sur la cochlée, à un instant donné, par une disposition particulière de cellules au repos et de cellules qui envoient une impulsion. Cette disposition est alors envoyée vers la mémoire instantanée (I). L'instant suivant la disposition change lorsqu'il s'agit d'un phénomène acoustique évolutif (comme la parole en particulier); cette nouvelle disposition est alors inscrite sur la verticale suivante et ainsi de suite. Prenons un exemple. A l'instant " a ", les cellules 2, 3 et 5 fonctionnent, faisant basculer les tores a2, a3 et a5. L'instant suivant (b), ce sont les tores b2, b4 et b8. Puis c'est le tour de la colonne c, d et ainsi de suite.

Finalement on aura une somme de points, c'est-à-dire une " image électrique " que nous retrouverons chaque fois que nous aurons eu à l'entrée de l'oreille le même signal acoustique.

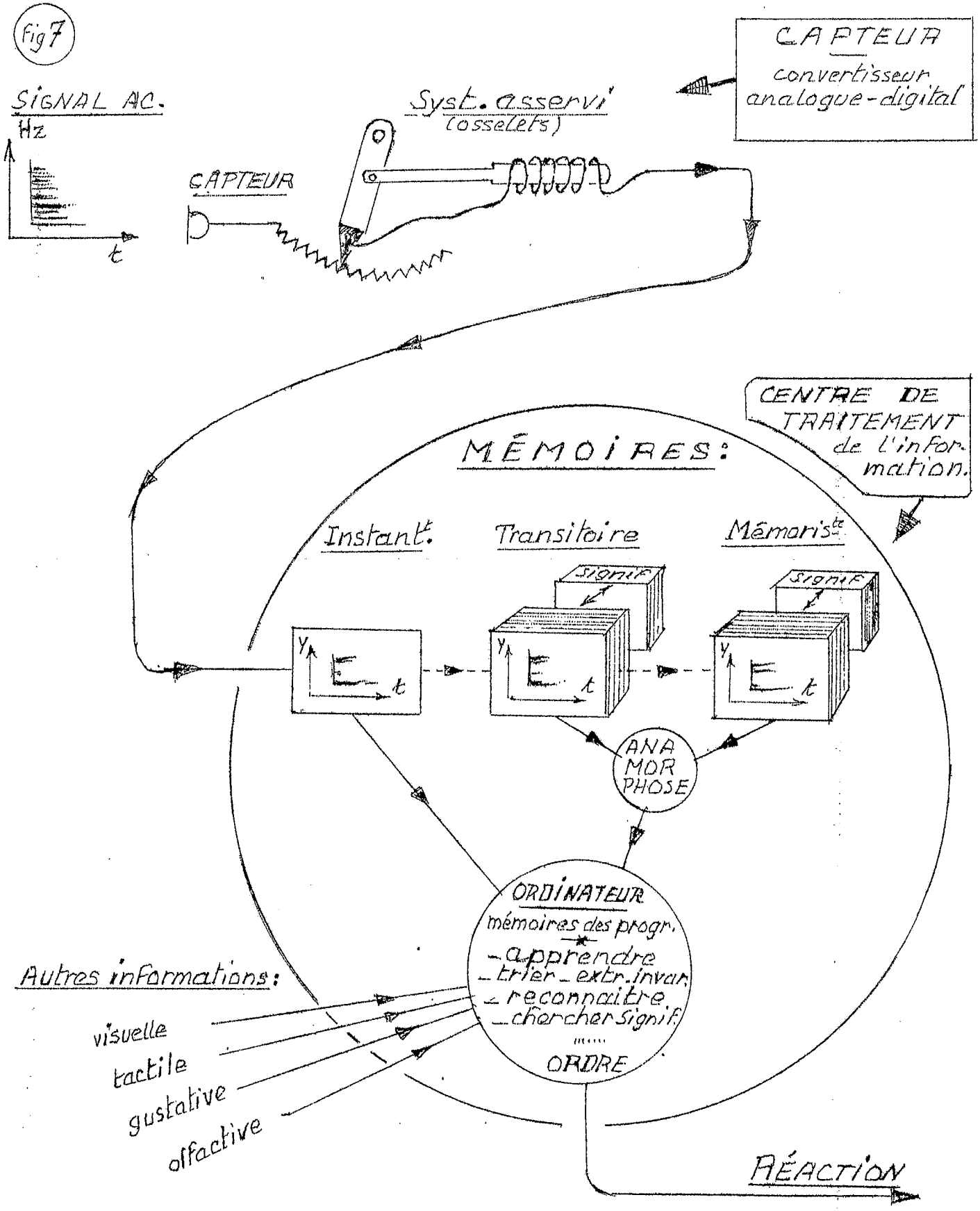
Pour terminer l'opération, on peut facilement imaginer qu'en arrivant à la dernière colonne de la matrice, un ordre soit envoyé pour " effacer " la matrice; on récupère alors sur le fil de lecture la configuration que l'on peut éventuellement " recopier " sur une deuxième matrice en cas de besoin.

Ces dispositifs, avec quelques variantes, sont maintenant réalisés dans la majorité des ordinateurs. On peut aisément imaginer que la nature ait utilisé des matériaux différents des ferrites employés dans les ordinateurs, et miniaturisé les " composants électroniques " à l'extrême. Nous y voici dès lors au coeur du problème : nous allons pouvoir enregistrer sur matrices objectives et recopier à volonté des informations; ces matrices simulent parfaitement nos mémoires.

b) Les mémoires (Fig.7). Trois types de mémoires sont indispensables pour comprendre ce que nous observons en audition :

..../

Fig 7



La forme acoustique en Hz/temps est d'abord codée en impulsions électriques par le capteur. Un système d'asservissement au niveau permet d'éviter la saturation de la mémoire instantanée. La configuration d'impulsions envoyée par le capteur est codée une seconde fois (forme temporelle sur la mémoire instantanée). Cette dernière, différente de celle du signal physique, permet cependant de la reconnaître

- une mémoire instantanée. C'est la matrice à tores dont nous avons parlé tout à l'heure (fig.6). Nous estimons sa capacité temporelle normale de quelques secondes; mais elle varie certainement à l'infini selon les individus. L'actualité acoustique s'inscrit ici au fur et à mesure; après 2 ou 3 secondes, tout est périodiquement effacé en bloc et on recommence. Mais en cas de besoin, si on y trouve un intérêt, la configuration électrique peut être récupérée par le fil de lecture et on peut l'envoyer sur l'une des matrices à tores disponibles dans une " mémoire transitoire ".

- la mémoire transitoire. Elle est du même type que la mémoire instantanée. Mais on peut y stocker une quantité d'information, d'images, variable d'un individu à l'autre, mais en tout cas limitée. C'est l'opération que nous faisons lorsque nous écoutons par exemple attentivement un cours qui nous intéresse : nous " retenons " les signaux. Notons qu'à chaque matrice de signes peut être adjointe, dans une mémoire parallèle, la signification de chaque signe; celle-ci nous permet de " comprendre " : tel signal acoustique appelle telle signification. On peut ainsi imaginer toute une hiérarchie de mémoires associées les unes aux autres, remplies de signes autres qu'acoustiques, envoyés par des capteurs différents comme l'oeil, le nez etc... Pour les autres sens, le mécanisme de perception est strictement le même : seuls les capteurs sont sensibles à des phénomènes différents. De toutes façons le traitement de toutes ces informations sera fait par le seul et même organe de calcul ultérieurement.

La mémoire transitoire peut être effacée à volonté pour des raisons variées, par exemple :

- si on retrouve plusieurs fois la même image. On efface alors les "doubles", les " triples " etc... pour récupérer de la place dans la mémoire.

- si rien n'est intéressant. On peut tout effacer; la mémoire est alors disponible entièrement.

- si quelquechose s'est avéré d'importance capitale. On sait qu'il faut alors le " retenir " toute la vie; on fait alors une copie définitive qui ne s'effacera plus et on l'envoie dans la mémoire mémorisante.

- La mémoire mémorisante. Elle est peut être d'une nature spéciale; mais elle peut très bien être similaire aux précédentes, comportant cependant, comme dans certaines mémoires d'ordinateurs, un câblage tel qu'il soit impossible de l'effacer. L'image peut cependant " pâlir " par vieillissement, ou être détruite par voie chimique ou par traumatisme. En principe elle dure toute la vie; c'est là que sont nos souvenirs d'enfance en particulier et beaucoup de choses que nous avons apprises lorsque nous étions jeunes. Le contenu de cette mémoire est en

..../

fait le portrait de l'individu, car il détermine toutes ses réactions.

- La mémoire des programmes de traitement. Les mémoires précédentes renferment toutes les données que nous a fournies le monde extérieur. Qu'allons-nous faire de toute cette information ? Nous n'en ferons rien si nous ne possédons pas de programmes de traitement. Nous avons appris à l'école et tout au long de notre vie à faire des opérations variées, tant manuelles que mentales, mais qui supposent toutes un certain ordre de manipulation. Un ordinateur, une unité de traitement, ne peut, on le sait, faire que des opérations très simples. Mais par ailleurs tout problème compliqué, mathématique ou logique, peut toujours se réduire en dernière analyse à un très grand nombre de réponses par " oui " ou " non " . L'unité de traitement possède de ce point de vue l'intéressante propriété de pouvoir faire ces opérations à une rapidité vertigineuse. Celle que nous avons placée dans notre modèle, permet pour celà de faire rapidement toutes les opérations mentales qui nous intéressent ici.

c) Les opérations

Bornons-nous à celles qui concernent directement l'objet de cette étude, c'est-à-dire l'intelligibilité de la parole. Avec les fonctions précédentes nous pouvons :

- entendre. C'est laisser s'inscrire et s'effacer périodiquement l'actualité acoustique, sans souci de retenir quoi que ce soit.

- apprendre. C'est stocker signes et significations

- trier. C'est faire l'inventaire en T et rejeter ce qui est sans intérêt parce qu'on a besoin de place.

- extraire des stéréotypes : c'est extraire l'invariant de toutes une série de formes identiques, aux dimensions près, en n'en retenant que les points communs et significatifs. C'est un peu ce que fait le caricaturiste qui repère le minimum de points nécessaires et suffisants pour que l'on reconnaisse une forme, une personne. Les stéréotypes stockés en T ou M serviront alors de référence pour traiter l'information de la façon la plus économique.

- reconnaitre. Une forme apparait en I; on appelle les stéréotypes les plus probables selon le contexte, et on les superpose aux formes actuelles afin d'en extraire un taux de corrélation. Si tous les points du stéréotype se retrouvent dans le signal actuel, la forme est reconnue sans ambiguïté.

Mais nous avons vu plus haut que les formes actuelles, pour un même mot, pouvaient varier dans une très large mesure. C'est ici qu'intervient le système anamorphoseur. C'est l'équi-

..../

valent du bouton des récepteurs de télévision qui permet de comprimer ou d'étirer l'image en hauteur ou en largeur. Grâce à lui on réduit le stéréotype aux dimensions du signal actuel apparu sur I; et dès lors on peut relever le taux de corrélation comme précédemment. Insistons sur le fait qu'une corrélation de 100 % est tout à fait inutile pour reconnaitre le signal; ce taux peut être très faible, pourvu qu'aucun autre stéréotype n'atteigne ce taux. C'est pourquoi tout mot, si mal articulé soit-il, peut tout de même être reconnu.

- juger. C'est mesurer l'écart entre deux formes; il suffit de faire la différence entre deux matrices à tores, opération facilement concevable. Cette différence peut être quantitative (nombre de tores en plus ou en moins) ou qualitative (situation des tores différents dans la configuration considérée.

- associer, combiner. C'est prendre des éléments contenus dans les mémoires - donc appris - et par essais successifs apprécier ce " qui fait bien ensemble " en fonction de certains critères d'association également appris, donc conventionnels par définition. C'est le problème de la composition musicale ou littéraire.

On peut encore incorporer dans ce modèle de fonctionnement tous les mécanismes mentaux ou musculaires que nous trouvons chez l'homme. La réaction musculaire ou mentale dépend évidemment du conditionnement de l'individu qui a associé dans sa mémoire tel signe avec telle signification, tel sous-programme de commande musculaire ou mental.

Tel qu'il est, ce modèle va nous permettre à présent de poser clairement le problème de l'intelligibilité de la parole, d'en mettre en évidence les variables et les conditions pratiques.

III - VARIABLES ET CONDITIONS DE L'INTELLIGIBILITE

Il est évidemment impossible de décrire toutes les possibilités combinatoires entre les variables que nous avons mises en lumière dans notre modèle; nous nous contenterons donc de quelques cas particuliers montrant comment on peut tenter d'approcher le problème de l'intelligibilité dans le but de comprendre un certain nombre de mécanismes.

1er CAS : Tous les éléments de la chaîne de communication sont normalisés, et le canal n'intervient pas. L'opération de reconnaissance est alors la plus simple possible. Une simple superposition entre stéréotype et signal suffit; pas d'anamorphose, une corrélation de 100 %, aucune ambiguïté : c'est " oui " ou " non ". Il s'agit bien sûr d'un cas théorique ....

..../

2° CAS : Toute la chaine est normalisée, sauf le capteur; le canal n'intervenant pas. Nous voici tout de suite plongé dans une complication inextricable, car entre le sourd total et le bien-entendant la destruction du signal par filtrage peut faire varier l'intelligibilité de 100 à 0 %. Une simple déficience dynamique du système ossiculaire peut rendre inintelligible tout ce qui est trop intense : soit la voix elle même, soit le bruit de fond qui y est mélangé. Comme on l'a vu plus haut, le taux de corrélation peut être très faible sans pour celà que la reconnaissance soit rendue impossible; mais l'ambiguïté peut intervenir, et la confusion de termes peut se produire dès lors entre deux ou plusieurs termes. Rappelons de ce point de vue que les risques de confusion sont fortement diminués par le contexte, c'est-à-dire la prévisibilité; l'intelligibilité du message peut être totale, malgré l'absence de nombreux éléments des formes. Lorsqu'un visiteur entre au laboratoire et dit " Bonjour Monsieur; comment allez-vous ? " il suffit de quelques traces infimes pour comprendre le message. Inversement lorsqu'on me pose une question sur un certain sujet alors que je suis en train de réfléchir longuement sur un autre problème particulier, je puis très bien ne pas comprendre une phrase, même si elle est parfaitement articulée et si j'en connais tous les mots, car je cherche alors dans une certaine mémoire des formes qui ne s'y trouvent pas.

3° CAS : Toute la chaîne est normalisée, mais le canal intervient. A lui seul il peut modifier l'intelligibilité de 100 à 0 % par filtrage ou masquage. Tout dépend alors de l'émergence de la forme des mots à chaque instant sur la forme composite du bruit de fond. Insistons encore sur le fait qu'il ne s'agit pas d'un problème d'intensité, de rapport signal/bruit, ou d'un problème de bandes fréquentielles. Dans la réalité l'intelligibilité pose donc chaque fois un cas particulier, impossible à trancher simplement lorsque le bruit de fond est évolutif, mais dont la méthode du sonagraphe permet une approche objective réaliste puisque son originalité est, précisément, de mettre en évidence l'émergence d'une forme sur un fond acoustique.

Un cas particulier est celui d'une espèce de bruit blanc intense. Nous pouvons projeter sur celui-ci n'importe quel stéréotype : la corrélation sera toujours de 100 % par définition; c'est pourquoi nous pouvons positivement " entendre des voix " dans certaines conditions de bruit, par exemple en métro ou en chemin de fer. Mais le contexte visuel nous renseigne sur l'illusion... Rappelons encore quel rôle important joue ici la prévisibilité et en particulier la connaissance préalable de ce que nous allons entendre, surtout quand le vocabulaire et les phrases possibles sont limités. On vérifie couramment qu'il est possible de comprendre totalement une phrase rien que d'après sa durée totale et sans percevoir aucune forme distincte.. Le cas n'est pas rare dans certains métiers bruyants.

...../

4° CAS : Tout est normalisé, sauf le centre de traitement. Mille cas combinatoires peuvent alors se présenter. Imaginons un individu doté d'une mémoire instantanée très courte (50 milli-secondes); il ne pourra jamais entendre un mot entier, ni à fortiori le retenir... donc il ne saura jamais parler, même si son appareil phonatoire est parfait à tous points de vue ! Si d'autre part la mémoire I a une très faible marge dynamique, elle sera continuellement saturée, au moindre petit bruit, et le taux d'intelligibilité peut tomber à zéro si le signal présenté de nombreuses pointes de niveau.

On peut encore imaginer d'autres cas; par exemple celui de l'individu qui ne dispose que d'une bande étroite ne permettant de recevoir qu'une information limitée ! C'est le sourd partiel pour qui l'intelligibilité varie largement selon le cas.

Pour les mémoires T et M elles ont de même des capacités très variables d'un individu à l'autre. Une mémoire T inexistante, celà signifie que l'individu ne pourra rien retenir et ne pourra donc ni acquérir de l'expérience, ni même apprendre à parler : il vivra dans l'actualité immédiate comme l'animal. Par contre il sera " heureux " dans la mesure justement où il oubliera tous les évènements désagréables au fur et à mesure. Certaines expériences semblent montrer que l'abus systématique d'alcool altère justement cette mémoire. Fait curieux en apparence seulement : si l'individu était normal auparavant, il aura des références dans la mémoire mémorisante; il pourra donc sembler vivre tout à fait normalement, faire ce qu'il faisait autrefois; mais il ne peut plus apprendre de mots nouveaux, une langue nouvelle; la notion d'intelligibilité se confond alors avec celle de " trou de mémoire ".

De son côté le système anamorphoseur peut intervenir; si ses performances sont faibles, l'individu comprendra mal certains types de voix !

Enfin, si le classement des stéréotypes dans les mémoires a été fait de façon anarchique (mauvaise pédagogie lors de l'apprentissage de la langue) ou si l'unité de calcul est petite, la parole peut devenir inintelligible lorsque le débit est rapide, car soit le temps d'accès aux mémoires, soit l'opération mentale seront trop longs, et pendant ce temps d'autres mots auront défilé sans être reconnus ou retenus !

Les performances de l'unité de traitement conditionnent en tout premier lieu ce qu'il est convenu d'appeler l'intelligence d'un sujet. Il est certain que ces performances ne serviront à rien, si le sujet n'a stocké au préalable en mémoire des données et des programmes en nombre et en qualité suffisants. Mais toutes choses égales, le sujet bien nanti de ce point de vue " comprendra " toujours plus vite qu'un autre un mot plus ou moins déformé ou détruit : nous l'avons constamment vérifié lors de nos tests sur la parole synthétique.

..../

On pourrait encore définir de très nombreux cas combinatoires entre l'unité de traitement, les capacités et le contenu des diverses mémoires, les propriétés de l'anamorphoseur etc.. tout celà convergerait pour montrer que l'intelligibilité est très largement variable d'un individu à l'autre, indépendamment des caractéristiques du signal physique ou du canal.

Insistons en passant sur le fait qu'il ne s'agit pas pour nous de mettre des noms anatomiques sur les fonctions que nous situons dans le cerveau; mais il est certain qu'elles existent, localisées pour certaines d'entre elles comme la mémoire I, l'anamorphoseur, l'organe de traitement etc... , disposées différemment selon les individus pour ce qui est des éléments contenus en T ou M. Quoiqu'il en soit, nous retrouvons ici tous les problèmes de l'informatique, science récente susceptible d'apporter beaucoup d'idées nouvelles aux physiologistes, mais qui pourrait de son côté, tirer grand profit des observations et des connaissances de ceux-ci.

Résumons. Chaque élément de la chaîne de communication est susceptible de jouer un rôle déterminant dans la reconnaissance des formes du langage parlé; mais d'un individu à l'autre les différences sont tellement considérables que seuls des tests subjectifs statistiques très larges peuvent avoir une signification, étant admis que tout dépend en dernière analyse de la prévisibilité.

Toutes les considérations précédentes se sont lentement élaborées à l'occasion d'une série de travaux variés ou l'intelligibilité était en cause; il n'est donc pas hors de propos de donner quelques précisions sur ceux-ci.

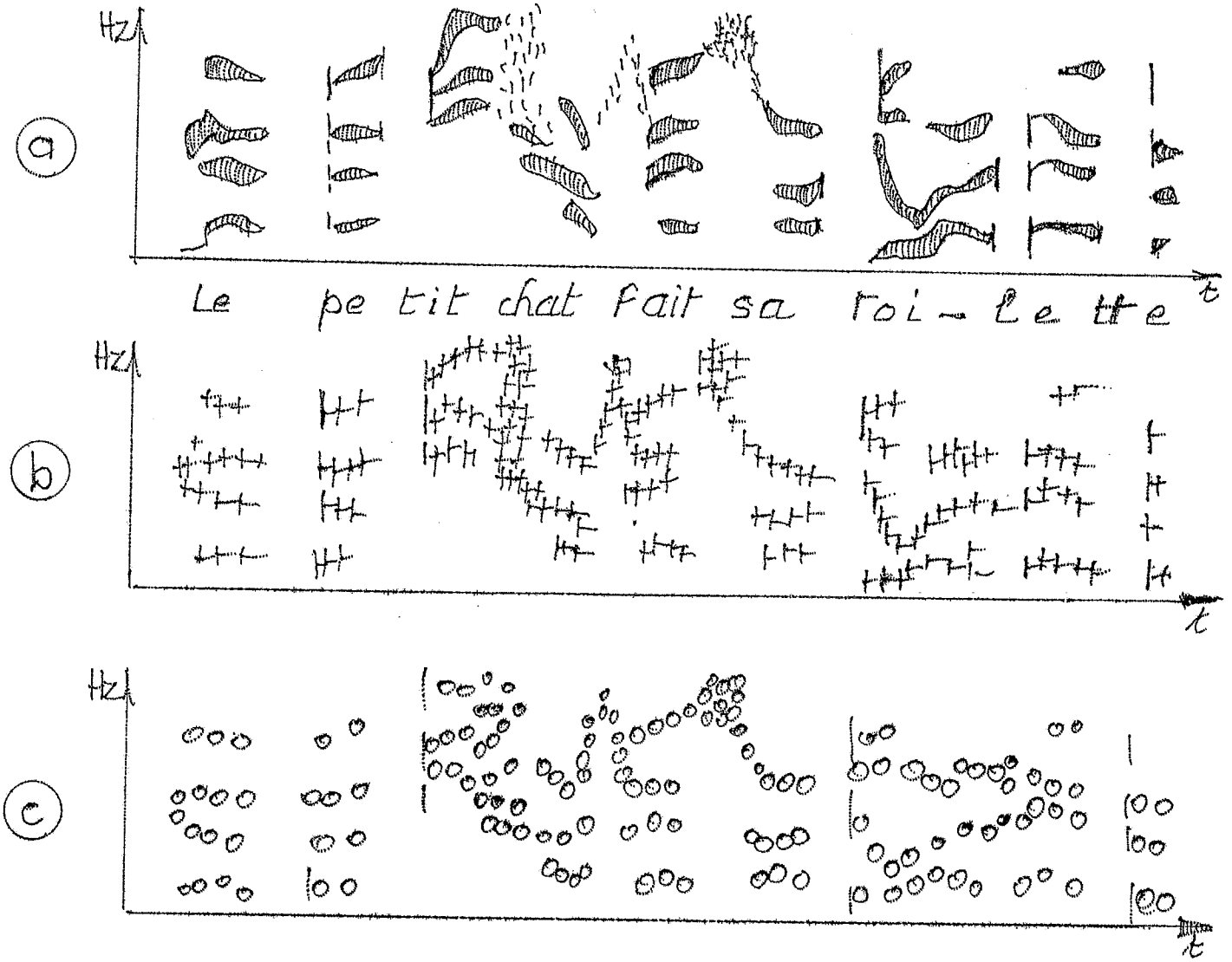
IV - QUELQUES EXEMPLES PRATIQUES DE PROBLEMES D'INTELLIGIBILITE

Notre programme général de recherche et divers hasards heureux nous ont amenés à nous intéresser à divers aspects pratiques de l'intelligibilité de la parole.

1°) LA SYNTHESE DE LA PAROLE. Nous avons fait sur ce sujet de nombreux travaux et des publications (bib. 6, 11, 12, 13, 14) auxquelles nous renvoyons pour le détail. A l'origine, notre but était de faire une étude exhaustive de la structure physique de la parole, problème qui nous est rapidement apparu comme d'une complication extraordinaire résultant en particulier des différences entre locuteurs et de l'existence d'une redondance énorme. La première idée qui nous est donc venue, c'était de chercher à simplifier le problème, de chercher à extraire des formes acoustiques globales de la parole normale ce qui était important pour l'intelligibilité, c'est-à-dire le squelette informatif sémantique. Celà nous a conduit

...../

Fig 8



Les formes sémantiques d'une phrase, intelligible quand elles sont "relues" à l'icophone, le restent même lorsqu'elles sont très grossièrement représentée par des croix (b) ou de gros points. (c). En fait la configuration est la même et si, de plus, on connait la phrase d'avance, les mots semblent parfaitement clairs. Dans le cas contraire, et si on fait d'abord écouter (c) à un sujet non prévenu, la phrase peut être totalement inintelligible.
Le problème des formes visuelles est strictement identique.

à nous intéresser tout spécialement à la parole chuchotée dont nous avions observé qu'elle déterminait sur les sonagrammes des graphismes présentant toutes les caractéristiques d'une forme, au sens de la Gestalttheorie. La parole chuchotée étant totalement intelligible dans nos langues, ces formes représentent précisément le squelette informatif que nous recherchions. Nous avons vérifié effectivement qu'il existait des formes fortes, difficiles à détruire et qu'on reconnait à peu près toujours dans les conditions les plus défavorables.

D'autre part, la méthode de synthèse de parole que nous avons élaborée nous a permis d'expérimenter sur l'anamorphose des formes de parole. Avec l'Icophone, c'est particulièrement commode. Rappelons qu'il s'agit d'un synthétiseur qui lit des formes graphiques grâce à une série de photodiodes dont chacune commande un générateur de sons sinusoïdaux lorsqu'un point de l'image passe devant elle. Tous les signaux sont ensuite mélangés, et on " regarde " proprement la forme dessinée avec l'oreille ! Il est ainsi bien facile d'allonger ou d'élargir une forme quelconque, puis de tester instantanément le résultat à l'audition. La mise au point du répertoire des di-grammes phonétiques pour entrée en machine, nous a directement confronté avec le problème intelligibilité-anamorphose. Il en est découlé toute une série d'observations originales portant sur divers points que nous avons déjà eu l'occasion d'effleurer occasionnellement plus haut.

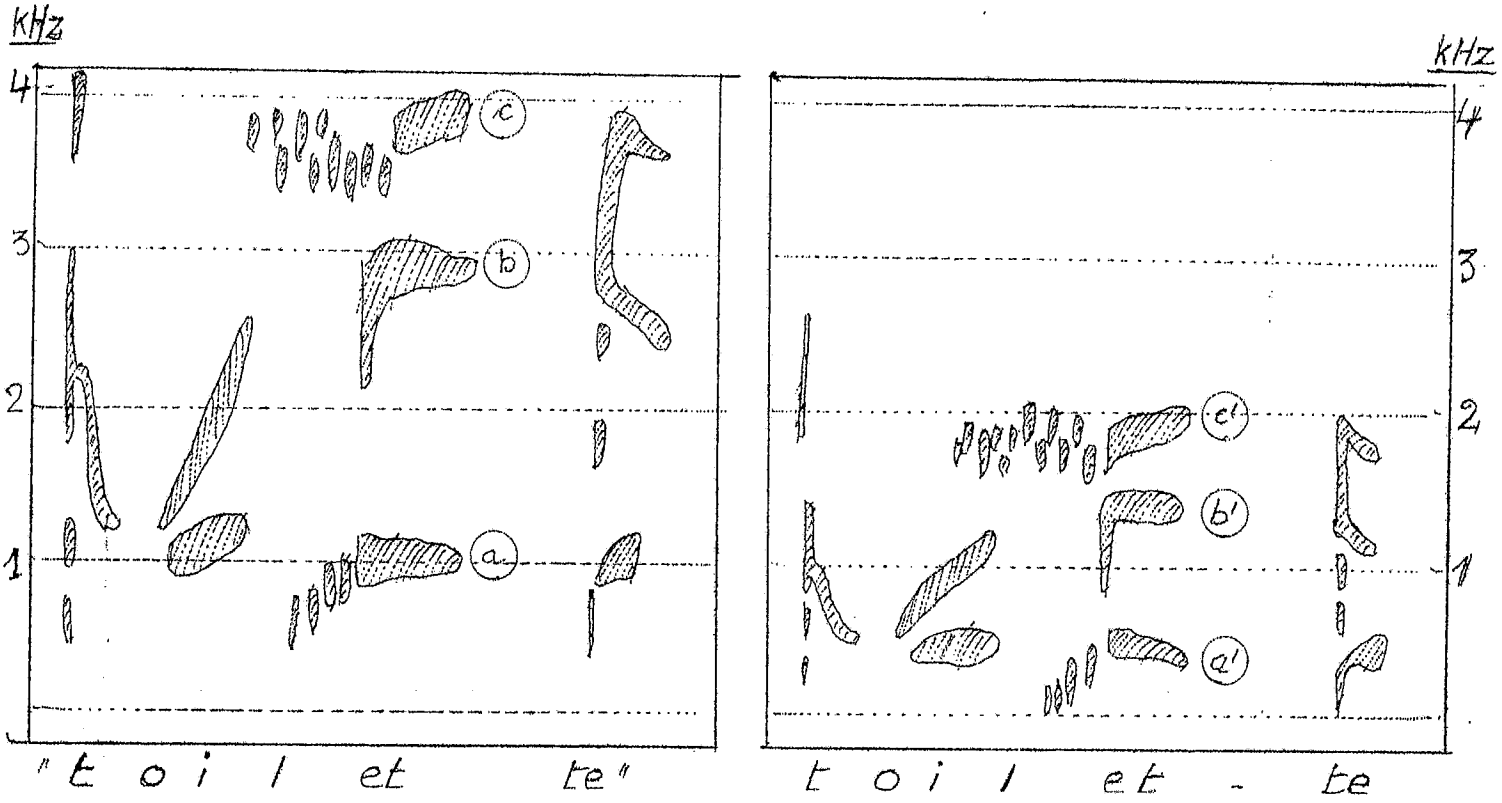
- rôle de la prévisibilité. Elle abrège et facilite grandement l'opération de reconnaissance des formes sémantiques. Nous avons vérifié la toute-puissance de la suggestion. On prévient par exemple un sujet qu'il va entendre le mot "maison". L'auditeur prépare dès lors le stéréotype dans sa mémoire; il a tout le temps qu'il faut pour celà. Tout est prêt pour l'expérience, et on présente alors le mot annoncé, dessiné sur l'icogramme de façon extrêmement grossière : la reconnaissance du mot par le sujet est immédiate et totale. Si l'auditeur n'est pas prévenu, il ne comprend strictement rien, même pas une voyelle ou un phonème. Nous avons été surpris très longtemps et intrigué par ce phénomène tant que nous ne possédions pas un modèle de fonctionnement du système auditif adéquat. Le mot est compris en bloc, comme une totalité, ou n'est pas compris; s'il est prévisible, l'intelligibilité est totale, sinon elle peut être nulle, pour le même signal physique.

- la reconstitution mentale d'une forme partiellement détruite par filtrage ou masquage. Avec la méthode de l'Icophone, l'expérimentation est particulièrement facile. On efface simplement des bandes sur le ruban transparent qui porte les formes. On peut aussi couper tout simplement telle ou telle cellule, le dessin restant intégral; cette méthode est même plus souple. Enfin il est bien facile de détruire la forme d'un mot par des bruits variés, dessinés

..../

cf fig 3d

EXEMPLE REEL D'ANAMORPHOSE correspondant à la figure 3d



Voici,à gauche,la forme sémantique du mot"toilette",réalisée à l'aide de digrammes phonétiques,selon la méthode de synthèse que nous avons imaginée.La fréquence est en ordonnée,le temps en abscisse.Lorsqu'on lit cette forme à l'aide de l'ICOPHONE,on reconnait sans ambigüité le mot.L'expérience montre que si on anamorphose graduellement cette forme dans le sens vertical,le mot reste intelligible dans de très larges limites.Par exemple,à droite,on vérifie que la forme du son "et" est comprimée verticalement du simple au double par rapport avec l'échantillon représenté à gauche.(a'c' est la moitié de ac).Celà signifie que le "formant" c' (2000 Hz environ) qui est le formant 3 ici, est beaucoup plus grave que le formant 2 (b) de l'échantillon de gauche,sans que le mot cesse d'être reconnu pour autant,

Conclusion: il est impossible de definir un "formant" par sa fréquence et la forme est définie non par une frequence,mais par un rapport de frequences.

On vérifie qu'il en est de même pour l'anamorphose temporelle. Finalement la forme sémantique globale est reconnue dans la mesure ou ses rapports de fréquences et ses rapports de durees restent à l'intérieur de certaines limites auxquelles nous sommes habituées et qui correspondent aux différences dimensionnelles des appareils phonatoires humains.

sur une autre bande transparente qu'on superpose tout simplement à la première et qu'on peut d'ailleurs faire glisser pour vérifier auditivement le rôle de tel bruit en tel endroit du mot ou d'une phrase. Les observations que nous avons pu faire ainsi sont très intéressantes. Si le mot est connu au préalable, on peut amputer ou masquer la forme d'une manière extraordinaire, la détruire presque intégralement; le mot reste intelligible. La figure 8 montre un exemple précis d'une expérience. On a dessiné la phrase " le petit chat fait sa toilette "; puis on l'a redessinnée avec des tirets ou de très gros points. Quand on a vu au préalable la forme globale de la phrase normale, on retrouve avec les tirets et les points la même forme générale, plus ou moins " quantifiée "; à l'audition l'expérience montre qu'on comprend parfaitement la phrase dans tous les cas.

Mais si on procède de façon inverse, en faisant venir un sujet non averti pour lui soumettre d'abord les formes quantifiées, celui-ci ne comprend généralement rien du tout. Cependant, pour une quantification assez fine, la forme prend subitement corps et alors le sujet comprend toute la phrase d'un seul coup. Nous avons vérifié d'autre part que des étrangers, parlant assez mal le français, comprennent de toute façon beaucoup moins vite et beaucoup moins bien. Les stéréotypes qu'ils ont en mémoire ne sont pas " épurés ", car ils n'ont pas entendu répéter les mots autant de fois qu'un autochtone; ils ont besoin de la forme intégrale pour la reconnaitre. Chaque phonatome doit être reconnu individuellement et dans ces conditions on s'explique pourquoi l'intelligibilité est beaucoup plus faible pour un sujet étranger.

Une autre expérience a été faite systématiquement par Melle CASTELLENGO. On dessine d'abord un mot normal, bien compris avec l'Icophone. Puis on reprend les formes en les allongeant de plus en plus dans le sens de la hauteur (en étirant l'échelle des fréquences), sans modifier la durée des phonatomes et des mots. On vérifie alors que toutes ces formes sont encore comprises, si le mot est connu en particulier, même pour un degré d'anamorphose énorme où, par exemple, le formant 1 d'un phonatome dépasse en fréquence le formant 2 de la forme normale ! L'expérience est remarquable et montre effectivement pourquoi il est illusoire de vouloir définir la structure physique d'un mot à l'aide de fréquences absolues, comme on le fait couramment. L'étirage de l'image dans le sens horizontal où l'on allonge l'échelle de temps est beaucoup plus facile, puisqu'il suffit de régler la vitesse de défilement de l'image. Les conclusions sont identiques dans une très large marge d'allongement. Ceci met en lumière le rôle important du système anamorphoseur du centre de traitement et de nombreuses observations nous font penser que lorsqu'une personne connue nous aborde, le système s'"accorde" sur la voix du locuteur que nous connaissons, avant qu'il n'ait ouvert la bouche... Pour un inconnu, il faut quelques instants pour réaliser cet accord; dans ces conditions, nous ne comprenons par les premiers mots, et sommes généralement obligés de les faire

...../

répéter, s'ils étaient importants. Très généralement la formule de salutation suffit pour nous laisser le temps d'accorder notre anamorphoseur; elle nous permet ainsi de comprendre sans complications des locuteurs fabriquant pour le même mot des formes très différentes. Ce mécanisme pourrait fort bien expliquer pourquoi nous réussissons à comprendre une personne qui nous parle dans un brouhaha de conversations (problème de la cocktail-party) : il nous suffit d'accorder notre anamorphoseur sur la voix de cette personne; dès lors toutes les autres voix perdent considérablement en intelligibilité et nous ne retenons que celle que nous avons choisie.

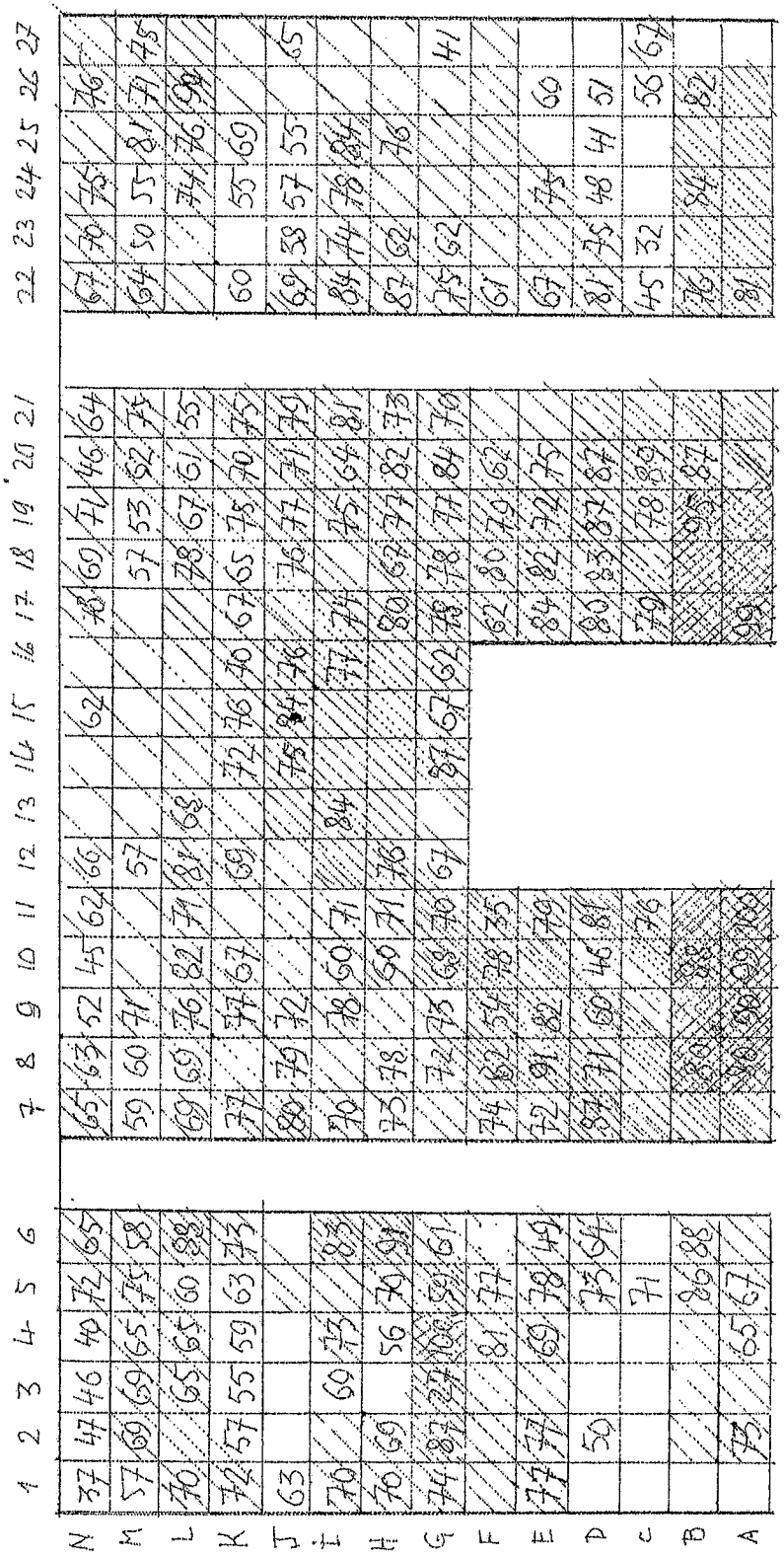
Lors des recherches sur la parole synthétique, que nous continuons à développer d'ailleurs, nous sommes constamment aux prises avec le problème de l'intelligibilité et nous savons maintenant à quel point il faut se défier de la suggestion et de la prévisibilité. Nous pensons que de nombreux chercheurs en synthèse de parole se sont laissé induire en erreur de ce point de vue; étant prévenus, ils comprenaient parfaitement les mots, et s'ils n'ont pas chaque fois fait vérifier l'intelligibilité par des sujets non prévenus, ils ont pu tirer des conclusions erronnées quant à leur méthode de synthèse.

Ces observations ont été complétées par d'autres, qui sont apparues à l'occasion de recherches en acoustique des salle de parole, que nous avons faites récemment.

2°) L'INTELLIGIBILITE EN SALLE DE PAROLE. Ce devrait être le principal souci des architectes qui ont à réaliser des salles de conférence, des amphithéâtres. Malheureusement ce sont des considérations d'ordre financier, visuel, architectural qui priment très généralement; la première surtout, que l'on prétend impérieuse. Or il faut insister dès le départ sur le fait qu'on peut faire une bonne ou une mauvaise salle de parole exactement au même prix ! La vérité c'est qu'il n'existe actuellement aucune doctrine suffisante en acoustique des salles. Les projets, lorsqu'ils existent, se contentent de mentionner la conformité avec la formule de SABINE, relativement à la durée de réverbération, dont l'intérêt est certain, mais qui ne représente qu'une partie infime du problème. En fait, depuis KNUDSEN (Le projet en acoustique architecturale) on ne semble guère avoir avancé. Quelques résultats nouveaux sont bien apparus; SPANDÖCK en Allemagne a repensé le problème des modèles réduits; SCHROEDER aux USA tente d'utiliser les possibilités offertes par les ordinateurs, machines à traiter la complexité. Car c'est de celà qu'il s'agit ! Le problème physique des effets de la salle sur les formes de la parole est dès le départ d'une difficulté inouïe. Difficulté accrue du fait que tout change dès que le locuteur change de place ou de direction; et les résultats varient encore d'un point d'audition à un autre dans une large mesure.

..../

AMPHITHÉATRE 32 Faculté des Sciences de Paris. (Tests d'intelligibilité)



90 à 100%

80 à 90%

70 à 80%

60 à 70%

< 60%

place du locuteur

Les hachures donnent une idée de l'allure d'ensemble de l'intelligibilité. Mais il était nécessaire de rajouter les nombres exacts de l'une des expériences réalisée en présence de 117 étudiants (pour 240 places). Avec un nombre très faible de sujets, les résultats seraient nécessairement moins bons.

On notera des valeurs locales anormales, par exemple aux points F.11 ou H4 et qui sont dues à des déficiences auditives du sujet. Mais lorsqu'on trouve des valeurs très faibles pour plusieurs places voisines, on peut tenir pour assuré que la salle est mauvaise en cet endroit (ex. N1 N2 N3 N4 ou D24 D25 D26 etc.) Il faut de larges statistiques pour obtenir des résultats significatifs par cette méthode.

Finalement il apparait impossible de traiter le problème comme un problème isolé. Réaliser une bonne salle de parole suppose une connaissance approfondie non seulement des problèmes architecturaux ou physiques (absorption, réverbération, échos, etc...) mais aussi et surtout la possession d'une connaissance de la structure physique de la parole et de la perception du langage parlé. Nous l'avons bien vu lorsque nous avons été confrontés avec la réalité ....

L'occasion s'est présentée à propos d'une série d'amphithéâtres construits ces dernières années à la Faculté des Sciences (Halle aux Vins) et dont certains sont très bons, d'autres beaucoup moins. Nous avons dàs lors réfléchi à la possibilité d'imaginer une méthode plus efficace que celle des logatomes classiques, doublée d'une méthode objective suffisamment simple pour être utilisable et suffisamment adaptée au problème pour pouvoir confirmer les résultats de la méthode subjective.

a) La méthode subjective. La méthode des logatomes est actuellement la seule pratique pour obtenir un aperçu statistique des qualités d'intelligibilité d'une salle de parole. Nous avons formulé plus haut un certain nombre de réserves à son égard. Comme on ne voit cependant pas d'autre possibilité pour tester la réalité pratique, la seule solution consiste à essayer de l'améliorer. Ce qui ne va pas dans cette méthode, c'est qu'elle a été élaborée sans tenir compte de la structure sémantique de la parole. Comme nous avons fait des recherches détaillées sur cette question et obtenu des résultats intéressants, il était normal de tenter de l'es appliquer ici.

Une salle de parole est faite pour comprendre des mots, c'est-à-dire pour percevoir et reconnaitre les formes acoustiques normales du langage parlé. On se trouve généralement dans de telles salles pour apprendre des choses que l'on ne sait pas; souvent on y prononce des mots nouveaux, des chiffres, donc en principe des formes imprévisibles. Il est donc de toute nécessité de supprimer toute prévisibilité des mots, sinon les tests n'auront plus guère de sens. Mais par contre il est raisonnable de mettre ces mots dans une super-structure, une phrase, dont la syntaxe est prévisible. En tout cas il ne faut pas utiliser des sons, des syllabes isolées sinon on se place dans des conditions artificielles et on ne pourra plus tirer de conclusions relativement à la parole normale, fluide, comportant des formes plus ou moins franchement séparées. Nous savons bien qu'un discours normal comprend généralement un nombre limité de types de phrases comportant par exemple la succession : article, nom, verbe, compléments variés; ou bien : verbe, adjectif, nom etc... Pour concilier les conditions requises (imprévisibilité des mots, prévisibilité de la structure de la phrase) il suffit de prendre ou de construire des phrases telles que l'ensemble, réalisé avec des mots du dictionnaire, n'ait pas de sens, mais ait une allure de phrase normale. On

...../

peut affirmer alors que si la phrase est comprise, elle le sera à fortiori dans le cas où intervient la prévisibilité. La transcription sera bien entendu phonétique en cas de besoin : on ne cherche pas à tester l'orthographe !

Partant de ce raisonnement nous avons mis sur pied deux tests pratiques, utilisés concuremment, et dont on additionne les résultats :

- Nous utilisons une série de phrases normales, dans lesquelles nous repérons 50 phonatomes variés, donc des phrases à syntaxe classique, mais qui n'ont pas de signification, les mots étant choisis en conséquence. Par exemple : " les nuages rageaient contre un temps cuisiné. " etc... On peut prendre des textes littéraires hermétiques, si on le désire, qui ont le mérite d'être tous faits; ainsi nous avons utilisé des textes de MALLARME, ne comportant donc que des " mots de la tribu ".

- Une deuxième série de tests utilise des phrases supprimant la prévisibilité à l'intérieur des mots; on fabrique des phrases de syntaxe toujours classique, mais on utilise des mots forgés de toutes pièces, mais " sonnant français ". On notera de ce point de vue qu'une salle jugée bonne pour le français n'est pas nécessairement un optimum pour l'anglais ou l'allemand dont la structure phonétique n'est pas identique. Si on est en mal d'imagination, il existe des textes tout faits, par exemple dans RABELAIS ou MICHAUX et autres; mais on se rappellera qu'il ne faut jamais prendre deux fois de suite le même texte avec les mêmes sujets. On dicte ce texte et on y pointe encore 50 phonatomes significatifs.

Finalement on additionne les résultats des 2 épreuves et on obtient directement un pourcentage d'intelligibilité hautement significatif, en particulier si on a fait lire les deux textes par deux personnes différentes dont les voix et la façon de parler et d'articuler sont très dissemblables.

On peut alors définir la qualité d'une salle de parole en fonction de ces résultats. On dessine pour celà un plan de la salle avec des cases représentant les places numérotées. On porte dans chaque case le pourcentage d'intelligibilité; on peut aussi hachurer différemment selon le taux. Finalement on obtient une configuration hautement significative des qualités de la salle (fig.9). Les résultats que nous avons obtenus, nous ont empiriquement conduit aux taux d'intelligibilité suivants :

	95 à 100 %		la salle est parfaite
de	90 à 95 %		la salle est bonne
de	80 à 90 %		salle acceptable lorsqu'il s'agit d'un vocabulaire connu par les auditeurs
de	70 à 80 %		salle déficiente
en dessous de 70 %			,la salle est mauvaise

...../

Nous avons donc fait ces tests dans plusieurs amphithéâtres de la Faculté des Sciences, établi les taux, porté ceux-ci sur un plan, à chaque place, pour détecter les bonnes ou les mauvaises zones d'intelligibilité. La comparaison entre divers amphis a été tout à fait intéressante et significative, car elle correspond avec ce que disent les usagers habituels de la salle : locuteurs et auditeurs.

Mais pour que de tels essais aient une valeur, il faut une salle occupée normalement, soit au 3/4 des places environ. Ceci est toujours une affaire assez compliquée et nous avons cherché à remplacer cette méthode par une autre, qui pourrait être faite en salle vide, et fournirait des documents objectifs confirmant les résultats de la méthode précédente. Nous avons donc imaginé et expérimenté dans les mêmes salles en utilisant diverses méthodes dont nous allons décrire celle qui nous a semblé la plus intéressante.

b) La méthode objective. Nos recherches sur la structure sémantique de la parole nous ayant montré qu'il s'agissait d'un problème de reconnaissance de formes par un récepteur humain, une méthode objective de l'intelligibilité en salle consiste à mettre au point une méthode qui rende compte des déformations et amputations de ces formes.

Nous utilisons alors trois magnétophones autonomes de qualité professionnelle, avec trois microphones similaires. Le premier est placé à très petite distance de la bouche du locuteur (20 cm) : on élimine ainsi à ce point, l'influence de la salle. Un décibelmètre portatif est placé à un mètre de la bouche du locuteur. Celui-ci lit alors normalement deux ou trois phrases types, choisies pour contenir des formes acoustiques variées, recouvrant toute la bande de la parole. Le locuteur règle l'intensité de sa voix en regardant le décibelmètre; par exemple à 80 dB, et on enregistre un texte très court .

Simultanément les deux autres magnétophones enregistrent les mêmes phrases en deux points différents de la salle. L'idéal serait de disposer 5 magnétophones; mais on peut toujours recommencer le test en divers points de toutes façons.

On tire alors les sonagrammes de toutes ces phrases et on les compare.

L'expérience nous a rapidement montré que ces documents étaient d'un intérêt extraordinaire. Prenons un exemple :

Soit la même phrase, que nous avons donc relevée à la source et en l'un des points de la salle. On recopie le premier sonagramme (ou on le photographie) sur papier calque en lavis rouge; le sonagramme au point d'audition est

..../

de même recopié en vert. On superpose les deux calques et on les observe dans une boite à lumière, par transparence.

En brun-noir (somme du vert et du rouge) on aura tout ce qui est commun au signal à la source et au signal à distance; c'est de cette zone brune que le " centre de traitement " du récepteur devra extraire les éléments de forme permettant de reconnaitre le mot. Comme dans un sonagramme convenablement établi on lit tout ce qu'on voit et inversement, on peut rapidement se faire une idée de l'intelligibilité dans la mesure où on sait lire un sonagramme et y reconnaitre la forme sémantique des mots utilisés; celle-ci a été étudiée au préalable en laboratoire.

En rouge, on lira tout ce que la salle a détruit (qui n'existe que dans le signal de près)

En vert, on aura ce que la salle ajoute : échos, résonances, trainages etc...)

Nous n'avons fait qu'ébaucher cette méthode, mais il est facile de comprendre son intérêt dans la mesure où elle fournit un document objectif sur les modifications de forme de mots prononcés normalement et dans leur intégralité.

Comme il est facile de quantifier une image sonagraphique, on entrevoit une solution intéressante pour traiter le problème avec l'aide d'un ordinateur. C'est un problème auquel nous pensons, mais il nous faudra d'abord développer nos recherches dans les autres amphithéâtres et d'autres salles, ce que nous avons prévu à notre programme.

En attendant, nous avons cherché une méthode encore plus simple qui résumerait les précédentes. L'idée est apparue à travers nos résultats dans le domaine de la parole. Lorsqu'on observe attentivement des séquences de parole synthétique bien faite à l'Icophone, on voit qu'il est possible de " résumer " la parole en deux points :

Les formes comportent normalement deux éléments : un spectre de raies assez large (entre 100 et 6000 Hz environ) et des phénomènes impulsionnels très brefs (explosives en particulier), dont la durée est de quelques millisecondes. Nous avons donc imaginé des signaux-types réglables, qui simulent les deux caractéristiques essentiel du signal physique de la parole.

Nous utilisons alors :

- un instrument de musique simplifié, saxophone à trois notes différentes, qui fournit des spectres de raies riches, tout à fait comparables à ceux de la voix humaine.

- une crécelle ordinaire en bois.

Avec l'instrument de musique nous jouons les trois notes, successivement, les unes après les autres, en laissant un temps d'arrêt chaque fois. Puis on joue les mêmes notes répétées suc-

..../

cessivement le plus vite possible avec des coups de langue appropriés. Pour la crécelle, nous la faisons tourner cran par cran, puis nous accélérons progressivement.

Les trois magnétophones sont placés comme précédemment ainsi que le décibelmètre qui permet de régler l'intensité des sons musicaux,

On tire ensuite les sonagrammes de l'opération complète et on procède par simple comparaison visuelle ou par recopie en couleur des documents superposés et observés par transparence.

Nous avons observé que ces deux tests aboutissaient au même résultat que les tests précédents beaucoup plus compliqués et plus longs à dépuiller. Diverses notions ont très clairement émergé de ces essais, permettant de préjuger de l'intelligibilité à partir de ces tests faits en salle vide. On peut les résumer ainsi :

La parole est d'abord un phénomène fréquentiel; le filtrage des formes et les résonances de la salle apparaissent très nettement et peuvent être chiffrés à partir des sons musicaux; on voit comment et dans quelle mesure avec une trop grande réverbération ou avec des échos, les sons successifs rapides s'interpénètrent. Les sons musicaux simulant assez bien les sons " voisés ", si deux sons successifs émis à 200 millisecondes d'intervalle ne se discernent plus sur le sonagramme, on peut être certain que l'intelligibilité sera très mauvaise. On peut vérifier sur le document ce qui se passe dans les diverses régions fréquentielles, voir, par exemple s'il existe dans une région aigüe assez de redondance pour oompenser le mauvais pouvoir séparateur dans telle région grave. De toutes façons on aura toujours le signal au niveau de la source comme référence.

Pour ce qui est de l'aspect micro-temporel de la parole (explosives, par exemple), les sonagrammes comparatifs de crécelle seront extrêmement parlants. De près, on verra une série de hachure verticales se rapprochant de plus en plus; en un certain point, par exemple lorsque les traits sont distants de 10 millisecondes, les hachures commencent à fusionner et devenir indistinctes. Sur le sonagramme relevé à la place de l'auditeur, cette fusion interviendra bien plus tôt en raison de la réverbération de la salle; par exemple si la fusion se fait dès 50 ou à fortiori 100 millisecondes, on peut être assuré que toutes les plosives seront détruites, et avec elles les phonatomes qui en contiennent. Il est donc possible de chiffrer très précisément le " pouvoir séparateur impulsionnel " dans la salle, hautement significatif de l'intelligibilité.

Signalons que le test des sons musicaux rejoint une méthode que nous avons étudiée pour déterminer la qualité d' une église du point de vue de l'orgue qui s'y trouve (bib.16). En effet un son tenu que l'on bloque subitement avec la langue sur l'anche du " saxophone " utilisé, met en évidence sur

..../

le sonagramme un diagramme fréquence-temps tout à fait caractéristique de l'acoustique de la salle considérée; on y lit très simplement la durée des résonances à chaque fréquence. D'autre part, avec la crécelle on est immédiatement renseigné sur le filtrage de la salle et sur les échos que l'on peut d'ailleurs chiffrer facilement en donnant après la crécelle, quelques coups de claquette de faible niveau.

Nous avons en tout cas vérifié que moyennant un peu d'entrainement, cette méthode simple et expéditive permettait de définir comparativement les propriétés des salles de parole du point de vue intelligibilité. Nous poursuivons actuellement nos recherches sur ce point et signalons que la méthode est utilisable pour contrôler éventuellement l'efficacité d'un " traitement " insonorisant d'une salle dans la mesure où on peut faire les enregistrements avant et après l'opération. Nos résultats, que nous ne pouvons détailler ici, feront l'objet de publications ultérieures; nous en avons parlé pour montrer de quelle manière on peut envisager une approche réaliste du problème de l'intelligibilité dans les salles de parole et pour donner des exemples concrets de définition objective des paramètres d'un canal " naturel " dont le rôle peut être déterminant en fin de la chaîne de communication des messages parlés.

D'autres occasions nous ont d'ailleurs été fournies qui nous ont permis d'approfondir nos observations générales sur l'intelligibilité de la parole dans le bruit.

3°) L'INTELLIGIBILITE DE LA PAROLE DANS LE BRUIT

Un problème particulièrement intéressant nous a été posé par un fabricant de magnétophones spéciaux pour aviation. On sait que pendant la durée du vol, la conversation entre les membres de l'équipage est intégralement enregistrée afin qu'il existe un document en cas d'accident. Si l'atterrissage est normal, la bande est effacée en bloc à l'arrivée. Or dans la cabine de pilotage le bruit est toujours très élevé, et détruit par conséquent largement les formes sémantiques de la parole. Toutes les considérations vues plus haut restent valables ici, mais il convient de faire diverses observations.

D'abord, le vocabulaire dans la cabine de pilotage est très limité et très prévisible; l'intelligibilité est donc bien meilleure pour le personnel navigant que pour un auditeur quelconque. Le problème posé dans ces conditions au fabricant de magnétophones est très particulier. En effet, la technique électro-acoustique admet à priori que l'appareil le meilleur est toujours celui qui est techniquement le plus perfectionné, le plus " fidèle ". C'est une opinion contre laquelle nous nous sommes élevé souvent déjà. Un appareil, une méthode ne sont pas les meilleurs quand ils sont les plus parfaits, mais quand ils sont les mieux adaptés au problème à traiter. Ainsi, nous avons l'exemple du système auditif, dont on sait qu'il est techniquement très imparfait, mais qui s'avère parfaitement adapté quand

..../

il s'agit d'extraire une forme déformée et amputée sur un fond compliqué et mouvant. Le but n'est pas de fournir des mesures précises de fréquence ou de niveau, mais de mettre en évidence et de traiter des formes.

Le problème qui nous était posé est devenu rapidement clair, à l'observation d'une série de sonagrammes d'enregistrements faits dans des cabines de pilotage. On y voit des bandes de bruit intenses sur lesquelles on repère bien des formes vocales que nous savions lire sur le sonagramme, des mots que nous savions reconnaître; mais à l'audition ces bandes saturent l'oreille du fait de leur intensité et masquent ainsi le reste de la forme qui existe très visiblement dans le signal physique. Il faut donc éviter cette saturation, ce que nous avons facilement réussi à faire en coupant les bandes de bruit intense avec un filtre de réjection; en procédant ainsi, on ampute bien entendu les formes des mots plus ou moins largement, mais comme le vocabulaire est limité et largement prévisible, notre système central extraira, à partir de ce qui reste de la forme des mots, un taux de corrélation suffisant pour les reconnaître par confrontation avec les stéréotypes stockés en mémoire.

Conclusion : dans ce cas particulier, il faut fabriquer un " mauvais " magnétophone, au sens usuel du terme, muni de filtres de réjection permettant d'éliminer les bandes nuisibles du bruit. Si ce magnétophone doit être universel, il suffira de le doter de filtres de réjection à bandes réglables en fréquence et en largeur; on adaptera alors les filtres aux particularités du bruit rayonné par tel ou tel type d'avion et que l'on aura analysé au préalable au laboratoire. Signalons en passant que c'est là un problème relativement facile à trancher, le bruit d'un avion étant un signal quasi-invariable.

Les expériences que nous avons faites sur ce sujet ont été extraordinairement démonstratives; même pour des personnes non informées du vocabulaire utilisé par le pilote, la réjection des bandes nuisibles restitue une parole totalement intelligible en pratique.

Nous avons cité ce cas parce que nous l'estimons important : il montre quels aspects paradoxaux peut revêtir le problème de l'intelligibilité et indique une voie qui serait peut-être intéressante et permettrait de tenter une approche de certains problèmes mal résolus dans le domaine de la communication des messages parlés, entre autres celui du brouillage des émissions radio.

De toutes façons, nous pensons à présent avoir cerné l'essentiel du problème de l'intelligibilité et nous pouvons d'ores et déjà tirer un certain nombre de conclusions générales.

V. CONCLUSION GENERALE

Nous avons proposé ici un modèle de fonctionnement de la chaine de communication des messages parlés qui nous a permis de poser dans son ensemble le problème de l'intelligibilité. Nous avons insisté sur le fait que, si l'anatomie et la physiologie de l'émetteur sonore, l'appareil phonatoire, étaient bien connus, la question de la structure physique sémantique des sign aux de la parole l'était beaucoup moins. Les travaux que nous avons faits de ce point de vue nous ont permis de montrer que le problème de la parole est un cas particulier du problème très général de la génération, transmission perception, reconnaissance et intégration, de formes; celles-ci correspondent à des archétypes au niveau de la mémoire des mouvements, à des "programmes de fabrication" des signaux de la parole. Ces archétypes appris se traduisent, au niveau des mémoires du récepteur, par des stéréotypes utilisés comme références pour comparer avec les sign aux qui arrivent à l'oreille et les reconnaitre par corrélation.

Nous avons insisté sur le fait que le seul chainon relativement bien connu était le canal, mais que le fonctionnement du récepteur restait largement un mystère. Nous en avons alors proposé un modèle de fonctionnement, basé sur ce que nous savons de l'informatique, et avons défini les fonctions nécessaires et suffisantes pour que le système puisse expliquer les réactions que nous observons dans la réalité journalière.

Cette simulation ne prétend pas à la similitude anatomique, mais à la similitude des fonctions. Telle qu'elle est actuellement, elle s'est avérée extrêmement précieuse pour comprendre les nombreux problèmes que pose l'intelligibilité de la parole. Il doit être clair que ce modèle de fonctionnement représente une hypothèse de travail, un outil à penser, dont la valeur ne réside nullement dans le fait qu'elle soit exacte, mais dans le fait qu'elle soit prégnante et qu'elle ouvre des voies nouvelles d'investigation, permettant de repenser l'ensemble du problème compliqué dont il est question ici.

Nous connaissons bien les critiques que pourraient être amenés à formuler certains spécialistes concernés directement ou indirectement par ces questions. De toutes façons nous n'avons cherché qu'à donner un aperçu d'ensemble de nos préocupations, de nos résultats et de nos idées: divers points particuliers feront l'objet de publications ultérieures. Par exmeple nous envisageons une étude de l'intelligibilité dans le chant, qui représente un cas particulier d'anamorphose temporelle des formes sémantiques. Une autre étude est en cours relativement

.../

au fonctionnement électro-mécanique de l'oreille etc.

Nous espérons simplement que les idées émises ici, en raison même de leur allure peu orthodoxe, retiendront l'attention de certains et donneront l'impulsion à des recherches susceptibles de faire progresser le problème de l'intelligibilité de la parole et, beaucoup plus ambitieusement, celui des mécanismes perceptifs et intégratifs humains en général.

LEIPP

PARIS, 4 Juin 1968

BIBLIOGRAPHIE

limitée essentiellement aux publications du Groupe d'Acoustique sur l'intelligibilité (n°s 3 à 16)

1° CANAC(F) — L'acoustique des Théâtres antiques. Ses ensignements Ed. CNRS 1967

2. LEHMANN (R) — Etude psycho-physique de l'intelligibilité du langage
Annales Télécomm. 1962

3.LEIPP (E) — La chaine de communication du message musical
Cahiers d'Etude de la Radio-Télévision
Flammarion Paris 1960

4. LEIPP (E) — Die subjektive Bowertung der Musik
7. Tonmeistertagung. Cologne. Oct. 1966
Ed. Westdeutscher Rundfunk 1968

5. LEIPP (E) — Le problème du bruit
Bulletin GAM n° 20 mai 1966 Ed. Int. Fac.Sciences Paris

6. LEIPP (E) — Information sémantique et parole; essai d'une gestaltthéorie de la parole. Bulletin GAM n° 22 Juin 1965

7. LEIPP (E) — Les variables de l'audition musicale
Conférence Journées d'Etude. CHIRON,Paris 1966

8. LEIPP (E) — Acoustique et Orthophonie
Rééducation Orthophonique n° 27 Paris 1967

9. LEIPP (E) — Un modèle fonctionnel de l'audition
Exposé, Séminaire Dr. HECAEN Hôpital Ste Anne.

10. LEIPP (E) — Mécanique et acoustique de l'appareil phonatoire
Bulletin GAM n° 32 Déc. 1967

11. LEIPP (E) - CASTELLENGO (M) - LIENARD (J.S)
Parole et Gestaltthéorie
Colloque GALF Lannion mai 1966. Ed Interne Fac.Sciences

12. LEIPP (E) - CASTELLENGO (M) - LIENARD (J.S) - SAPALY (J)
Structure physique et contenu sémantique de la parole
Colloque GALF Grenoble Avril 1967. Ed.Int. Fac.Sciences
A paraître dans Revue Française d'acoustique

13. LEIPP (E) — Contenu informatif de la parole
Comptes Rendus Congrès Intern.Ac. BUDAPEST Oct.1967

14. LEIPP (E) - CASTELLENGO (M) - LIENARD (J.S)
La synthèse de la parole à partir de digrammes phonétiques C.R. Congrès International Acoustique TOKIO (sept. 1968)

15. LEIPP (E) — Le problème de la perception des signaux acoustiques par effet de contraste. Annales Télécomm. T.20 n° 5-6. 1965

16. LEIPP (E)
a) Méthode d'appréciation des quakités musicales d'un ensemble orgue - salle. C.R. ICA Liège Sept 1965

b) Peut-on apprécier objectivement les qualités musicales d'un orgue. Revue ORGUE n°118 (Avril Juin 1966 p. 70....

N O T E

du 10 Novembre 1968

Le texte qui précède a été rédigé en mai dernier et devait faire l'objet, en juin, de notre dernière réunion du G.A.M. Entre temps de nombreux éléments nouveaux sont apparus dans nos recherches sur l'intelligibilité.

Les travaux conjugués du Groupe d'Acoustique et du Groupe de Mécanique Appliquée du Laboratoire de Mécanique Physique d'une part, du Groupe de calcul hybride du Centre de Calcul Analogique du C.N.R.S. d'autre part ont permis très récemment, de réussir la synthèse en temps réel d'une parole basée sur les phonatomes en utilisant un ordinateur IBM 1130. A notre réunion du 8 Novembre nous avons eu le plaisir de présenter le premier échantillon de parole synthétique réalisé ainsi. La possibilité de fabriquer en continu et en temps réel des mots et des phrases, à la cadence où on les pense, est un événement important, dans la mesure, surtout, où toute analyse préalable des phrases et des mots est inutile ici.

Dans cette méthode, au lieu de dessiner les " formes " des mots au pinceau et à l'encre, en associant les diagrammes phonétiques tels que nous les avions définis, l'opération est faite quasi instantanément. On tape phonétiquement le texte sur un clavier classique de machine à écrire. Avec un peu d'entraînement on arrive à suivre la cadence normale de la parole, comme en sténotypie GRANDJEAN par exemple. Grâce à une sortie " son " spécialement construite (Icophone 03) on entend directement la parole synthétique moyennant agencement d'un programme spécial dont la mise au point a demandé bien entendu de longues semaines de travail. Nous pouvons reprendre ainsi certains essais d'intelligibilité de façon beaucoup plus efficace.

Ce supplément d'information explique diverses remarques de la discussion qui a suivi notre réunion du 8 Novembre dernier. Nous pensons d'ailleurs refaire ultérieurement une réunion sur ce thème.

DISCUSSION

de la réunion du 8/11/1968

M. GUEU - Votre méthode de synthèse aura certainement des incidences dans le problème très actuel de traduction automatique des langues. Si j'ai bien compris, la parole que vous synthétisez est une espèce de parole chuchotée, où l'on élimine l'influence des cordes vocales.

M. LEIPP - Exactement. Je dois préciser que toute notre méthode de synthèse est basée sur les particularités de la voix chuchotée. Celà nous a permis de simplifier énormément le problème. En effet les " dessins " que nous réalisons pour simuler la voix ne représentent qu'une partie infime de l'information contenue dans la parole normale; ils représentent une véritable " caricature " où nous ne conservons que ce qui est strictement indispensable pour retrouver l'information sémantique d'un mot. Nous opérons précisément comme le caricaturiste qui a su extraire les quelques traits nécessaires et suffisants pour reconnaître le personnage qu'il veut représenter. Pour le " récepteur ", une caricature n'est pas nécessairement plus inintéressante qu'une photographie : elle représente au contraire, une économie. L'important y est mis en relief, et nous ne perdons pas de temps à traiter la complexité de l'image réelle dont la redondance ne représente d'intérêt que dans la mesure où elle permet de résister à la destruction des formes lorsqu'elles passent dans un canal de mauvaise qualité !

M. MOLES - Vous savez combien l'idée de modèle et d'exploitation systématique de modèle me tient à coeur. L'insertion en France dans le domaine technique, de la théorie de la Gestalt est l'ABC d'une grande partie de la psychologie mondiale.

Je rappelle que la simulation d'un processus de reconnaissance par passage d'une mémoire transitoire à une mémoire permanente a été prévue la première fois par GRAY et WALTER avec ses modèles. D'autre part, l'ICOPHONE est assez voisin du Play back de HASKINS.

Mais les spécialistes qui ont tenté de faire de la parole synthétique avec le Play back, sont partis des données que leur ont fourni les linguistes, en les admettant comme des dogmes. L'intérêt de votre travail est d'avoir abordé le problème en rejetant comme erronnée l'existence même de ce qu'on appelle les phonèmes et en la remplaçant par la notion d'éléments phonétiques, de phonatomes qui représentent le mouvement entre deux phonèmes. La meilleure façon de suivre les maîtres est souvent de les contester... Je crois que vous l'avez fait d'une façon efficace et économique.

Je voudrais encore attirer l'attention sur le fait qu'il existe maintenant des méthodes de transformation de Gestalt, de formes, par voie mathématique. Il serait intéressant de passer à l'ICOPHONE les figures de transformation ainsi obtenues et d'écouter comment " ça sonne ". Celà permettrait sans doute d'explorer des lois de transitions sonores sur lesquelles on ne connait strictement rien actuellement.

DISCUSSION (suite)

M. LEIPP - Je vais répondre le plus brièvement possible à vos observations :

- Le modèle fonctionnel que j'ai élaboré n'est original et intéressant que dans la mesure où ses réactions se raccordent avec celles que nous observons normalement chez l'homme et en fournissent une explication raisonnable, s'appuyant d'ailleurs sur un certain nombre de données que nous ont apportées récemment les psycho-physiologues.

Pour ce qui est du play back de HASKINS, l'ICOPHONE est plus élaboré et plus fiable en raison de sa conception et des composantes électroniques utilisées, celà au dire même de l'utilisateur principal, DELATTRE, qui nous a rendu visite il y a quelques mois.

Quant aux idées reçues, nous pensons que le rôle même de tout laboratoire de recherche est précisément de ne pas les recevoir ou, du moins, de les tester de façon permanente, sinon l'efficacité du laboratoire sera de pauvre espèce !

Si nous soutenons que les " phonèmes " n'ont aucune existence réelle dans la structure physique de la parole normalement articulée, ce n'est pas pour le simple plaisir de contredire les théories établies. Tout notre travail repose sur l'observation que la forme acoustique d'un mot résulte d'un ordre de mouvements cohérent de l'appareil phonatoire déterminé par un programme appris, stocké dans notre mémoire.

Ainsi, prenons les phonèmes " ch " et " a ". Ils représentent chacun un élément statique. L'expérience montre qu'on n'entendra jamais le mot " cha " en associant ces deux phonèmes. Pour nous le phonème n'existe pas, mais " cha ", association des deux phonèmes qui représente le mouvement entre les positions " ch " et " a " représente vraiment l'atome, insécable, de la parole, atome que nous appelons " élément phonétique " ou"phonatome ".

Lorsque nous utilisons donc des phonatomes pour faire de la parole synthétique, non seulement nous faisons une économie extraordinaire, mais, de plus, nous obtenons une parole fluide composée d'éléments dynamiques qui s'enchainent " naturellement " et celà sans avoir besoin de passer par une analyse préalable du mot au sonagraphe... C'est grâce à cette économie que nous avons pu aboutir à des résultats intéressants avec des moyens limités.

Nous sommes d'autre part tout à fait convaincus que le problème des " Formes " vocales mérite d'être repris à l'aide des méthodes mathématiques de transformation, mais ce n'est pas notre affaire.

M. PIMONOV - Si on veut économiser de l'énergie, il faut utiliser des spectres discrets; en chuchotant, vous utilisez beaucoup plus d'énergie qu'en parole normale.

M. LEIPP - Oui, bien entendu ! Mais celà est exact lorsque vous réalisez les mêmes niveaux. Nous savons justement que la parole nor-

...../

DISCUSSION (suite)

male porte plus loins parce que la forme sémantique est modulée par le spectre de raies des cordes vocales. L'intérêt de la parole chuchotée est justement de permettre d'envoyer à un récepteur une forme sémantique avec une énergie tellement faible que les autres récepteurs, placés un peu plus loin, s sont incapables de la percevoir.

M. PIMONOV - Voici une autre observation : je pense que la notion de " forme " a été le piège tendu aux techniciens et aux théoriciens par les philosophes et les psychologues du siècle dernier ! Il faut maintenant en sortir.

Un même objet vu de face ou de profil n'a généralement rien de commun du point de vue forme : vous reconnaissez l'objet, pas la forme. Donc dans votre mémoire ce n'est pas la forme que vous conservez, mais la description de l'objet considéré. Grâce à celà, vous pouvez reconnaitre l'objet même s'il présente un aspect tout à fait différent.

M. LEIPP - La nature résoud toujours ses problèmes par la voie la plus économique. Qu'est-ce qui est plus économique ? Stocker dans un tiroir que nous appellerons " mémoire " une description numérique sous forme de tableaux de chiffres ou de graphiques ou y stocker l'objet lui-même ? Je pencherais assez pour la deuxième réponse ! Je conçois donc aisément que nous puissions stocker dans notre mémoire un objet quelconque (codé, celà n'a pas d'importance, en configuration " volumique "). Ce stockage n'est évidemment possible que si nous avons appris, au préalable, à connaître la chose dans ses trois dimensions. Si nous l'avons observé sous tous ses angles. Cet objet ainsi stocké en mémoire, nous pouvons à loisir, en cas de besoin, " l'appeler " et le regarder sous tous les angles afin de superposer l'image qu'il donne avec celle qui apparait à un moment donné sur notre " mémoire instantanée ". Si nous stockons l'objet lui-même, considéré comme gestalt à trois dimensions, nous aurons réalisé une extraordinaire économie de place comparativement à la description par liste numérique. Enfin tout celà relève de l'hypothèse.

M. MIZZI - Vous utilisez des mots de combien de bits ?

M. QUINIO - 16 bits par mot.

M. MIZZI - Quel est votre algorithme pour pouvoir travailler en temps réel ? Comment faites-vous pour retrouver les diagrammes phonétiques dans votre fichier ?

M. QUINIO - Il nous a fallu une zone mémoire et une zone tampon . On manipule les zones tampon pour décharger une des zones chargées.

Pour ce qui est de la recherche de l'élément désiré, nous savons où il est. Nous avons optimisé le temps de recherche en utilisant les résultats d'un travail fait naguère sur CAB 500 par J.S. LIENARD et TEIL; les éléments phonétiques sont classés par taux d'occurence décroissant dans la langue Française.

...../

DISCUSSION (suite)

M. MIZZI - Peut-on penser qu'avec 300 ou 400 éléments phonétiques on puisse vraiment représenter la plupart des mots français ?

M. LEIPP - On peut en être certain puisque nous l'avons démontré par la pratique. Il ne s'agit pas de la plupart des mots, mais de leur intégralité, y compris les néologismes et tout ce qu'on voudra. Moyennant quelques additifs au répertoire des diagrammes phonétiques que nous avons maintenant en trois exemplaires différents, on peut sans autre complication parler en allemand, en anglais ou toute autre langue. Nous avons donné quelques échantillons tout à l'heure pour le montrer.

M. MIZZI - Dans les problèmes de documentation automatique, c'est alors un progrès extraordinaire.

M. SIESTRUNCK - Nous avons déjà fait un peu de prospective sur les possibilités de cette méthode de synthèse de parole. D'ici peu elle sera opérationnelle. Nous savons que l'avenir est plein de promesses, dans la mesure où il nous sera matériellement possible de développer ces questions.

Paris le 10 Novembre 1968